

Subspace Learning in Krein Spaces: Complete Kernel Fisher Discriminant Analysis with Indefinite Kernels

Stefanos Zafeiriou*

Department of Computing, Imperial College London
London SW7 2AZ, U.K.
s.zafeiriou@imperial.ac.uk
<http://ibug.doc.ic.ac.uk/people/szafeiriou>

Abstract. Positive definite kernels, such as Gaussian Radial Basis Functions (GRBF), have been widely used in computer vision for designing feature extraction and classification algorithms. In many cases non-positive definite (npd) kernels and non metric similarity/dissimilarity measures naturally arise (e.g., Hausdorff distance, Kullback Leibler Divergences and Compact Support (CS) Kernels). Hence, there is a practical and theoretical need to properly handle npd kernels within feature extraction and classification frameworks. Recently, classifiers such as Support Vector Machines (SVMs) with npd kernels, Indefinite Kernel Fisher Discriminant Analysis (IKFDA) and Indefinite Kernel Quadratic Analysis (IKQA) were proposed. In this paper we propose feature extraction methods using indefinite kernels. In particular, first we propose an Indefinite Kernel Principal Component Analysis (IKPCA). Then, we properly define optimization problems that find discriminant projections with indefinite kernels and propose a Complete Indefinite Kernel Fisher Discriminant Analysis (CIKFDA) that solves the proposed problems. We show the power of the proposed frameworks in a fully automatic face recognition scenario.

Keywords: subspace learning, indefinite kernels, face recognition.

1 Introduction

In many computer vision applications we encounter the following problem. Given a high dimensional visual representation of objects we wish to find a condensed representation that captures their underlying, possibly non-linear, structure. The aforementioned problem is usually tackled by the application of linear and non-linear dimensionality reduction techniques, also referred to as subspace learning techniques. Research on subspace learning mainly revolves around two main interrelated directions, that is (a) subspace learning using kernels [1–4, 7–9] and (b) manifold learning [10, 11].

* This work was funded by Junior Research Fellowship of Imperial College London.

Linear dimensionality reduction is usually performed by finding a set of projections bases while low-dimensional feature extraction is performed by applying these learned bases onto a vector representation of the data. Kernel-based subspace learning methods mainly extend their linear counterparts using (conditionally) positive definite (pd) functions as kernels [1–4, 7–9]. A pd kernel is interpreted as an inner product in a Hilbert space [12]. Kernel-based subspace learning algorithms perform an implicit mapping of the input data into a high-dimensional Hilbert space (also referred to as feature space) and use the reproducing properties of pd kernels to express the projections as a linear combination of the data in the feature space. Dimensionality reduction is then performed by projecting the data in the feature space using the learned bases. All computations are efficiently performed via the inner product of the feature space (the so-called kernel trick).

Notable kernel-based methods include the Kernel Principal Component Analysis (KPCA) [1] and Kernel Fisher Discriminant Analysis (KDA) [2, 7–9]. KDA finds a set of projection bases by maximizing the trace of between-class scatter matrix while minimizing the trace within-class scatter matrix of low-dimensional space. The solution of the KDA optimization problem has resulted in a wealth of research works dealing with the problem of how the range and the useful null space of the within-class scatter matrix can be used for discovering projection bases. The most popular methods discard discriminative information, either in one space or the other [3, 7–9]. A complete framework which extracts features from both spaces was proposed in [2].

The above noted kernel subspace-learning techniques are applicable only in the case of pd kernels. This imposes limitations to their applicability, since many non-pd (npd) kernels arise as similarity measures. For example, in [13–15] the authors tried to incorporate invariance or robustness into the measure. Another family of useful npd kernels are the compact support (cs) kernels [16]. Popular non-Euclidean (nonmetric) similarities/dissimilarities, such as Hausdorff distances [17] and Kullback-Leibler divergence between probability distributions, can be used to define npd kernels [18, 19]. Hence, there is both practical and theoretical need to properly handle all these measures and npd kernels in order to extract discriminant features using an KDA framework with npd kernels. One way to deal with this is to approximate the npd kernel with a positive definite (pd) one and use this kernel instead [6].

