

# Principal Component Analysis With Complex Kernels

Athanasios Papaioannou, *Student Member, IEEE*, Stefanos Zafeiriou, *Member, IEEE*,

**Non-linear complex representations, via the use of complex kernels, can be applied to model and capture the nonlinearities of complex data. Even though, the theoretical tools of Complex Reproducing Kernel Hilbert Spaces (CRKHS) have been recently successfully applied to the design of digital filters and regression and classification frameworks there is limited research on component analysis and dimensionality reduction in CRKHS. The aim of this paper is to properly formulate the most popular component analysis methodology, i.e. Principal Component Analysis (PCA), in CRKHS. In particular, we define a general Widely Linear Kernel Complex PCA (WLCKPCA) framework. Furthermore, we show how to efficiently perform Widely Linear PCA (WLPCA) in small sample sized problems. Finally, we show the usefulness of the proposed framework in robust reconstruction using Euler data representation.**

**Index Terms**—Principal Component Analysis, Complex Kernels, Pattern Recognition, Machine vision

## I. INTRODUCTION

**P**RINCIPAL component analysis (PCA), which appeared first in [1], is one of the most widely used data analysis tool for dimensionality reduction, lossy data compression, feature extraction and data visualization [2]. Arguably, the application of PCA to face representation and dimensionality reduction [3] was the starting point for the vast development of the very popular field of component analysis and subspace learning [4]–[8].

PCA is used to transform a number of observed correlated variables into a smaller number of uncorrelated ones, which account for as much of the total variance as possible. The PCA is originally defined on real-valued random variables [9]. However, complex data arise in many applications such as analysis of signal from radars, sonars, image processing etc. Recently, there is a lot of research towards developing neural networks with complex data [10]–[13]. In the case of PCA, in order to deal with complex-valued data, the direct extension of real PCA was firstly proposed, the so-called Circular Complex Principal Component Analysis (CCPCA) [14], [15]. In CCPCA only the Hermitian covariance matrix is taken into account and hence the information of the pseudo-covariance matrix is not taken into consideration. The underlying assumption of CCPCA is that the data are proper and circular<sup>1</sup>. In order to amend this and exploit the general non-circularity of complex data [17], [18] a Widely Linear PCA (WLPCA) was proposed in [15]. In [15] only the case that the signal dimensionality is smaller than the number of samples was considered. Recently, a Probabilistic Complex Non-circular PCA (PCNCPA) [19] was proposed by formulating and solving a Maximum Likelihood (ML) optimization problem. The PCNCPA, even though it does not assume a widely linear model,

Manuscript received December 1, 2012; revised December 27, 2012. This paper was completed as part of the postgraduate programme co-financed by the Act “Scholarships programme SSF (State Scholarships Foundation/IKY) with an individualized assessment process of the Academic Year 2011-2012” from resources of the Operational Programme “Education and Lifelong Learning”, of the European Social Fund (ESF), the NSRF 2007-2013.

The authors are with the Department of Computing, Imperial College London, United Kingdom, 180 Queen’s Gate, London SW7 2AZ, UK (email: apapaion@imperial.ac.uk)

<sup>1</sup>A complex random signal is proper if it is uncorrelated with its complex conjugate. A signal is circular iff  $\mathbf{x}$  and  $\mathbf{x}' = e^{j\alpha}\mathbf{x}$  have the same probability distribution for any given real  $\alpha$  [16]

it takes into account the pseudo-covariance matrix. Furthermore, due to its probabilistic nature it explicitly models data noise but does not have a closed form solution (i.e., optimization is performed using gradient descent rules), converges to a local optimum and, hence needs careful data initialization.

The majority of the complex PCA [15], [16] approaches proposed assume linear data dependencies. But, in many applications non-linear data dependencies naturally arise. The non-linear nature of various phenomena constituted the extension of real PCA to kernel PCA [20] one of the most popular methodologies for non-linear dimensionality reduction and feature extraction. To the best of our knowledge, even though nonlinear complex kernels have been recently studied for the design of digital filters and regression frameworks [21]–[23], very limited research has been conducted on non-linear component analysis using positive definite complex kernels and in the limited work conducted circularity was always assumed [24], [25]. In this paper we aim at advancing the state-of-the-art by formulating a general Widely Linear Complex Kernel PCA (WLCKPCA) methodology. In particular the contributions of the paper are

- we formulate the general framework of performing PCA using the widely linear model in Complex Reproducing Kernel Hilbert Spaces (CRKHS).
- we show how, by using this framework, WLPCA can be performed in Small Sample Size (SSS) problems (i.e., where the data dimensionality  $M$  is significantly greater than the number of training samples  $N$ ). SSS problems often arise in computer vision problems such as image reconstruction and recognition where number of image’s pixels is much greater than number of images, namely  $M \gg N$ .
- we show that by using the proposed WLPCA models and a newly proposed complex robust kernel state-of-the-art results can be achieved in face reconstruction.

The remainder of the paper is organized as follows. In Section II, we briefly describe the theory of Reproducing Complex Kernel Hilbert Spaces (RKHS). In Section III, we discuss circular complex PCA in RCKHS. In section IV, we formulate the proposed widely linear PCA model in RCKHS. In Section V, we show how in the case of a special kernel where the non-linear mapping to the feature space is known, robust reconstruction can be performed by the proposed widely linear PCA model. Experimental results are described in Section VI. Finally, conclusions are drawn in Section VII.

## II. COMPLEX RKHS

The theory of Reproducing Kernel Hilbert Spaces was generally defined using both real and complex kernels [26]. Even though, frameworks for Complex Kernel Least Mean Squares (CKLMS) [21], [27] and for Support Vector Machines (SVMs) [28] have been proposed, there is very limited work on component analysis and subspace learning with complex kernels and in the limited studies circularity of the data was assumed [24], [25], [29]. Before defining the proposed WLKPCA model we will make a brief introduction on complex RKHS focusing on the differences between the pure real space.

Let  $\mathbb{C}$  be the set of complex numbers. Let a function  $k : \mathbb{C}^M \times \mathbb{C}^M \rightarrow \mathbb{C}$  ( $\mathbb{C}^M$  is the so-called input space) with  $k(\mathbf{x}_i, \mathbf{x}_j) =$

$k^*(\mathbf{x}_j, \mathbf{x}_i)$  where  $\mathbf{x}_i, \mathbf{x}_j \in \mathbb{C}^M$  and  $*$  is the complex conjugation in  $\mathbb{C}$ . Let us assume a finite set  $\mathcal{T} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$  with  $\mathbf{x}_i \in \mathbb{C}^M$ . Let us also define the Hermitian matrix  $\mathbf{K} = [k(\mathbf{x}_i, \mathbf{x}_j)] \in \mathbb{C}^{N \times N}$ .  $\mathbf{K}$  is positive semi-definite (psd) iff for all  $\mathbf{c} \in \mathbb{C}^N$  it satisfies  $\mathbf{c}^H \mathbf{K} \mathbf{c} \geq 0$ , where  $H$  is the Hermitian transposition operation. The kernel  $k$  is psd iff for all sets  $\mathcal{T}$  the corresponding kernel matrix  $\mathbf{K}$  is psd.

