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

ARTICLE INFO

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

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

ABSTRACT

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
1063-5203/© 2015 Elsevier Inc. All rights reserved.

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

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

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*

$$A = \tilde{U}\tilde{\Lambda}\tilde{V}^T \tag{5}$$

with $\tilde{U} \in \mathbb{R}^{N\times r}$ ($\tilde{U}^T\tilde{U} = I$), $\tilde{\Lambda} \in \mathbb{R}^{r\times r}$ a positive diagonal matrix and $\tilde{V} \in \mathbb{R}^{M\times r}$ ($\tilde{V}^T\tilde{V} = I$) where the matrices $\tilde{U}, \tilde{V}, \tilde{\Lambda}$ are defined by the shifted eigenvalue problem

$$\begin{aligned} A\tilde{V} &= \tilde{U}\tilde{\Lambda} \\ A^T\tilde{U} &= \tilde{V}\tilde{\Lambda} \end{aligned} \tag{6}$$

with the additional condition that all diagonal elements of $\tilde{\Lambda}$ are non-zero positive numbers.

As stated by Lanczos [14, p. 672] it is remarkable that the principal axes associated with the zero eigenvalue do not participate at all in the formation of the matrix A.

In the next section we will show now that the shifted eigenvalue problem (6) can be obtained from a variational principle, different from (3).

3. A new variational principle

For the given matrix A, let us define two sets of data points, corresponding to the rows and columns of the matrix, respectively: $x_i = A^T\epsilon_i$, $z_j = A\varepsilon_j$ for $i = 1, \ldots, N$, $j = 1, \ldots, M$ where $\epsilon_i, \varepsilon_j$ denote standard basis vectors, of dimension N and M, respectively. The vectors ϵ_i and ε_j are the column vectors of the identity matrices I_N and I_M, respectively. Related to these data points the following linear feature maps are defined:

$$\begin{aligned} \varphi(x_i) &= C^T x_i = C^T A^T \epsilon_i \\ \psi(z_j) &= z_j = A\varepsilon_j \end{aligned} \tag{7}$$

where $\varphi : \mathbb{R}^M \to \mathbb{R}^N$, $\psi : \mathbb{R}^N \to \mathbb{R}^N$. We call the matrix $C \in \mathbb{R}^{M\times N}$ a compatibility matrix, which enables that the data points $x_i \in \mathbb{R}^M$, $z_j \in \mathbb{R}^N$ can be compared with each other after applying the feature maps.

Consider now the following constrained optimization problem (primal problem)

$$\begin{aligned} \min_{w,v,e,r} J(w, v, e, r) = & -w^T v + \frac{1}{2}\gamma \sum_{i=1}^{N} e_i^2 + \frac{1}{2}\gamma \sum_{j=1}^{M} r_j^2 \\ \text{subject to} \quad & e_i = w^T\varphi(x_i),\ i = 1, \ldots, N \\ & r_j = v^T\psi(z_j),\ j = 1, \ldots, M \end{aligned} \tag{8}$$

with $e_i, r_j \in \mathbb{R}$ and $w, v \in \mathbb{R}^N$. Thanks to the compatibility matrix C the feature maps φ, ψ and their corresponding vectors w, v have compatible dimensions. The non-zero and finite parameter γ is trading-off the first term versus the other terms in the objective. The solution to this problem can be related then to Lanczos' decomposition theorem:

Theorem 2. *If the following compatibility condition holds for the matrix C*

$$ACA = A \tag{9}$$

then the Karush–Kuhn–Tucker conditions to (8) result in the shifted eigenvalue problem

$$\begin{aligned} A[\beta] &= [\alpha]\tilde{\Lambda} \\ A^T[\alpha] &= [\beta]\tilde{\Lambda} \end{aligned} \tag{10}$$

where all diagonal elements of $\tilde{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_r)$ *are non-zero positive numbers and* $\lambda_l = 1/\gamma_l$ *for* $l = 1, \ldots, r$ *with* r *the rank of matrix* A*. The matrices* $[\alpha] \in \mathbb{R}^{N \times r}, [\beta] \in \mathbb{R}^{M \times r}$ *consist of* r *Lagrange multiplier vectors* $\alpha = (\alpha_1, \ldots, \alpha_N)^T$ *and* $\beta = (\beta_1, \ldots, \beta_M)^T$ *as its columns, related to the first and second set of constraints in (8), respectively.*