The need to properly handle npd kernels, instead of approximating them with pd ones, has initiated a number of studies on the proper design of classification algorithms [20]. In particular in [21] a geometrical interpretation of learning a large margin classifier with indefinite kernels has been discussed. In [21] classification frameworks based on two-class Kernel Fisher Discriminant Analysis (KFDA) and in [18] Kernel Quadratic Discriminant (KQD) analysis with indefinite kernels were proposed. In this paper we study feature extraction with npd (or simple indefinite) kernels. We first formulate Indefinite Kernel Principal Component Analysis (IKPCA) in Krein spaces. A Krein space is a vector space \mathcal{K}

equipped with an indefinite inner product¹. The npd kernel is interpreted as the indefinite inner product of the Krein space. Furthermore, we define optimization problems for extracting discriminant projections in Krein spaces. In particular, we formulate a Complete Indefinite Kernel Discriminant Analysis (CIKDA) in Krein spaces.

We would like to highlight that in [5, 18, 21] only classifiers based on quadratic discriminant functions and two class classifier based on IKFDA in Krein spaces were proposed. Our paper takes a different direction. That is, we propose sub-space learning algorithms in Krein Spaces for feature extraction and object representation. To the best of our knowledge this is the first time that discriminant feature extraction is performed in Krein spaces. Summarizing the contributions of this paper are: (a) an Indefinite Kernel Principal Component Analysis in Krein Spaces² (b) a Complete Indefinite Kernel Fisher Discriminant Analysis (ICKFDA) in Krein spaces. We furthermore propose npd kernels that contrary to [18] achieve state-of-the-art performance in fully automatic face recognition.

2 Krein Spaces

Krein spaces are important as they provide feature-space representations of dissimilarities and provide us with insights on the geometry of classifiers defined with non-positive kernels [18, 21].

An abstract space \mathcal{K} is a Krein space over reals \mathfrak{R} if there exists an (indefinite) inner product $\langle \cdot, \cdot \rangle_{\mathcal{K}} : \mathcal{K} \times \mathcal{K} \rightarrow \mathfrak{R}$ with the following properties [22]:

$$\begin{aligned} \langle \mathbf{x}, \mathbf{y} \rangle_{\mathcal{K}} &= \langle \mathbf{y}, \mathbf{x} \rangle_{\mathcal{K}} \\ \langle c_1 \mathbf{x} + c_2 \mathbf{z}, \mathbf{y} \rangle_{\mathcal{K}} &= c_1 \langle \mathbf{x}, \mathbf{y} \rangle_{\mathcal{K}} + c_2 \langle \mathbf{z}, \mathbf{y} \rangle_{\mathcal{K}} \end{aligned} \quad (1)$$

for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathcal{K}$ and $c_1, c_2 \in \mathfrak{R}$. \mathcal{K} is composed of two vector spaces, such that $\mathcal{K} = \mathcal{K}_+ \oplus \mathcal{K}_-$. \mathcal{K}_+ and \mathcal{K}_- describe two Hilbert spaces over \mathfrak{R} . We denote their corresponding positive definite inner products as $\langle \cdot, \cdot \rangle_{\mathcal{K}_+}$ and $\langle \cdot, \cdot \rangle_{\mathcal{K}_-}$, respectively. The decomposition of \mathcal{K} into two such subspaces defines two orthogonal projections: \mathbf{P}_+ onto \mathcal{K}_+ and \mathbf{P}_- onto \mathcal{K}_- , known as fundamental projections of \mathcal{K} . Using these projections, $\mathbf{x} \in \mathcal{K}$ can be represented as $\mathbf{x} = \mathbf{P}_+ \mathbf{x} + \mathbf{P}_- \mathbf{x}$. The identity matrix in \mathcal{K} is given by $\mathbf{I}_{\mathcal{K}} = \mathbf{P}_+ + \mathbf{P}_-$.