Similar to the case of real kernels a complex kernel can define an arbitrary dimensional complex Hilbert space  $\mathcal{F}$  ( $\mathcal{F}$  is the so-called feature space and we assume that it is isomorphic with  $\mathbb{C}^F$  with  $F \gg M$ ) and a function  $\phi : \mathbb{C}^M \rightarrow \mathcal{F}^2$  such that  $\phi(\mathbf{x}) = k(\cdot, \mathbf{x})$  (reproducing property). The norm and inner product in  $\mathcal{F}$  is defined by means of  $k$  as  $\langle \phi(\mathbf{x}), \phi(\mathbf{y}) \rangle_{\mathcal{H}} = k(\mathbf{y}, \mathbf{x}) = \phi(\mathbf{y})^H \phi(\mathbf{x})$ . The inner product is sesquilinear (i.e., linear in one argument and antilinear in the other) and Hermitian the following hold (this is not the case for real kernels)

$$\begin{aligned} \langle a\phi(\mathbf{x}) + b\phi(\mathbf{z}), \phi(\mathbf{y}) \rangle_{\mathcal{H}} &= a\langle \phi(\mathbf{x}), \phi(\mathbf{y}) \rangle_{\mathcal{H}} \\ &\quad + b\langle \phi(\mathbf{z}), \phi(\mathbf{y}) \rangle_{\mathcal{H}} \\ \langle \phi(\mathbf{x}), a\phi(\mathbf{y}) + b\phi(\mathbf{z}) \rangle_{\mathcal{H}} &= a^* \langle \phi(\mathbf{x}), \phi(\mathbf{y}) \rangle_{\mathcal{H}} \\ &\quad + b^* \langle \phi(\mathbf{x}), \phi(\mathbf{z}) \rangle_{\mathcal{H}} \\ \langle \phi(\mathbf{x}), \phi(\mathbf{y}) \rangle_{\mathcal{H}}^* &= \langle \phi(\mathbf{y}), \phi(\mathbf{x}) \rangle_{\mathcal{H}}. \end{aligned} \quad (1)$$

Some complex kernel which properties have been studied are the Szego and Bregman kernels [30]. Another interesting kernel is complex Gaussian kernel  $k(\mathbf{z}, \mathbf{w}) = \exp -\frac{\sum_{i=1}^M (z_i - w_i^*)^2}{\sigma^2}$  [27], [31]. Methods for complexifying real kernels such as the linear kernel, the polynomial kernel, the Gaussian kernel [32], [33] have been proposed. Recently, complex kernels have been proposed in the literature that can efficiently compute the dot product between two vectors with arbitrary number of linear complex filter responses (i.e., in this case  $\phi(\mathbf{x})$  and  $\phi(\mathbf{y})$  is the vector with arbitrary number of linear responses). For this particular kernel there is a closed-form solution given by  $\phi(\mathbf{x}) = \mathbf{H}^{1/2} \mathbf{f}(\mathbf{x})$  where  $\mathbf{H}$  is a matrix which contains the sum of the power spectrum of the filters and  $\mathbf{f}(\mathbf{x})$  is the 2D Discrete Fourier Transform (DFT) of the vectorized image  $\mathbf{x}$  [34]. Another category of recently introduced complex kernels are the robust kernels proposed in [24], [25] and are defined as  $k(\mathbf{x}, \mathbf{y}) = \sum_k \exp(\alpha \pi j(x_i - y_j))$ . This particular kernel has a closed form feature space representation

$$\phi(\mathbf{x}) = \frac{1}{\sqrt{2}} \exp(j\alpha \pi \mathbf{x}) = \frac{1}{\sqrt{2}} \begin{bmatrix} e^{j\alpha \pi x_1} \\ \vdots \\ e^{j\alpha \pi x_M} \end{bmatrix} \quad (2)$$

where the elements  $0 \leq x_i \leq 1$  and for  $0 \leq a < 2$  the mapping  $\phi$  is invertible and the invert is given by  $\mathbf{x} = \frac{1}{\pi a} \angle \phi(\mathbf{x})$ .

In the following we will briefly describe how Circular Complex Kernel PCA (CCKPCA) is formulated. Then, we will formulate a widely linear model for PCA with complex kernels. Before let us define the kernels matrices that we use hereafter. First, let us define the centralized kernel matrix  $\tilde{\mathbf{K}}$  as

$$\begin{aligned} \tilde{\mathbf{K}} &= [\langle \phi(\mathbf{x}_i) - \mathbf{m}_\phi, \phi(\mathbf{x}_j) - \mathbf{m}_\phi \rangle_{\mathcal{H}}] \\ &= \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle_{\mathcal{H}} - \frac{1}{N} \sum_{k=1}^N \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_k) \rangle_{\mathcal{H}} - \\ &\quad - \frac{1}{N} \sum_{k=1}^N \langle \phi(\mathbf{x}_k), \phi(\mathbf{x}_j) \rangle_{\mathcal{H}} + \frac{1}{N^2} \sum_{k=1}^N \sum_{l=1}^N \langle \phi(\mathbf{x}_k), \phi(\mathbf{x}_l) \rangle_{\mathcal{H}} \\ &= (\mathbf{I} - \frac{1}{N} \mathbf{E}) \mathbf{K} (\mathbf{I} - \frac{1}{N} \mathbf{E}) \end{aligned} \quad (3)$$

where  $\mathbf{E} = [1]$  and  $\mathbf{m}_\phi = \frac{1}{N} \sum_i \phi(\mathbf{x}_i)$ . Furthermore, assuming that the data in  $\mathcal{T}$  are ordered in a matrix  $\mathbf{X}_\phi = [\phi(\mathbf{x}_1), \dots, \phi(\mathbf{x}_N)]$ , then  $\mathbf{K} = \mathbf{X}_\phi^H \mathbf{X}_\phi$ . Similarly, assuming the centralized matrix  $\tilde{\mathbf{X}}_\phi = [\phi(\mathbf{x}_1) - \mathbf{m}_\phi, \dots, \phi(\mathbf{x}_N) - \mathbf{m}_\phi] = \mathbf{X}_\phi (\mathbf{I} - \frac{1}{N} \mathbf{E})$  then  $\tilde{\mathbf{K}} = \tilde{\mathbf{X}}_\phi^H \tilde{\mathbf{X}}_\phi$ .

<sup>2</sup>function  $\phi$  can be explicit or implicit depending on the choice of the kernel

Finally, we will encounter the following vectors  $\mathbf{g}(\mathbf{t})$  and  $\tilde{\mathbf{g}}(\mathbf{t})$  as  $\mathbf{g}(\mathbf{t}) = \mathbf{X}_\phi^H \phi(\mathbf{t}) = [k(\mathbf{t}, \mathbf{x}_i)]$  and

$$\begin{aligned} [\tilde{\mathbf{g}}(\mathbf{t})]_j &= [\tilde{\mathbf{X}}_\phi^H (\phi(\mathbf{t}) - \mathbf{m}_\phi)]_j \\ &= k(\mathbf{t}, \mathbf{x}_j) - \frac{1}{N} \sum_{i=1}^N k(\mathbf{t}, \mathbf{x}_i) \\ &\quad - \frac{1}{N} \sum_{k=1}^N k(\mathbf{x}_k, \mathbf{x}_j) + \frac{1}{N^2} \sum_{k=1}^N \sum_{l=1}^N k(\mathbf{x}_k, \mathbf{x}_l). \end{aligned} \quad (4)$$

From the above equations, we can conclude that  $\mathbf{g}(\mathbf{t})$  and  $\tilde{\mathbf{g}}(\mathbf{t})$  can be computed using only kernel  $k$ .