Proof. Take the Lagrangian

$$\mathcal{L}(w, v, e, r; \alpha, \beta) = J(w, v, e, r) - \sum_i \alpha_i \left(e_i - w^T \varphi(x_i)\right) - \sum_j \beta_j \left(r_j - v^T \psi(z_j)\right). \tag{11}$$

The Karush–Kuhn–Tucker conditions result into

$$\begin{array}{l} \frac{\partial \mathcal{L}}{\partial w} = 0 \;\Rightarrow\; v = \sum_i \alpha_i \varphi(x_i) \\ \frac{\partial \mathcal{L}}{\partial v} = 0 \;\Rightarrow\; w = \sum_j \beta_j \psi(z_j) \\ \frac{\partial \mathcal{L}}{\partial e_i} = 0 \;\Rightarrow\; \gamma e_i = \alpha_i,\ \forall i \\ \frac{\partial \mathcal{L}}{\partial r_j} = 0 \;\Rightarrow\; \gamma r_j = \beta_j,\ \forall j \\ \frac{\partial \mathcal{L}}{\partial \alpha_i} = 0 \Rightarrow e_i = w^T \varphi(x_i),\ \forall i \\ \frac{\partial \mathcal{L}}{\partial \beta_j} = 0 \Rightarrow r_j = v^T \psi(z_j),\ \forall j. \end{array} \tag{12}$$

Next one can eliminate w, v, e, r by substituting the first four expressions into the last two. This gives

$$\begin{array}{rcl} \left[\varphi(x_i)^T \psi(z_j)\right] [\beta] & = & [\alpha]\tilde{\Lambda} \\ \left[\psi(z_j)^T \varphi(x_i)\right] [\alpha] & = & [\beta]\tilde{\Lambda} \end{array} \tag{13}$$

which follows from collecting all solutions corresponding to non-zero λ (i.e. finite $\gamma = 1/\lambda$ values) of $\lambda \alpha_i = \sum_j \beta_j \psi(z_j)^T \varphi(x_i)$ for $i = 1, \ldots, N$ and $\lambda \beta_j = \sum_i \alpha_i \varphi(x_i)^T \psi(z_j)$ for $j = 1, \ldots, M$. Here $\left[\varphi(x_i)^T \psi(z_j)\right]$ denotes the matrix with ij-entry equal to $\varphi(x_i)^T \psi(z_j)$.

Using (9) one obtains

$$\varphi(x_i)^T \psi(z_j) = \epsilon_i^T ACA\varepsilon_j = \epsilon_i^T A\varepsilon_j = A_{ij}$$

resulting into (10). □

Corollary 1. *The eigenvector solutions to (10) and their corresponding eigenvalues lead to an objective value* J *evaluation in (8) equal to zero.*

Proof. From the Karush–Kuhn–Tucker conditions in the proof of Theorem 2 it follows that $w = \sum_j \beta_j \psi(z_j)$, $v = \sum_i \alpha_i \varphi(x_i)$ and $\gamma e_i = \alpha_i$, $\gamma r_j = \beta_j$. Substituting this in (8) gives

$$\begin{array}{rcl} -w^T v + \frac{1}{2}\gamma \sum_{i=1}^N e_i^2 + \frac{1}{2}\gamma \sum_{j=1}^M r_j^2 & = & -\sum_j \beta_j \psi(z_j)^T \sum_i \alpha_i \varphi(x_i) + \frac{\gamma}{2} \sum_i \frac{\alpha_i^2}{\gamma^2} + \frac{\gamma}{2} \sum_j \frac{\beta_j^2}{\gamma^2} \\ & = & -\lambda \beta^T \beta + \frac{\lambda}{2}(\alpha^T \alpha + \beta^T \beta) \\ & = & 0 \end{array} \tag{14}$$

using the fact that $\alpha^T \alpha = \beta^T \beta$ (see also [14, p. 668]) and all λ values are non-zero in connection to Theorems 1 and 2. □