Let us denote by $\mathbf{x}_+ \in \mathcal{K}_+$ and $\mathbf{x}_- \in \mathcal{K}_-$, the projections onto the subspaces $\mathbf{P}_+ \mathbf{x}$ and $\mathbf{P}_- \mathbf{x}$, respectively. Then, $\langle \mathbf{x}_+, \mathbf{y}_- \rangle_{\mathcal{K}} = 0$ for all $\mathbf{x}, \mathbf{y} \in \mathcal{K}$. Moreover, $\langle \mathbf{x}_+, \mathbf{y}_+ \rangle_{\mathcal{K}} > 0$ and $\langle \mathbf{x}_-, \mathbf{y}_- \rangle_{\mathcal{K}} < 0$ for any non-zero vectors \mathbf{x} and \mathbf{y} in \mathcal{K} . Therefore, \mathcal{K}_+ is a positive subspace, while \mathcal{K}_- is a negative subspace. The inner product of \mathcal{K} is defined as the difference of $\langle \cdot, \cdot \rangle_{\mathcal{K}_+}$ and $\langle \cdot, \cdot \rangle_{\mathcal{K}_-}$, *i.e.* for all $\mathbf{x}, \mathbf{y} \in \mathcal{K}$:

$$\langle \mathbf{x}, \mathbf{y} \rangle_{\mathcal{K}} = \langle \mathbf{x}_+, \mathbf{y}_+ \rangle_{\mathcal{K}_+} - \langle \mathbf{x}_-, \mathbf{y}_- \rangle_{\mathcal{K}_-} \quad (2)$$

¹ For more information regarding Krein spaces the interested reader can refer to [22].

² Although, methods similar to the proposed IKPCA were implied in previous works [18, 19] and in Chapter 6 of the PhD thesis [35] a complete formulation of IKPCA in Krein spaces has not been proposed before.

A Krein space \mathcal{K} has an associated Hilbert space $|\mathcal{K}|$ which can be found *via* the linear operator $\mathbf{J} = \mathbf{P}_+ - \mathbf{P}_-$, called the fundamental symmetry. This symmetry satisfies $\mathbf{J} = \mathbf{J}^{-1} = \mathbf{J}^T$ and describes the basic properties of a Krein space. Its connection to the original Krein space can be written in terms of a “conjugate” by using (2) and \mathbf{J} , as

$$\mathbf{x}^* \mathbf{y} \triangleq \langle \mathbf{x}, \mathbf{y} \rangle_{\mathcal{K}} = \mathbf{x}^T \mathbf{J} \mathbf{y} = \langle \mathbf{J} \mathbf{x}, \mathbf{y} \rangle_{|\mathcal{K}|}. \tag{3}$$

That is, \mathcal{K} can be turned into its associated Hilbert space $|\mathcal{K}|$ by using the positive definite inner product of the associated Hilbert space, $\langle \cdot, \cdot \rangle_{|\mathcal{K}|}$, as $\langle \mathbf{x}, \mathbf{y} \rangle_{|\mathcal{K}|} = \langle \mathbf{x}, \mathbf{J} \mathbf{y} \rangle_{\mathcal{K}}$.

In the following we are particularly interested in finite dimensional Krein spaces where \mathcal{K}_+ is isomorphic to \mathfrak{R}^p and \mathcal{K}_- is isomorphic to \mathfrak{R}^q . Such a Krein space describes a pseudo-Euclidean space and is characterized by its so-called signature, $(p, q) \in \mathbb{N}^2$, which indicates the dimensionality, p and q , of the positive and negative subspaces, respectively [18]. The fundamental symmetry is

$$\mathbf{J} = \begin{bmatrix} \mathbf{I}_p & \mathbf{0} \\ \mathbf{0} & -\mathbf{I}_q \end{bmatrix} \tag{4}$$

where \mathbf{I}_z is the identity matrix in $\mathfrak{R}^{z \times z}$ and $\mathbf{0}$ implies zero padding.