### III. CIRCULAR PRINCIPAL COMPONENT ANALYSIS WITH COMPLEX KERNELS

Let us assume set  $\mathcal{T}$  and a complex positive definite kernel  $k$  which defines an explicit or implicit mapping  $\phi(\cdot)$  to an CRKHS  $\mathcal{F}$ . We can define the total scatter matrix in  $\mathcal{F}$

$$\mathbf{S}_\phi = \sum_{i=1}^N (\phi(\mathbf{x}_i) - \mathbf{m}_\phi)(\phi(\mathbf{x}_i) - \mathbf{m}_\phi)^H = \tilde{\mathbf{X}}_\phi \tilde{\mathbf{X}}_\phi^H. \quad (5)$$

In CCPCA a set of bases  $\mathbf{U}_\phi$  in  $\mathcal{F}$  should be found by solving the following optimization problem

$$\begin{aligned} \mathbf{U}_\phi &= \arg \max_{\mathbf{U}_\phi} \text{tr}[\mathbf{U}_\phi^H \mathbf{S}_\phi \mathbf{U}_\phi] \\ \text{subject to } \mathbf{U}_\phi^H \mathbf{U}_\phi &= \mathbf{I} \end{aligned} \quad (6)$$

Since, the dimensionality of  $\mathcal{F}$  is larger than  $N$ , subspace  $\mathbf{U}_\phi$  can be always written as a linear combination of  $\tilde{\mathbf{X}}_\phi$  as  $\mathbf{U}_\phi = \tilde{\mathbf{X}}_\phi \mathbf{V}$ , where  $\mathbf{V} \in \mathbb{C}^{M \times N}$ . Hence optimization problem (6) can be reformulated as

$$\begin{aligned} \mathbf{V}_o &= \arg \max_{\mathbf{V}} \text{tr}[\mathbf{V}^H \tilde{\mathbf{X}}_\phi^H \tilde{\mathbf{X}}_\phi \tilde{\mathbf{X}}_\phi^H \tilde{\mathbf{X}}_\phi \mathbf{V}] \\ \text{subject to } \mathbf{V}^H \tilde{\mathbf{K}} \mathbf{V} &= \mathbf{I}. \end{aligned} \quad (7)$$

or equivalently

$$\begin{aligned} \mathbf{V}_o &= \arg \max_{\mathbf{V}} \text{tr}[\mathbf{V}^H \tilde{\mathbf{K}}^2 \mathbf{V}] \\ \text{subject to } \mathbf{V}^H \tilde{\mathbf{K}} \mathbf{V} &= \mathbf{I}. \end{aligned} \quad (8)$$

Computing the derivatives with regards to  $\mathbf{V}$  and  $\mathbf{V}^*$  and the Lagrange multiplies of the constraint we get that the solution of the problem is given by performing eigenanalysis on the centralized kernel matrix  $\tilde{\mathbf{K}} = \tilde{\mathbf{X}}_\phi^H \tilde{\mathbf{X}}_\phi = \mathbf{M} \mathbf{A} \mathbf{M}^H$  and  $\mathbf{U}_\phi = \tilde{\mathbf{X}}_\phi \mathbf{M} \mathbf{A}^{-\frac{1}{2}}$ . Feature extraction can be easily computed using only  $k$

$$\begin{aligned} \mathbf{y} &= \mathbf{U}_\phi^H (\phi(\mathbf{t}) - \mathbf{m}_\phi) \\ &= \mathbf{A}^{-\frac{1}{2}} \mathbf{M}^H \tilde{\mathbf{X}}_\phi^H (\phi(\mathbf{t}) - \mathbf{m}_\phi) \\ &= \mathbf{A}^{-\frac{1}{2}} \mathbf{M}^H \tilde{\mathbf{g}}(\mathbf{t}) \end{aligned} \quad (9)$$

where  $\tilde{\mathbf{g}}(\mathbf{t})$  can be computed by (4).

### IV. WIDELY LINEAR PRINCIPAL COMPONENT ANALYSIS WITH COMPLEX KERNELS

The above complex PCA is optimal under a noise model of a circular Gaussian. That is, it uses only the information of the covariance  $\Sigma_\phi = \tilde{\mathbf{X}}_\phi \tilde{\mathbf{X}}_\phi^H$ , assuming that the pseudo-covariance  $\mathbf{C}_\phi = \tilde{\mathbf{X}}_\phi \tilde{\mathbf{X}}_\phi^T$  is zero [17] [35]. In the following, in order to exploit the information of the pseudo-covariance we extend the widely linear PCA proposed in [36] in arbitrary CRKHS. The proposed extension also provides a methodology for performing a widely linear PCA in SSS problems. First, we show how arbitrary psd complex kernels can be efficiently incorporated. This leads to an implementation of Widely Linear Complex PCA (WLCPCA) of  $O(N^3)$  complexity. The proposed methodology can be also applied in the case of the linear kernel (i.e.,  $k(\mathbf{x}, \mathbf{y}) = \mathbf{y}^H \mathbf{x}$ ) to perform a widely linear PCA in  $O(N^3)$  (instead of the algorithm of  $O(M^3)$  proposed in [36]).

We incorporate the pseudo-covariance information by augmenting the complex random signal with its complex conjugate in the

CRKHS. To do so, in order to ease the derivations let us define  $\mathbf{z}(\mathbf{x}_j)$  or simply  $\mathbf{z}_j$  as  $\mathbf{z}_j = \phi(\mathbf{x}_j) - \mathbf{m}_\phi$  in  $\mathcal{F}$  and define the augmented signal

$$\underline{\mathbf{z}}_j \triangleq \frac{1}{\sqrt{2}} \begin{bmatrix} \phi(\mathbf{x}_j) - \mathbf{m}_\phi \\ \phi^*(\mathbf{x}_j) - \mathbf{m}_\phi^* \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{z}_j \\ \mathbf{z}_j^* \end{bmatrix} \in \mathbb{C}^{2F} \quad (10)$$

Before proceeding to the construction of the algorithm, we need to define the following real and imaginary parts of the complex mapped vector  $\mathbf{z}_j$

$$\begin{aligned} \mathbf{z}_j^r &= \text{Real}[\mathbf{z}_j] \in \mathbb{R}^F \\ \mathbf{z}_j^c &= \text{Imag}[\mathbf{z}_j] \in \mathbb{R}^F \end{aligned} \quad (11)$$

and the real composite vector  $\tilde{\mathbf{z}}_j$  which is obtained by stacking  $\mathbf{z}_j^r$  on  $\mathbf{z}_j^c$

$$\tilde{\mathbf{z}}_j \triangleq \begin{bmatrix} \mathbf{z}_j^r \\ \mathbf{z}_j^c \end{bmatrix} \in \mathbb{R}^{2F} \quad (12)$$

We can build the vector  $\underline{\mathbf{z}}_j$  using matrices  $\mathbf{z}_j^r$  and  $\mathbf{z}_j^c$  via the use of a  $2F \times 2F$  unitary matrix

$$\mathbf{T}_F = \frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{I}_F & j\mathbf{I}_F \\ \mathbf{I}_F & -j\mathbf{I}_F \end{bmatrix} \quad (13)$$

where  $\mathbf{I}_F$  is an  $F \times F$  identity matrix. It is easy to verify that  $\mathbf{T}_F^H \mathbf{T}_F = \mathbf{T}_F \mathbf{T}_F^H = \mathbf{I}$ . Having defined  $\mathbf{T}_F$  we can build

$$\underline{\mathbf{z}}_j = \frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{z}_j \\ \mathbf{z}_j^* \end{bmatrix} = \mathbf{T}_F \begin{bmatrix} \mathbf{z}_j^r \\ \mathbf{z}_j^c \end{bmatrix} = \mathbf{T}_F \tilde{\mathbf{z}}_j \quad (14)$$

The main matrices that will be used in the analysis are summarised in the below table.