Corollary 2. *Taking as objective function* $-J$ *instead of* J *(i.e. flipping the signs of the terms in the objective) results into the same solution (10).*

Corollary 3 *(Compatibility matrix). Equation (9) $ACA = A$ can be solved by the vec operation using the property that $\mathrm{vec}(ACA) = (A^T \otimes A)\mathrm{vec}(C)$ where C follows then from solving the linear system*

$$(A^T \otimes A)\mathrm{vec}(C) = \mathrm{vec}(A).$$

Given the matrix A one has then NM equations in the NM number of unknown elements of matrix C. However, solving for $\mathrm{vec}(C)$ would require then the inversion of a possibly large matrix of size $NM \times NM$. In case of ill-conditioning one can employ the pseudo-inverse for solving the linear system.

For a full rank matrix A the compatibility matrix C is directly obtained in terms of the pseudo-inverse of matrix A:

- *case $N > M$: take $C = A^\dagger = (A^T A)^{-1} A^T$ satisfying $ACA = A$;*
- *case $N < M$: take $C^T = (A^T)^\dagger = (AA^T)^{-1} A$;*
- *case $N = M$: take $C = A^{-1}$,*

which ensures that $A^T A$ is invertible in the different cases and the compatibility condition (9) is satisfied.

Remark 1 *(Compatible feature maps).* An alternative choice for the feature maps is

$$\begin{aligned} \varphi(x_i) &= C_1^T x_i = C_1^T A^T \epsilon_i \\ \psi(z_j) &= C_2 z_j = C_2 A \varepsilon_j \end{aligned} \tag{15}$$

with $C_1 \in \mathbb{R}^{M \times q}$, $C_2 \in \mathbb{R}^{q \times N}$ and $q \geq \max\{N, M\}$. In this case the compatibility condition becomes $AC_1C_2A = A$. If A is full rank then the matrices C_1, C_2 can be chosen to satisfy: $C_1C_2 = A^\dagger$ (if $N > M$); $C_1C_2 = (A^T)^\dagger$ (if $N < M$); $C_1C_2 = A^{-1}$ (if $N = M$). Depending on the choices of C_1, C_2 this leads then to different feature maps.

Remark 2 *(Primal and dual representations).* The approach is model-based where the model $\mathcal{M}$ has a primal (P) and dual (D) representation [23], related to the primal optimization problem (8) and its solution in terms of the Lagrange multipliers, respectively:

$$\begin{tikzcd} & (P):\ \begin{array}{l} e_i = w^T \varphi(x_i) \\ r_j = v^T \psi(z_j) \end{array} \\ \mathcal{M} \arrow[ur] \arrow[dr] & \\ & (D):\ \begin{array}{l} e_i = \sum_j \beta_j \psi(z_j)^T \varphi(x_i) \\ r_j = \sum_i \alpha_i \varphi(x_i)^T \psi(z_j). \end{array} \end{tikzcd} \tag{16}$$

One can also consider an out-of-sample extension related to this model which would correspond to either adding a row or a column to the matrix A.

Remark 3. The choice of the value γ in (8) is treated at the selection level. The value is chosen such that $\lambda = 1/\gamma$ is an eigenvalue of the shifted eigenvalue problem (10).

Remark 4 *(Eigenvalues in primal problem).* Elimination of e and r in the primal (8) gives the problem $\min_{w,v} -w^T v + \frac{1}{2}\gamma \sum_{i=1}^N [w^T \varphi(x_i)]^2 + \frac{1}{2}\gamma \sum_{j=1}^M [v^T \psi(z_j)]^2$ for which $1/\gamma$ can be chosen then to correspond to the eigenvalues in $R_x w = (1/\gamma) v$, $R_z v = (1/\gamma) w$ where $R_x = \sum_i \varphi(x_i)\varphi(x_i)^T$, $R_z = \sum_j \psi(z_j)\psi(z_j)^T$. Note that both $R_x, R_z \in \mathbb{R}^{N \times N}$.