A non-positive definite (npd) kernel k defines an implicit mapping $\psi : \mathfrak{R}^d \rightarrow \mathcal{K}$ into a (in)finite dimensional Krein space. Analogously to Hilbert space, our kernel is equivalent to the dot-product in feature space, *i.e.* $k(\mathbf{x}_i, \mathbf{x}_j) = \langle \psi(\mathbf{x}_i), \psi(\mathbf{x}_j) \rangle_{\mathcal{K}}$. The squared distance in feature space is given by

$$\begin{aligned} l^2(\mathbf{x}_i, \mathbf{x}_j) &= (\psi(\mathbf{x}_i) - \psi(\mathbf{x}_j))^* (\psi(\mathbf{x}_i) - \psi(\mathbf{x}_j)) \\ &= k(\mathbf{x}_i, \mathbf{x}_i) - 2k(\mathbf{x}_i, \mathbf{x}_j) + k(\mathbf{x}_j, \mathbf{x}_j). \end{aligned} \tag{5}$$

Also a non-negative dissimilarity measure $l^2(\mathbf{x}_i, \mathbf{x}_j)$ that satisfies the following properties (1) $l^2(\mathbf{x}_i, \mathbf{x}_i) = 0$, (2) $l^2(\mathbf{x}_i, \mathbf{x}_j) > 0, \forall \mathbf{x}_i \neq \mathbf{x}_j$ and (3) $l^2(\mathbf{x}_i, \mathbf{x}_j) = l^2(\mathbf{x}_j, \mathbf{x}_i)$ and does not satisfy the triangular inequality can define an npf kernel.

3 KPCA in Krein Spaces

Let $\mathbf{X} = [\mathbf{x}_1 \cdots \mathbf{x}_N] \in \mathfrak{R}^{d \times N}$ be a set of given samples and $\mathbf{X}_\psi = [\psi(\mathbf{x}_1) \cdots \psi(\mathbf{x}_N)]$ be their implicit mapping. Motivated by KPCA and pseudo-Euclidean embedding [18, 23], we formulate KPCA with Krein spaces.

Let us define the mean $\mathbf{m}^{\mathcal{K}}$, and the centralized matrix $\tilde{\mathbf{X}}_\psi$ as

$$\mathbf{m}^{\mathcal{K}} = \frac{1}{N} \mathbf{X}_\psi \mathbf{1}_N \quad \tilde{\mathbf{X}}_\psi = \mathbf{X}_\psi \mathbf{L} \tag{6}$$

where $\mathbf{L} \triangleq \mathbf{I}_N - \frac{1}{N} \mathbf{1}_N \mathbf{1}_N^T$ and $\mathbf{1}_N$ is an N -dimensional vector containing only ones [18]. We then define the total scatter matrix in \mathcal{K} as

$$\mathbf{S}_t^{\mathcal{K}} \triangleq \frac{1}{N} \sum_{i=1}^N (\psi(\mathbf{x}_i) - \mathbf{m}^{\mathcal{K}}) (\psi(\mathbf{x}_i) - \mathbf{m}^{\mathcal{K}})^* = \frac{1}{N} \tilde{\mathbf{X}}_\psi \tilde{\mathbf{X}}_\psi^* = \frac{1}{N} \tilde{\mathbf{X}}_\psi \tilde{\mathbf{X}}_\psi^T \mathbf{J} = \mathbf{S}_{|\mathcal{K}|} \mathbf{J} \tag{7}$$

where $\mathbf{S}_{|\mathcal{K}|}$ is the total scatter matrix in the associated Hilbert space $|\mathcal{K}|$.

In a similar way to that of KPCA in Hilbert space, we generalize KPCA in Krein space as follows. We wish to compute a set of projections $\mathbf{U}_o = [\mathbf{u}_1 \cdots, \mathbf{u}_N]$ with $\mathbf{u}_i \in \mathcal{K}$ such that³

$$\mathbf{U}_o = \arg \max_{\mathbf{U}} \text{tr} (\mathbf{U}^* \mathbf{S}_t^{\mathcal{K}} \mathbf{U}) \quad \text{s.t. } \mathbf{U}^* \mathbf{U} = \mathbf{J}. \tag{8}$$