Summary of used symbols	
$\underline{\Sigma}_{\tilde{\mathbf{z}}\tilde{\mathbf{z}}}$	covariance matrix of the real composite vector $\mathbf{X}_{\tilde{\mathbf{z}}} \mathbf{X}_{\tilde{\mathbf{z}}}^T \stackrel{EVD}{=} \mathbf{W}_\phi \mathbf{\Lambda} \mathbf{W}_\phi^T$
$\underline{\Sigma}_{\underline{\mathbf{z}}\underline{\mathbf{z}}}$	augmented covariance matrix $\mathbf{X}_{\underline{\mathbf{z}}} \mathbf{X}_{\underline{\mathbf{z}}}^H \stackrel{EVD}{=} \mathbf{U}_\phi \mathbf{\Lambda} \mathbf{U}_\phi^H \stackrel{AEVD}{=} \mathbf{U}_\phi \mathbf{L} \mathbf{U}_\phi^H$
$\tilde{\mathbf{K}}^R$	centralized kernel of the real composite vector $\mathbf{X}_{\tilde{\mathbf{z}}}^T \mathbf{X}_{\tilde{\mathbf{z}}} = \text{Real}[\tilde{\mathbf{K}}] \stackrel{EVD}{=} \mathbf{\Xi} \mathbf{\Lambda} \mathbf{\Xi}^T$
$\mathbf{\Xi}$	eigenvectors of $\tilde{\mathbf{K}}^R$
$\mathbf{\Lambda}$	eigenvalues of $\tilde{\mathbf{K}}^R$
$\mathbf{W}_\phi$	eigenvectors of $\underline{\Sigma}_{\tilde{\mathbf{z}}\tilde{\mathbf{z}}}$ , $\mathbf{W}_\phi = \mathbf{X}_{\tilde{\mathbf{z}}} \mathbf{\Xi} \mathbf{\Lambda}^{-\frac{1}{2}}$
$\mathbf{U}_\phi$	eigenvectors of $\underline{\Sigma}_{\underline{\mathbf{z}}\underline{\mathbf{z}}}$ , $\mathbf{U}_\phi = \mathbf{T}_F \mathbf{W}_\phi$
$\underline{\mathbf{U}}_\phi$	augmented eigenvectors of $\underline{\Sigma}_{\underline{\mathbf{z}}\underline{\mathbf{z}}}$ , $\underline{\mathbf{U}}_\phi = \mathbf{U}_\phi \mathbf{T}_F^H$

Let us define the covariance matrix of the real composite vector  $\tilde{\mathbf{z}}_j$

$$\underline{\Sigma}_{\tilde{\mathbf{z}}\tilde{\mathbf{z}}} = \frac{1}{N} \sum_{j=1}^n \tilde{\mathbf{z}}_j \tilde{\mathbf{z}}_j^T = \mathbf{X}_{\tilde{\mathbf{z}}} \mathbf{X}_{\tilde{\mathbf{z}}}^T = \begin{bmatrix} \Sigma_{rr} & \Sigma_{rc} \\ \Sigma_{rc}^T & \Sigma_{cc} \end{bmatrix} \quad (15)$$

where  $\Sigma_{rr} = \mathbf{X}_{z^r} \mathbf{X}_{z^r}^T$ ,  $\Sigma_{cc} = \mathbf{X}_{z^c} \mathbf{X}_{z^c}^T$ ,  $\Sigma_{rc} = \mathbf{X}_{z^r} \mathbf{X}_{z^c}^T$ ,  $\mathbf{X}_{\tilde{\mathbf{z}}} = \sqrt{\frac{1}{n}} [\mathbf{z}_1 \dots \mathbf{z}_N]$ ,  $\mathbf{X}_{z^r} = \sqrt{\frac{1}{N}} [\mathbf{z}_1^r \dots \mathbf{z}_N^r]$  and  $\mathbf{X}_{z^c} = \sqrt{\frac{1}{N}} [\mathbf{z}_1^c \dots \mathbf{z}_N^c]$ . Using the unitary transform in (13) the augmented covariance matrix of the complex augmented random vectors  $\underline{\mathbf{z}}_j$  can be defined as

$$\begin{aligned} \underline{\Sigma}_{\underline{\mathbf{z}}\underline{\mathbf{z}}} &= \frac{1}{N} \sum_{j=1}^N \underline{\mathbf{z}}_j \underline{\mathbf{z}}_j^H = \mathbf{X}_{\underline{\mathbf{z}}} \mathbf{X}_{\underline{\mathbf{z}}}^H \\ &= \begin{bmatrix} \Sigma_{\underline{\mathbf{z}}\underline{\mathbf{z}}} & \mathbf{C}_{\underline{\mathbf{z}}\underline{\mathbf{z}}} \\ \mathbf{C}_{\underline{\mathbf{z}}\underline{\mathbf{z}}}^* & \Sigma_{\underline{\mathbf{z}}\underline{\mathbf{z}}}^* \end{bmatrix} \\ &= \mathbf{T}_F \underline{\Sigma}_{\tilde{\mathbf{z}}\tilde{\mathbf{z}}} \mathbf{T}_F^H \end{aligned} \quad (16)$$

where  $\Sigma_{\underline{\mathbf{z}}\underline{\mathbf{z}}} = \frac{1}{N} \sum_{j=1}^N \mathbf{z}_j \mathbf{z}_j^H$  is the covariance matrix and  $\mathbf{C}_{\underline{\mathbf{z}}\underline{\mathbf{z}}} = \frac{1}{N} \sum_{j=1}^N \mathbf{z}_j \mathbf{z}_j^T$  is the complementary covariance matrix or relation

[18] or pseudo-covariance matrix [14]. The corresponding problem is to find a projection matrix  $\mathbf{U}_\phi$  such that

$$\begin{aligned} \mathbf{U}_\phi &= \arg \max_{\mathbf{U}} \text{tr}[\mathbf{U}_\phi^H \underline{\Sigma}_{\underline{\mathbf{z}}\underline{\mathbf{z}}} \mathbf{U}_\phi] \\ \text{subject to } &\mathbf{U}_\phi^H \mathbf{U}_\phi = \mathbf{I}. \end{aligned} \quad (17)$$

In order to solve the above optimization problem we need to perform eigenvalue decomposition (EVD) to the augmented covariance matrix  $\underline{\Sigma}_{\underline{\mathbf{z}}\underline{\mathbf{z}}}$ . This matrix can be written using EVD as follows:

$$\underline{\Sigma}_{\underline{\mathbf{z}}\underline{\mathbf{z}}} = \mathbf{U}_\phi \mathbf{\Lambda} \mathbf{U}_\phi^H \quad (18)$$

where  $\mathbf{U}_\phi^H \mathbf{U}_\phi = \mathbf{I}$

In the following we will show how to find the positive eigenvalues of this matrix and represent  $\mathbf{U}_\phi$  as a linear combination of the augmented vectors  $\underline{\mathbf{z}}$  as  $\mathbf{U}_\phi = \mathbf{X}_{\underline{\mathbf{z}}} \mathbf{V}$ .