Please cite this article in press as: J.A.K. Suykens, SVD revisited: A new variational principle, compatible feature maps and nonlinear extensions, Appl. Comput. Harmon. Anal. (2015), http://dx.doi.org/10.1016/j.acha.2015.09.004

4. Nonlinear extensions

Singular value decomposition has also been studied within the context of integral equations since the work of Schmidt [18,19], related to unsymmetric kernels. Here we will propose an alternative way to introduce kernels into the formulation and to consider nonlinear extensions to the singular value decomposition.

The data sets $\{x_i\}_{i=1}^N$, $\{z_j\}_{j=1}^M$ are obtained from the given matrix A. From these a matrix B is then obtained

$$B_{ij} = f(x_i, z_j) \tag{17}$$

with $f : \mathbb{R}^M \times \mathbb{R}^N \to \mathbb{R}$ where the map f is realized through compatible feature maps (15) $\varphi(x_i) = C_1^T x_i = C_1^T A^T \epsilon_i$, $\psi(z_j) = C_2 z_j = C_2 A \varepsilon_j$ for which $\varphi : \mathbb{R}^M \to \mathbb{R}^q$, $\psi : \mathbb{R}^N \to \mathbb{R}^q$ with $q \geq \max\{N, M\}$. The nonlinear mapping is obtained then through the compatible feature maps e.g. by

$$B_{ij} = K(\varphi(x_i), \psi(z_j)) \tag{18}$$

with a kernel function $K : \mathbb{R}^q \times \mathbb{R}^q \to \mathbb{R}$. This kernel function is not restricted to be positive definite. It will typically also be unsymmetric given that N and M are usually different. In fact a general nonlinear function can be taken for it.

Both with respect to matrix A and B data sets are defined then, together with the corresponding compatible feature maps:

$$\begin{array}{cc}
\text{Matrix } B\text{:} & \text{Matrix } A\text{:} \\
\\
\underline{x}_i = B^T \epsilon_i & x_i = A^T \epsilon_i \\
\underline{z}_j = B \varepsilon_j & z_j = A \varepsilon_j \\
\\
\underline{\varphi}(\underline{x}_i) = \underline{C}_1^T B^T \epsilon_i & \varphi(x_i) = C_1^T A^T \epsilon_i \\
\underline{\psi}(\underline{z}_j) = \underline{C}_2 B \varepsilon_j & \psi(z_j) = C_2 A \varepsilon_j
\end{array} \tag{19}$$

where $\underline{C}_1, \underline{C}_2$ are the compatibility matrices related to the B matrix with corresponding compatible feature maps $\underline{\varphi}, \underline{\psi}$.

The nonlinear mapping (18) has the following property then

$$B_{ij} = \underline{\varphi}(\underline{x}_i)^T \underline{\psi}(\underline{z}_j) = K(\varphi(x_i), \psi(z_j)) \tag{20}$$

provided that the compatibility condition $B\underline{C}_1\underline{C}_2 B = B$ holds.

Corollary 4 *(Symmetric matrix case). In the special case $A = A^T$ the data sets $\{x_i\}$, $\{z_j\}$ coincide. There is no compatibility issue in this case as φ is the only feature map here. One has then*

$$B_{ij} = \underline{\varphi}(\underline{x}_i)^T \underline{\varphi}(\underline{x}_j) = K(\varphi(x_i), \varphi(x_j)) \tag{21}$$

with a general kernel and $\underline{x}_i = B^T \epsilon_i, x_i = A^T \epsilon_i$.