We write the set of projections as a linear combination of samples as $\mathbf{U} = \tilde{\mathbf{X}}_{\psi} \mathbf{Q}$, and (8) becomes:

$$\begin{aligned} \mathbf{Q}_o &= \arg \max_{\mathbf{Q}} \text{tr} \left(\mathbf{Q}^T \tilde{\mathbf{X}}_{\psi}^T \mathbf{J} \tilde{\mathbf{X}}_{\psi} \tilde{\mathbf{X}}_{\psi}^T \mathbf{J} \tilde{\mathbf{X}}_{\psi} \mathbf{Q} \right) \\ &= \arg \max_{\mathbf{Q}} \text{tr} \left(\mathbf{Q}^T \tilde{\mathbf{K}} \tilde{\mathbf{K}} \mathbf{Q} \right) \end{aligned} \quad \text{s.t. } \mathbf{Q}^T \tilde{\mathbf{X}}_{\psi}^T \mathbf{J} \tilde{\mathbf{X}}_{\psi} \mathbf{Q} = \mathbf{Q}^T \mathbf{K} \mathbf{Q} = \mathbf{J} \tag{9}$$

where $\tilde{\mathbf{K}} = \tilde{\mathbf{X}}_{\psi}^* \tilde{\mathbf{X}}_{\psi}$ is the centralized kernel matrix. The eigendecomposition of $\tilde{\mathbf{K}}$ then yields the solution of the above

$$\tilde{\mathbf{K}} = \mathbf{V} \mathbf{\Lambda} \mathbf{V}^T = \mathbf{V} |\mathbf{\Lambda}|^{\frac{1}{2}} \mathbf{J} |\mathbf{\Lambda}|^{\frac{1}{2}} \mathbf{V}^T \tag{10}$$

where $\mathbf{\Lambda}$ is a diagonal matrix whose main diagonal consists of p positive and q negative eigenvalues ($p+q \leq N$) in the following order: first, positive eigenvalues with decreasing values, then negative ones with decreasing absolute values and finally zero values. Matrix $|\mathbf{\Lambda}|$ is the diagonal matrix containing the absolute values of the eigenvalues. The fundamental symmetry, matrix \mathbf{J} , is defined as in (4), and (p, q) is the pseudo-Euclidian space’s signature. Consequently, we obtain the optimal solution of (9) from $\mathbf{Q}_o = \mathbf{V}_{p+q} |\mathbf{\Lambda}_{p+q}|^{-\frac{1}{2}}$ and the optimal projection matrix from $\mathbf{U}_o = \tilde{\mathbf{X}}_{\psi} \mathbf{V}_{p+q} |\mathbf{\Lambda}_{p+q}|^{-\frac{1}{2}}$, where $\mathbf{\Lambda}_{p+q}$ contains the non-zero eigenvalues and \mathbf{V}_{p+q} denotes the corresponding eigenvectors.

Let $\mathbf{y} \in \mathcal{C}^d$ be a new sample, and $\hat{\mathbf{y}} = \psi(\mathbf{y}) \in \mathcal{K}$ denotes its mapping. Then, the part of $\hat{\mathbf{y}}$ which belongs to the positive subspace \mathfrak{R}^p is given by:

$$\begin{aligned} \hat{\mathbf{y}}_+ &= |\mathbf{\Lambda}_p|^{-\frac{1}{2}} \mathbf{V}_p^T \mathbf{M}^T \mathbf{X}_{\psi}^* \psi(\mathbf{y}) \\ &= |\mathbf{\Lambda}_p|^{-\frac{1}{2}} \mathbf{V}_p^T \mathbf{M}^T \begin{bmatrix} \langle \psi(\mathbf{x}_1), \psi(\mathbf{y}) \rangle_{\mathcal{K}} \\ \dots \\ \langle \psi(\mathbf{x}_N), \psi(\mathbf{y}) \rangle_{\mathcal{K}} \end{bmatrix} = |\mathbf{\Lambda}_p|^{-\frac{1}{2}} \mathbf{V}_p^T \mathbf{M}^T \begin{bmatrix} k(\mathbf{x}_1, \mathbf{y}) \\ \dots \\ k(\mathbf{x}_N, \mathbf{y}) \end{bmatrix} \end{aligned} \tag{11}$$

where $\mathbf{\Lambda}_p$ contains only the positive eigenvalues, and \mathbf{V}_p denotes the corresponding eigenvectors. Similarly, we can compute the features $\hat{\mathbf{y}}_- \in \mathfrak{R}^q$ using

$$\hat{\mathbf{y}}_- = |\mathbf{\Lambda}_q|^{-\frac{1}{2}} \mathbf{V}_q^T \mathbf{M}^T \mathbf{X}_{\psi}^* \psi(\mathbf{y}) \tag{12}$$