In order to compute  $\mathbf{V}$  we exploit the relationship between  $\underline{\Sigma}_{\underline{\mathbf{z}}\underline{\mathbf{z}}}$  and  $\underline{\Sigma}_{\tilde{\mathbf{z}}\tilde{\mathbf{z}}}$  from (16). First, we exploit the relationship between the eigenvalues of  $\mathbf{X}_{\tilde{\mathbf{z}}} \mathbf{X}_{\tilde{\mathbf{z}}}^T$  and  $\mathbf{X}_{\tilde{\mathbf{z}}}^T \mathbf{X}_{\tilde{\mathbf{z}}}$ . That is, based on linear algebra we can show that these matrices have the same positive eigenvalues [24] and the corresponding eigenvectors are related by  $\mathbf{W}_\phi = \mathbf{X}_{\tilde{\mathbf{z}}} \mathbf{\Xi} \mathbf{\Lambda}^{-\frac{1}{2}}$  where  $\mathbf{\Xi}$  and  $\mathbf{\Lambda}$  are the eigenvectors and eigenvalues of

$$\begin{aligned} \tilde{\mathbf{K}}^R &= \mathbf{X}_{\tilde{\mathbf{z}}}^T \mathbf{X}_{\tilde{\mathbf{z}}} = \mathbf{X}_{z^r}^T \mathbf{X}_{z^r} + \mathbf{X}_{z^c}^T \mathbf{X}_{z^c} \\ &= \text{Real}[\tilde{\mathbf{K}}] = \mathbf{\Xi} \mathbf{\Lambda} \mathbf{\Xi}^T \end{aligned} \quad (19)$$

where  $\tilde{\mathbf{K}}$  is the centralized kernel matrix in (3) and  $\mathbf{W}_\phi$  are the eigenvectors of  $\underline{\Sigma}_{\tilde{\mathbf{z}}\tilde{\mathbf{z}}}$ .

The eigenvalues and eigenvectors of the augmented matrix can be shown that are related with the respective eigenvalues and eigenvectors of the matrix  $\underline{\Sigma}_{\tilde{\mathbf{z}}\tilde{\mathbf{z}}}$  as follow

$$\begin{aligned} \underline{\Sigma}_{\underline{\mathbf{z}}\underline{\mathbf{z}}} &= \mathbf{X}_{\underline{\mathbf{z}}} \mathbf{X}_{\underline{\mathbf{z}}}^H \\ &= \mathbf{T}_F \mathbf{X}_{\tilde{\mathbf{z}}} \mathbf{X}_{\tilde{\mathbf{z}}}^T \mathbf{T}_F^H \\ &= \mathbf{T}_F \mathbf{W}_\phi \mathbf{\Lambda} \mathbf{W}_\phi^T \mathbf{T}_F^H \\ &= \mathbf{U}_\phi \mathbf{\Lambda} \mathbf{U}_\phi^H \end{aligned} \quad (20)$$

where  $\mathbf{U}_\phi^H \mathbf{U}_\phi = \mathbf{I}$ .

The above Eq. 20 summarizes the relationship of the spectrum of  $\underline{\Sigma}_{\tilde{\mathbf{z}}\tilde{\mathbf{z}}}$  in the real domain with  $\underline{\Sigma}_{\underline{\mathbf{z}}\underline{\mathbf{z}}}$  in complex domain. Furthermore, it was made apparent that in order to perform eigenanalysis to these two matrices we have just to calculate eigenvalues and eigenvectors of kernel matrix  $\tilde{\mathbf{K}}^R$ .

Since we want to derive a Widely Linear model as in [15], we need to arrange eigenvalues (and accordingly eigenvectors) in an augmented manner, using an Augmented Eigenvalue Decomposition (AEVD) which not only provides a more condensed representation but also improves the result [36].

Now we will attempt to relate AEVD with our previous computations of  $\underline{\Sigma}_{\underline{\mathbf{z}}\underline{\mathbf{z}}}$  and  $\underline{\Sigma}_{\tilde{\mathbf{z}}\tilde{\mathbf{z}}}$

$$\begin{aligned} \underline{\Sigma}_{\underline{\mathbf{z}}\underline{\mathbf{z}}} &= \mathbf{T}_F \underline{\Sigma}_{\tilde{\mathbf{z}}\tilde{\mathbf{z}}} \mathbf{T}_F^H \\ &= \mathbf{T}_F \mathbf{W}_\phi \mathbf{\Lambda} \mathbf{W}_\phi^T \mathbf{T}_F^H \\ &= \mathbf{T}_F \mathbf{W}_\phi \mathbf{T}_F^H \mathbf{T}_F^H \mathbf{\Lambda} \mathbf{T}_F^H \mathbf{T}_F^H \mathbf{W}_\phi^T \mathbf{T}_F^H \\ &= \underline{\mathbf{U}}_\phi \mathbf{L} \underline{\mathbf{U}}_\phi^H \end{aligned} \quad (21)$$

where

$$\underline{\mathbf{U}}_\phi = \mathbf{T}_F \mathbf{W}_\phi \mathbf{T}_F^H = \mathbf{U}_\phi \mathbf{T}_F^H \quad (22)$$

and

$$\mathbf{L} = \mathbf{T}_F^H \mathbf{\Lambda} \mathbf{T}_F = \begin{bmatrix} \mathbf{\Lambda}_1 + \mathbf{\Lambda}_2 & \mathbf{\Lambda}_1 - \mathbf{\Lambda}_2 \\ \mathbf{\Lambda}_1 - \mathbf{\Lambda}_2 & \mathbf{\Lambda}_1 + \mathbf{\Lambda}_2 \end{bmatrix} \quad (23)$$

where  $\mathbf{\Lambda}_1, \mathbf{\Lambda}_2$  contain the eigenvalues of  $\underline{\Sigma}_{\tilde{\mathbf{z}}\tilde{\mathbf{z}}}$  ordered as

$$\mathbf{\Lambda}_1 = \text{Diag}(\lambda_1, \lambda_3, \dots, \lambda_{2k-1}) \quad (24)$$

$$\mathbf{\Lambda}_2 = \text{Diag}(\lambda_2, \lambda_4, \dots, \lambda_{2k}) \quad (25)$$

where  $k = N/2$ ,

Having an expression for  $\mathbf{U}_\phi$  and choosing  $k \leq N/2$  we can perform feature extraction from a test sample as

$$\begin{aligned} y^{WL} &= \mathbf{U}_\phi^H \mathbf{z}(\mathbf{t}) \\ &= \mathbf{T}_k^H \mathbf{W}_\phi^T \mathbf{T}_F^H \mathbf{z} \\ &= \mathbf{T}_k^H \Lambda_{2k}^{-\frac{1}{2}} \Xi_{2k}^T \mathbf{X}_z^T \mathbf{T}_F^H \mathbf{z} \\ &= \mathbf{T}_k^H \Lambda_{2k}^{-\frac{1}{2}} \Xi_{2k}^T \mathbf{X}_z^H \mathbf{z} \\ &= \mathbf{T}_k^H \Lambda_{2k}^{-\frac{1}{2}} \Xi_{2k}^T \text{Real}[\tilde{\mathbf{g}}(\mathbf{t})] \end{aligned} \quad (26)$$

where  $\tilde{\mathbf{g}}(\mathbf{t})$  is computed by (4).