Corollary 5 *(Special case: kernel eigenvalue decomposition). A further special case to the symmetric case in Corollary 4 is obtained by taking a positive definite kernel applied to $\varphi(x_i), \varphi(x_j)$ by $K(\varphi(x_i), \varphi(x_j))$. A more specific case is $K(x_i, x_j) = \varphi(x_i)^T \varphi(x_j)$ with $x_i = A^T \epsilon_i$, $\varphi(x_i) = x_i$. For the latter the primal problem (8) reduces to $\min_{w,e} -w^T w + \frac{\gamma}{2} \sum_i e_i^2$ subject to $e_i = w^T x_i$ for $i = 1, \ldots, N$. This corresponds to the primal formulation given in [22] to kernel principal component analysis [16] for the case of a linear*

Please cite this article in press as: J.A.K. Suykens, SVD revisited: A new variational principle, compatible feature maps and nonlinear extensions, Appl. Comput. Harmon. Anal. (2015), http://dx.doi.org/10.1016/j.acha.2015.09.004

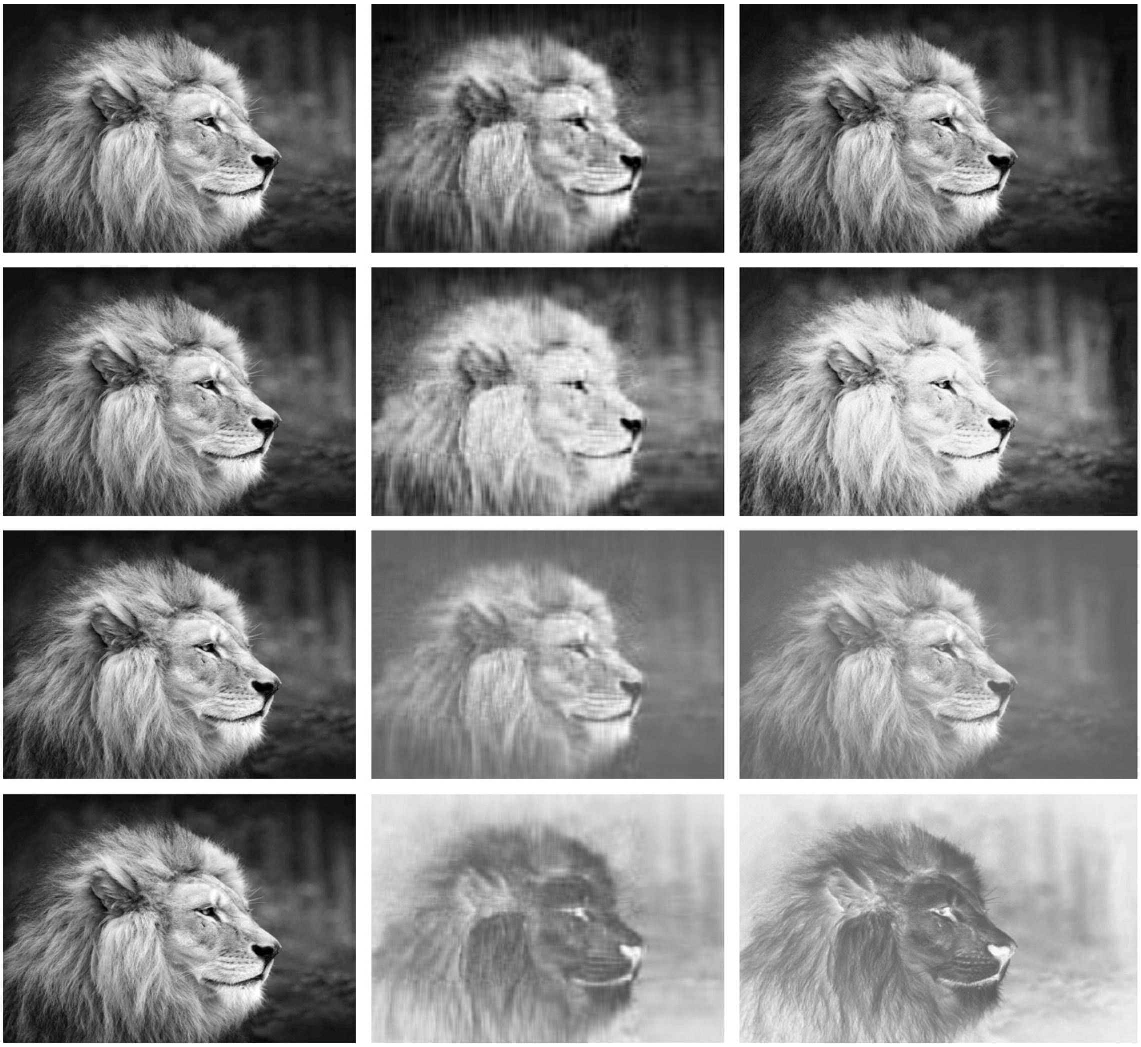

Fig. 1. SVD and nonlinear SVD on lion image: Column 1: original image; column 2: reconstruction based on 20 components; column 3: 100 instead of 20 components; Row 1: SVD; row 2: combination of linear and polynomial kernel with $c = 0.5, d = 2$; row 3: exponential with $\eta = 1$; row 4: exponential with $\eta = -1$.

kernel. The dual problem in this case becomes $[x_i^T x_j]\alpha = (1/\gamma)\alpha$ *with* $x_i = A\epsilon_i$ *and* $[x_i^T x_j]$ *denoting the kernel matrix with* ij*-th entry* $x_i^T x_j$.

5. Illustrative examples

We illustrate now the extension to nonlinear SVD on a few examples. In Figs. 1 and 2 a lion and butterfly image example is taken (source: free-picture.net[1,2]) with image size matrices 325×520. In this case $N < M$. From the image matrix A the data sets $\{x_i\}_{i=1}^N$ and $\{z_j\}_{j=1}^M$ are obtained. The matrix B is then obtained by

$$B_{ij} = K(\varphi(x_i), \psi(z_j)) \tag{22}$$

[1] http://www.free-picture.net/animals/lion/lion-predator__1372073224.jpg.html.

[2] http://www.free-picture.net/Butterflies/butterfly-26.jpg.html.

Please cite this article in press as: J.A.K. Suykens, SVD revisited: A new variational principle, compatible feature maps and nonlinear extensions, Appl. Comput. Harmon. Anal. (2015), http://dx.doi.org/10.1016/j.acha.2015.09.004