## V. RECONSTRUCTION WITH WIDELY LINEAR KERNEL PCA

In the previous section we showed how dimensionality reduction can be computed using widely linear kernel PCA for arbitrary psd complex kernels. One of the main uses of PCA is to perform robust reconstruction [24], [37]–[39]. Unfortunately, for arbitrary complex kernels reconstruction would require inversion of, the unknown and not computable function, function  $\phi$  which is, in generally, infeasible. One way to deal with this is to resort to the problem of pre-image computation (requires the solution of separate optimization problem) [24], [40], [41]. Recently, a robust kernel, defined in (2), has been proposed based on the Euler representation of data for which a simple and analytic pre-image computation exists [24]. In the following, we show how the above framework can be applied to perform robust image reconstruction in the case of the Euler kernel (2).

To reconstruct a test vector  $\mathbf{t}$  using the augmented components  $\mathbf{U}_\phi$  of the widely linear representation we need to exploit the properties of the widely linear transform [36]. The augmented component matrix  $\mathbf{U}_\phi = \mathbf{T}_p \mathbf{X}_z \Xi_p \Lambda_p^{-\frac{1}{2}} \mathbf{T}_2^H$  has the following block pattern

$$\mathbf{U}_\phi = \begin{bmatrix} \mathbf{U}_1 & \mathbf{U}_2 \\ \mathbf{U}_2^* & \mathbf{U}_1^* \end{bmatrix} \quad (27)$$

The reconstruction of the augmented test image  $\mathbf{t}$  can be performed as

$$\begin{aligned} \hat{\mathbf{z}}(\mathbf{t}) &= \mathbf{U}_\phi \mathbf{U}_\phi^H \mathbf{z}(\mathbf{t}) \\ &= \begin{bmatrix} \mathbf{U}_1 & \mathbf{U}_2 \\ \mathbf{U}_2^* & \mathbf{U}_1^* \end{bmatrix} \begin{bmatrix} \mathbf{U}_1^H & \mathbf{U}_2^T \\ \mathbf{U}_2^H & \mathbf{U}_1^T \end{bmatrix} \mathbf{z} \\ &= \begin{bmatrix} \mathbf{U}_1 \mathbf{U}_1^H + \mathbf{U}_2 \mathbf{U}_2^H & \mathbf{U}_1 \mathbf{U}_2^T + \mathbf{U}_2 \mathbf{U}_1^T \\ \mathbf{U}_2^* \mathbf{U}_1^H + \mathbf{U}_1^* \mathbf{U}_2^H & \mathbf{U}_2^* \mathbf{U}_2^T + \mathbf{U}_1^* \mathbf{U}_1^T \end{bmatrix} \mathbf{z}(\mathbf{t}) \end{aligned} \quad (28)$$

From above we can conclude that

$$\hat{\mathbf{z}}(\mathbf{t}) = (\mathbf{U}_1 \mathbf{U}_1^H + \mathbf{U}_2 \mathbf{U}_2^H) \mathbf{z}(\mathbf{t}) + (\mathbf{U}_1 \mathbf{U}_2^T + \mathbf{U}_2 \mathbf{U}_1^T) \mathbf{z}^*(\mathbf{t}) \quad (29)$$

and the reconstructed image  $\tilde{\mathbf{t}}$  can be given by  $\frac{1}{\alpha\pi} \angle \hat{\mathbf{z}}(\mathbf{t})$ . An interesting property of widely linear PCA approaches is that they provide features up to  $k < N/2$  (while circular and other PCA variants up to  $N$ ) [36]. In that way, since it allows a kind of multiplexes of the components, produces more condensed representations. The complete algorithm for learning the subspace and reconstructing using the widely linear kernel PCA for the kernel in (2) is given in Algorithm 1.

## Algorithm 1 WIDELY LINEAR KERNEL PCA WITH EULER KERNEL

**Input:** A set of  $n$  images  $I_j, j = 1, \dots$ , of  $M$  pixels, the number  $m$  of principal components and parameter  $\alpha$ .

**Output:** The principal subspace  $\mathbf{U}_\phi$  and eigenvalues  $\Sigma$ .

- 1) Represent  $I_j$  in the range  $[0, 1]$  and obtain  $\mathbf{x}_j$  by writing  $I_j$  in lexicographic order.
- 2) Compute  $\mathbf{z}_j$ ; using (2) and (10) form the augmented matrix of the transformed data  $\mathbf{X}_z \in \mathbb{C}^{2M \times N}$
- 3) Compute the kernel matrix  $\mathbf{K} = [k(\mathbf{x}_i, \mathbf{x}_j)]$  and find the eigendecomposition of  $\mathbf{K}^R = \text{Real}[\mathbf{K}] = \Xi \Lambda \Xi^H$ .
- 4) Find the  $m$ -reduced set,  $\Xi_m \in \mathbb{C}^{m \times N}$  and  $\Lambda_m \in \mathbb{R}^{m \times m}$ .
- 5) Compute  $\mathbf{U}_\phi = \mathbf{X}_z \Xi_m \Lambda_m^{-\frac{1}{2}} \mathbf{T}_2^H = \begin{bmatrix} \mathbf{U}_1 & \mathbf{U}_2 \\ \mathbf{U}_2^* & \mathbf{U}_1^* \end{bmatrix}$ .
- 6) Reconstruct a test sample  $\mathbf{t}$  first by using the widely linear model (29).
- 7) Fast pre-image computation: go back to the pixel domain using  $\tilde{\mathbf{t}} = \frac{\angle \hat{\mathbf{z}}(\mathbf{t})}{\alpha\pi}$

## VI. EXPERIMENTAL RESULTS

One of the main applications of PCA is to perform robust reconstruction. In this section we evaluate and assess the robustness of the proposed widely linear PCA using the Euler kernel for the application on image de-noising based on subspace-based image reconstruction.

For comparison, we select a number of standard and state-of-the-art approaches such as PCA, R1-PCA [37], PCA-L1 [38], HQ-PCA [39] and Euler PCA [24] (the last approach is the circular alternative of the proposed widely linear model). We use the same parameters and the same convergence for R1-PCA, PCA-L1, HQ-PCA as in [24].

Our data consists of a subset of the popular AR database [42]. In particular, we use a total of 100 images of size  $101 \times 91$  of different subjects as shown in Fig. 1. Furthermore, the XM2VTS database [43] is used. We use a total of 295 images of size  $101 \times 91$  of different subjects as shown in Fig. 2.



Fig. 1. Cropped example images from AR database.



Fig. 2. Cropped example images from XM2VTS database.

For corruption we used skin-like occlusions. We created two different sets. In the first one we corrupted randomly 10% of images and in the second on 50%. As in [24], we occlude a subset of the training data with hand signs of the American fingerspelling alphabet. Examples of the corrupted images can be seen in the Fig. 3. The chosen letter, its orientation and its position are randomized, and the skin color is adjusted to fit the subject.

Our evaluation is based on the the angular error [37]. Angular error between the corrupted subspace  $\mathbf{B}^{cor}$  and the uncorrupted subspace  $\mathbf{B}^{orig} = [\mathbf{b}_1^{orig} \dots \mathbf{b}_m^{orig}] \in \mathbb{R}^{p \times m}$  is used as follows

$$e_a(m) = m - \sum_{l=1}^m \sum_{s=1}^m \cos^2(\mathbf{b}_l^{orig}, \mathbf{b}_s^{cor}). \quad (30)$$



Fig. 3. Corrupted example images from AR database.

In the following, expect for the proposed widely linear PCA, we used up to  $k = 50$  components. Since, in our methodology components are multiplexed we allowed up to 25 components (which is exactly equivalent with using 50 in other methods). In all figures the error rate is plotted versus the number of components (up to 50). For visualization purposes we show the performance of our 25 components in the same scale (i.e., in the corresponding graphs 10 components means 5, 20 means 10, 50 correspond to 25 and so on).

Fig. 4 shows the angular error for the AR database. HQ-PCA and PCA-L1 perform the worst, especially when we have 10% of occluded images. For the first 30 components, HQ-PCA is worse than PCA-L1. For more than 30 components, HQ-PCA outperforms PCA-L1. The other four methods performs similarly well. It can be said that Euler PCA and Widely Linear PCA perform slightly better than R1-PCA and standard PCA. It seems that Widely Linear PCA's performance is almost the same with Euler PCA's performance, where in some components Widely Linear PCA is better (e.g. for 25 and 35 components in the case of 10% of occluded images) while in some other components Euler PCA outperforms Widely Linear PCA (e.g. for 15 and 45 components in the case of 10% of occluded images). We conclude to the same results as previous if we see the angular error for the XM2VTS database in Figure 5. Again, HQ-PCA and PCA-L1 perform the worst while the other methods achieve a similar performance.

Finally, visual reconstructions for some images and for all the tested methods and for the same number of components are shown in Fig. 6. As can be seen, the proposed widely linear PCA produces visually the best reconstruction.

## VII. CONCLUSION

In this brief we showed, theoretically, how to perform widely linear PCA with complex kernels. A byproduct of our analysis is also the solution of widely linear PCA in small sample size (SSS) problems (ie., where the data dimensionality is larger than the available samples. We applied the proposed analysis to a recently proposed robust complex kernel. Empirical results showed the proposed widely linear kernel PCA framework not only outperforms its circular counterpart but many state-of-the-art robust PCA methodologies. Future work on the topic includes the extension of pre-image based reconstruction techniques for arbitrary complex kernels.

## REFERENCES

- [1] K. Pearson, "On lines and planes of closest fit to systems of points in space," *Philosophical Magazine*, vol. 2, pp. 559–572, 1901.
- [2] T. Jolliffe, "Principal component analysis," *Berlin: Springer*, pp. 559–572, 2002.
- [3] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of cognitive neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.
- [4] E. Koktopoulou, J. Chen, and Y. Saad, "Trace optimization and eigenproblems in dimension reduction methods," *Numerical Linear Algebra with Applications*, vol. 18, no. 3, pp. 565–602, 2011.
- [5] X. He, S. Yan, Y. Hu, P. Niyogi, and H.J. Zhang, "Face recognition using laplacianfaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 3, pp. 328–340, 2005.
- [6] H. Li, T. Jiang, and K. Zhang, "Efficient and robust feature extraction by maximum margin criterion," *IEEE Transactions on Neural Networks*, vol. 17, no. 1, pp. 157–165, 2006.
- [7] P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997.
- [8] Fernando De la Torre, "A least-squares framework for component analysis," *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol. 34, no. 6, pp. 1041–1055, 2012.
- [9] A. Papoulis, *Probability & statistics*, vol. 2, Prentice-Hall, 1990.
- [10] R. Savitha, S. Suresh, and N. Sundararajan, "Projection-based fast learning fully complex-valued relaxation neural network," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 24, no. 4, pp. 529–541, 2013.
- [11] A. Hirose and S. Yoshida, "Generalization characteristics of complex-valued feedforward neural networks in relation to signal coherence," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 4, pp. 541–551, 2012.
- [12] Jin Hu and Jun Wang, "Global stability of complex-valued recurrent neural networks with time-delays," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 6, pp. 853–865, 2012.
- [13] Jun Li, Jian Yang, and Yongfeng Diao, "Continuous attractors of recurrent neural networks with complex-valued weights," in *Neural Networks (IJCNN), The 2012 International Joint Conference on*, 2012, pp. 1–8.
- [14] F.D. Neeser and J.L. Massey, "Proper complex random processes with applications to information theory," *IEEE Transactions on Information Theory*, vol. 39, no. 4, pp. 1293–1302, 1993.
- [15] P.J. Schreier and L.L. Scharf, "Second-order analysis of improper complex random vectors and processes," *IEEE Transactions on Signal Processing*, vol. 51, no. 3, pp. 714–725, 2003.
- [16] T. Adali, P.J. Schreier, and L.L. Scharf, "Complex-valued signal processing: The proper way to deal with impropriety," *IEEE Transactions on Signal Processing*, vol. 59, no. 11, pp. 5101–5125, 2011.
- [17] B. Picinbono, "On circularity," *IEEE Transactions on Signal Processing*, vol. 42, no. 12, pp. 3473–3482, 1994.
- [18] B. Picinbono and P. Bondon, "Second-order statistics of complex signals," *IEEE Transactions on Signal Processing*, vol. 45, no. 2, pp. 411–420, 1997.
- [19] Xi-Lin Li, T. Adali, and M. Anderson, "Noncircular principal component analysis and its application to model selection," *IEEE Transactions on Signal Processing*, vol. 59, no. 10, pp. 4516–4528, 2011.
- [20] B. Schölkopf, A. Smola, and K.R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural computation*, vol. 10, no. 5, pp. 1299–1319, 1998.
- [21] P. Bouboulis, K. Slavakis, and S. Theodoridis, "Adaptive learning in complex reproducing kernel hilbert spaces employing wirtinger's subgradients," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 3, pp. 425–438, 2012.
- [22] P. Bouboulis, S. Theodoridis, and M. Mavrouforakis, "The augmented complex kernel lms," *IEEE Transactions on Signal Processing*, vol. 60, no. 9, pp. 4962–4967, 2012.
- [23] K. Slavakis, P. Bouboulis, and S. Theodoridis, "Adaptive multiregression in reproducing kernel hilbert spaces: the multiaccess mimo channel case," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 2, pp. 260–276, 2012.
- [24] S. Liwicki, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic, "Euler principal component analysis," *International Journal of Computer Vision*, vol. 101, no. 3, pp. 498–518, 2013.
- [25] G. Tzimiropoulos, S. Zafeiriou, and M. Pantic, "Subspace learning from image gradient orientations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 12, pp. 2454–2466, 2012.
- [26] N. Aronszajn, "Theory of reproducing kernels," *Trans. Amer. Math. Soc.*, vol. 68, no. 3, pp. 337–404, 1950.
- [27] P. Bouboulis and S. Theodoridis, "Extension of wirtinger's calculus to reproducing kernel hilbert spaces and the complex kernel lms," *IEEE Transactions on Signal Processing*, vol. 59, no. 3, pp. 964–978, mar 2011.
- [28] E.J. Bayro-Corrochano and N. Arana-Daniel, "Clifford support vector machines for classification, regression, and recurrence," *IEEE Transactions on Neural Networks*, vol. 21, no. 11, pp. 1731–1746, 2010.
- [29] Z. Wang, Q. Han, and X. Niu, "Complex kernel pca for multimodal biometric recognition," *IEICE Electronics Express*, vol. 6, no. 16, pp. 1131–1136, 2009.
- [30] Vern I. Paulsen, "An introduction to the theory of reproducing kernel hilbert spaces draft 2/20/06," 2006.
- [31] Ingo Steinwart, Don Hush, and Clint Scovel, "An explicit description of the reproducing kernel hilbert spaces of gaussian rbf kernels," *IEEE*