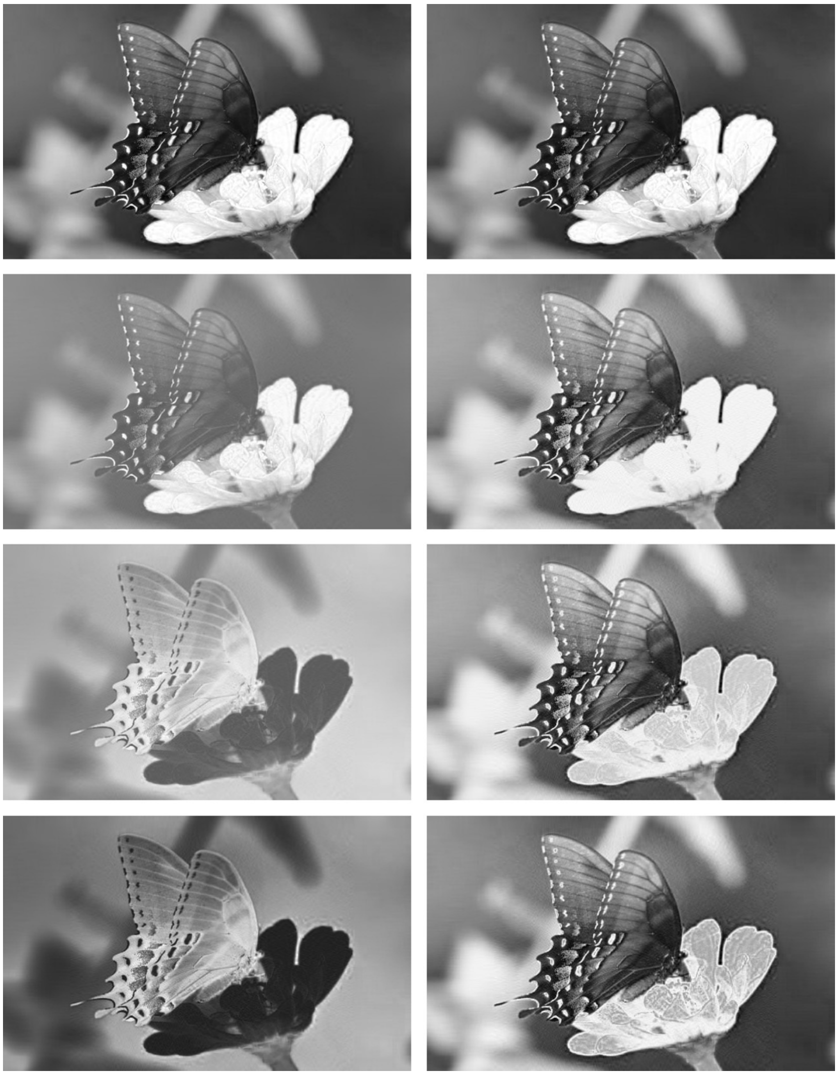

Fig. 2. Illustration of changing η in the exponential kernel (column 1) (row 1: original; row 2: $\eta = 1$; row 3: $\eta = -1$; row 4: $\eta = -2$) and d in the combined linear and polynomial kernel with $c = 0.5$ (column 2) (row 1: original; row 2: $d = 2$; row 3: $d = 4$; row 4: $d = 8$); 100 components are taken in all images.

with $\varphi(x_i) = C^T x_i$, $\psi(z_j) = z_j$ and $C^T = (A^T)^\dagger$ is taken, as explained in Corollary 3. The following kernels K are illustrated:

- linear:

$$K(\varphi(x_i), \psi(z_j)) = \varphi(x_i)^T \psi(z_j) = x_i^T C z_j$$

  corresponding to the SVD.
- combination of linear and polynomial:

$$K(\varphi(x_i), \psi(z_j)) = \varphi(x_i)^T \psi(z_j) - c \left(\varphi(x_i)^T \psi(z_j)\right)^d$$

  with c, d positive.
- exponential:

Please cite this article in press as: J.A.K. Suykens, SVD revisited: A new variational principle, compatible feature maps and nonlinear extensions, Appl. Comput. Harmon. Anal. (2015), http://dx.doi.org/10.1016/j.acha.2015.09.004

$$K(\varphi(x_i), \psi(z_j)) = \exp\left(\eta\, \varphi(x_i)^T \psi(z_j)\right)$$

with η positive or negative.

In Fig. 1 the nonlinear SVD is taken and reconstructions are shown for 20 and 100 components for the different kernels. In Fig. 2 the effect of the η value of the exponential kernel and d in the combination of the linear and polynomial kernel is illustrated. The pictures are obtained in Matlab using imshow by dividing the B_{ij} elements by $\max_{ij} B_{ij}$ after reading the original pictures by imread.

6. Conclusions

In this letter a new variational principle to the SVD has been proposed in connection to Lanczos' decomposition theorem and the shifted eigenvalue problem. The formulation is conceived within the setting of least squares support vector machines, with primal and dual representations. The unsymmetric nature of the problem leads to a compatibility condition that should hold related to the feature maps. The new formulation also enables to consider nonlinear SVDs. Possible future directions of research may include e.g. robustly weighted versions and new approaches to tensor decomposition methods.

Acknowledgments

The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007–2013) / ERC AdG A-DATADRIVE-B (290923). This paper reflects only the authors' views, the Union is not liable for any use that may be made of the contained information; Research Council KUL: GOA/10/09 MaNet, CoE PFV/10/002 (OPTEC), BIL12/11T; PhD/Postdoc grants; Flemish Government: FWO: PhD/Postdoc grants, projects: G.0377.12 (Structured systems), G.088114N (Tensor based data similarity); IWT: PhD/Postdoc grants, projects: SBO POM (100031); iMinds Medical Information Technologies SBO 2014; Belgian Federal Science Policy Office: IUAP P7/19 (DYSCO, Dynamical systems, control and optimization, 2012–2017).