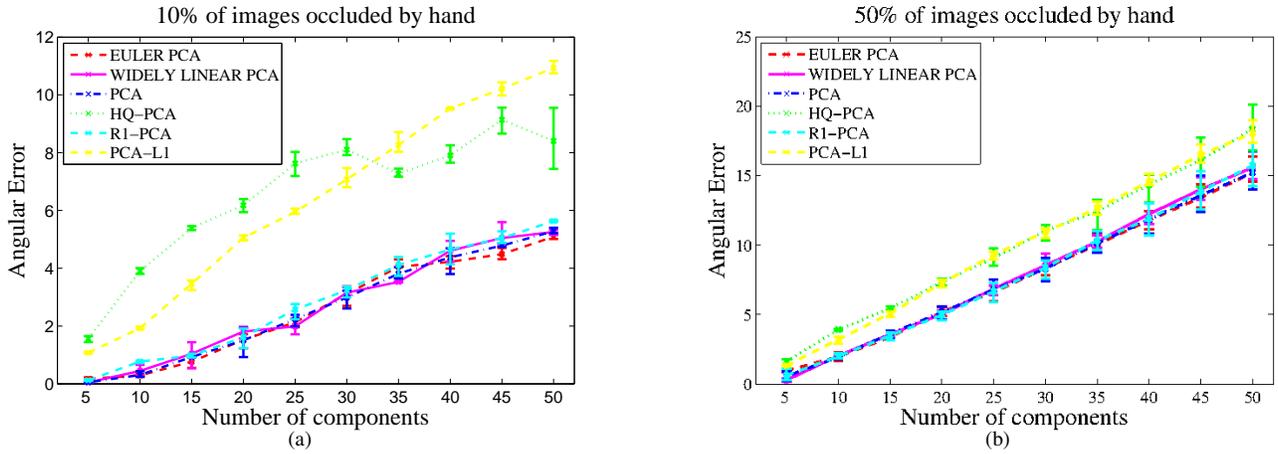


Fig. 4. Angular error with different rates of occluded images of AR database. In subfigure (a), 10% of images are selected randomly and corrupted by adding the skin occlusion, where in subfigure (b) half of the images are corrupted. Here, the mean value over 3 executions with different random patches is shown. Variance is indicated by error bars.

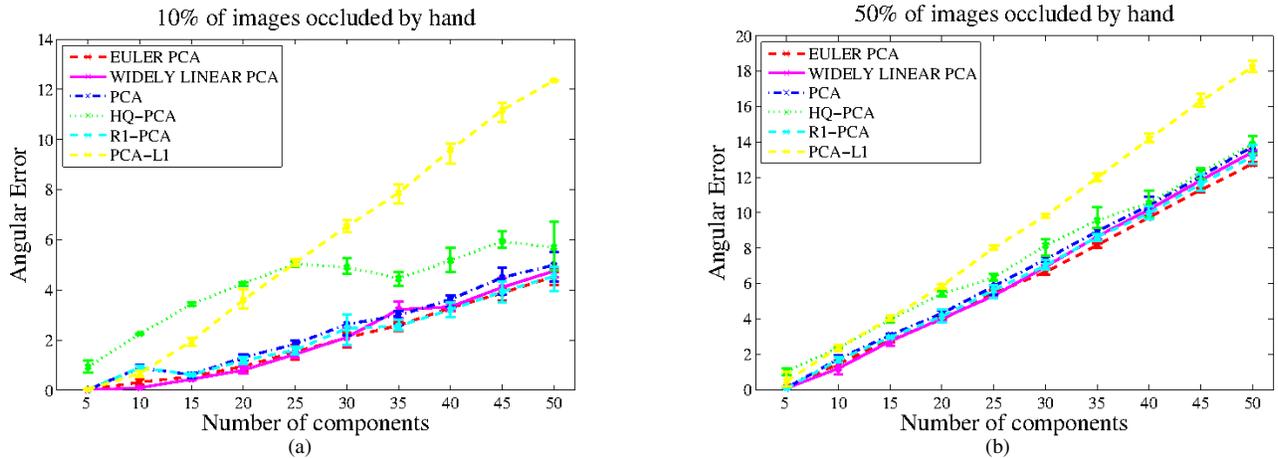


Fig. 5. Angular error with different rates of occluded images of XM2VTS database. In subfigure (a), 10% of images are selected randomly and corrupted by adding the skin occlusion, where in subfigure (b) half of the images are corrupted. Here, the mean value over 3 executions with different random patches is shown. Variance is indicated by error bars.

*Transactions on Information Theory*, vol. 52, no. 10, pp. 4635–4643, 2006.

[32] B. Schölkopf and A.J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*, Adaptive Computation and Machine Learning. Mit Press, 2002.

[33] Nello Cristianini and John Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*, Cambridge University Press, 2010.

[34] S. Zafeiriou, G. Tzimiropoulos, and M. Pantic, “Subspace analysis of arbitrarily many linear filter responses with an application to face tracking,” in *Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2011 IEEE Computer Society Conference on. IEEE, 2011, pp. 37–42.

[35] B. Picinbono and P. Chevalier, “Widely linear estimation with complex data,” *IEEE Transactions on Signal Processing*, vol. 43, no. 8, pp. 2030–2033, 1995.

[36] P.J. Schreier and L.L. Scharf, *Statistical Signal Processing of Complex-Valued Data: The Theory of Improper and Noncircular Signals*, Cambridge University Press, Cambridge, UK, 2010.

[37] Chris Ding, Ding Zhou, Xiaofeng He, and Hongyuan Zha, “R1-pca: Rotational invariant l1-norm principal component analysis for robust subspace factorization,” *ACM*, pp. 281–288, 2006.

[38] Nojun Kwak, “Principal component analysis based on l1-norm maximization,” *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol. 30, no. 9, pp. 1672–1680, 2008.

[39] Ran He, Bao-Gang Hu, Wei-Shi Zheng, and Xiang-Wei Kong, “Robust principal component analysis based on maximum correntropy criterion,” *IEEE Transactions on Image Processing*, vol. 20, no. 6, pp. 1485–1494, 2011.

[40] J.T.Y. Kwok and I.W.H. Tsang, “The pre-image problem in kernel methods,” *IEEE Transactions on Neural Networks*, vol. 15, no. 6, pp. 1517–1525, 2004.

[41] S. Zafeiriou and M. Petrou, “Nonlinear non-negative component analysis algorithms,” *IEEE Transactions on Image Processing*, vol. 19, no. 4, pp. 1050–1066, 2010.

[42] A.M. Martinez and R. Benavente, “The AR face database,” 1998, CVC Technical Report, 1998.

[43] K. Messer, J. Matas, J. Kittler, and K. Jonsson, “Xm2vtsdb: The extended m2vts database,” in *Second International Conference on Audio and Video-based Biometric Person Authentication, Washington D.C.*, pp. 72–77, 1999.



Fig. 6. Reconstruction of PCA, PCA-L1, R1-PCA, HQ-PCA, EULER-PCA and Widely Linear PCA after learning with 50% hand occluded images in AR database.