References

[1] C. Alzate, J.A.K. Suykens, A regularized kernel CCA contrast function for ICA, Neural Netw. 21 (2–3) (2008) 170–181.

[2] C. Alzate, J.A.K. Suykens, Multiway spectral clustering with out-of-sample extensions through weighted kernel PCA, IEEE Trans. Pattern Anal. Mach. Intell. 32 (2) (2010) 335–347.

[3] E. Beltrami, Sulle funzioni bilineari, G. Mat. Uso Stud. Univ. 11 (1873) 98–106.

[4] C. Cortes, V. Vapnik, Support vector networks, Mach. Learn. 20 (1995) 273–297.

[5] F. Cucker, D.-X. Zhou, Learning Theory: An Approximation Theory Viewpoint, Cambridge University Press, 2007.

[6] C. Eckart, G. Young, The approximation of one matrix by another of lower rank, Psychometrika 1 (3) (1936) 211–218.

[7] G.H. Golub, C.F. Van Loan, Matrix Computations, Johns Hopkins University Press, Baltimore, MD, 1989.

[8] D.R. Hardoon, S. Szedmak, J. Shawe-Taylor, Canonical correlation analysis: an overview with application to learning methods, Neural Comput. 16 (2004) 2639–2664.

[9] R.A. Horn, C.R. Johnson, Matrix Analysis, Cambridge University Press, 1985.

[10] H. Hotelling, Relation between two sets of variates, Biometrika 28 (1936) 322–377.

[11] C. Jordan, Mémoire sur les formes bilinéaires, J. Math. Pures Appl., Deuxieme Sér. 19 (1874) 35–54.

[12] C. Jordan, Sur la réduction des formes bilinéaires, C. R. Acad. Sci., Paris 78 (1874) 614–617.

[13] P.L. Lai, C. Fyfe, Kernel and nonlinear canonical correlation analysis, Int. J. Neural Syst. 10 (5) (2000) 365–377.

[14] C. Lanczos, Linear systems in self-adjoint form, Amer. Math. Monthly 65 (1958) 665–679.

[15] R. Mall, R. Langone, J.A.K. Suykens, Multilevel hierarchical kernel spectral clustering for real-life large scale complex networks, PLOS One 9 (6) (2014) e99966.

[16] B. Schölkopf, A. Smola, K.-R. Müller, Nonlinear component analysis as a kernel eigenvalue problem, Neural Comput. 10 (1998) 1299–1319.

[17] B. Schölkopf, A. Smola, Learning with Kernels, MIT Press, Cambridge, MA, 2002.

[18] E. Schmidt, Zur Theorie der linearen und nichtlinearen Integralgleichungen. I. Teil. Entwicklung willkürlichen Funktionen nach System vorgeschriebener, Math. Ann. 63 (1907) 433–476.

[19] G.W. Stewart, On the early history of the singular value decomposition, SIAM Rev. 35 (4) (1993) 551–566.

[20] G. Strang, Linear Algebra and Its Applications, Books/Cole, Cengage Learning, 2006.

[21] J.A.K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, J. Vandewalle, Least Squares Support Vector Machines, World Scientific, Singapore, 2002.

[22] J.A.K. Suykens, T. Van Gestel, J. Vandewalle, B. De Moor, A support vector machine formulation to PCA analysis and its kernel version, IEEE Trans. Neural Netw. 14 (2) (2003) 447–450.

[23] J.A.K. Suykens, C. Alzate, K. Pelckmans, Primal and dual model representations in kernel-based learning, Stat. Surv. 4 (2010) 148–183.

[24] V. Vapnik, Statistical Learning Theory, John Wiley & Sons, New York, 1998.

[25] G. Wahba, Spline Models for Observational Data, Ser. Appl. Math., vol. 59, SIAM, Philadelphia, 1990.

Please cite this article in press as: J.A.K. Suykens, SVD revisited: A new variational principle, compatible feature maps and nonlinear extensions, Appl. Comput. Harmon. Anal. (2015), http://dx.doi.org/10.1016/j.acha.2015.09.004