where $\mathbf{\Lambda}_q$ and \mathbf{V}_q corresponds to the negative eigenvalues. Furthermore, we can verify that the inner product of $\hat{\mathbf{x}}, \hat{\mathbf{y}} \in \mathcal{K}$ is equal to the kernel value as follows

$$\begin{aligned} \langle \hat{\mathbf{x}}, \hat{\mathbf{y}} \rangle_{\mathcal{K}} &= \hat{\mathbf{x}}^* \hat{\mathbf{y}} = \hat{\mathbf{x}}^T \mathbf{J} \hat{\mathbf{y}} = \psi(\mathbf{x})^* \tilde{\mathbf{X}}_{\psi} \mathbf{V} |\mathbf{\Lambda}|^{-\frac{1}{2}} \mathbf{J} |\mathbf{\Lambda}|^{-\frac{1}{2}} \mathbf{V}^T \tilde{\mathbf{X}}_{\psi}^* \psi(\mathbf{y}) \\ &= \psi(\mathbf{x})^T \mathbf{J} \mathbf{U}^* \mathbf{U} \mathbf{J} \psi(\mathbf{y}) = \psi(\mathbf{x})^T \mathbf{J} \psi(\mathbf{y}) = \langle \psi(\mathbf{x}), \psi(\mathbf{y}) \rangle_{\mathcal{K}} = k(\mathbf{x}, \mathbf{y}). \end{aligned} \tag{13}$$

³ Although pseudo-euclidean embedding has been proposed [19] the actual formulation of KPCA in Krein Spaces has not been previously proposed.

In order to establish a dimensionality reduction strategy, we can start by expanding the objective function of the optimization problem (8) as

$$\begin{aligned} \text{tr}(\mathbf{U}^* \mathbf{S}_{\mathcal{K}} \mathbf{U}) &= \text{tr}(\mathbf{Q}^T \tilde{\mathbf{K}} \tilde{\mathbf{K}} \mathbf{Q}) = \text{tr}(|\mathbf{\Lambda}|^{-\frac{1}{2}} \mathbf{V}^T \mathbf{V} \mathbf{\Lambda} \mathbf{V}^T \mathbf{V} \mathbf{\Lambda} \mathbf{V}^T \mathbf{V} |\mathbf{\Lambda}|^{-\frac{1}{2}}) \\ &= \text{tr}(|\mathbf{\Lambda}|) = \sum_{i=1}^N |\lambda_i|. \end{aligned} \tag{14}$$

As it can be observed, the actual functional to be minimized is based on the absolute eigenvalues, $|\lambda_i|$. Hence, the dimensionality reduction may be performed by removing the eigenvectors that correspond to the smallest in terms of magnitude eigenvalues. The signature of the reduced Krein space is then given by (p_1, q_1) with $p_1 \leq p$ and $q_1 \leq q$.

4 Discriminant Learning in Krein Spaces

Kernel Discriminant Analysis (KDA) in Hilbert spaces with positive definite (pd) kernels aims at finding discriminant projection bases by exploiting class-label information in the feature space. In the following we will formulate discriminant subspace learning algorithms by defining optimization problems based on the traces of the projected within and between class scatter matrices. We assume that our training set consists of C classes $\mathcal{C}_1, \dots, \mathcal{C}_C$. N_c denotes the cardinality of set \mathcal{C}_c . We define the between-class, within-class and total scatter matrices $\mathbf{S}_b^{\mathcal{K}}$, $\mathbf{S}_w^{\mathcal{K}}$ and $\mathbf{S}_t^{\mathcal{K}}$ in \mathcal{K} as

$$\mathbf{S}_b^{\mathcal{K}} \triangleq \sum_{c=1}^C N_c (\mathbf{m}_c^{\mathcal{K}} - \mathbf{m}^{\mathcal{K}}) (\mathbf{m}_c^{\mathcal{K}} - \mathbf{m}^{\mathcal{K}})^* \tag{15}$$

$$\mathbf{S}_w^{\mathcal{K}} \triangleq \sum_{c=1}^K \sum_{\mathbf{x}_i \in \mathcal{C}_c} (\psi(\mathbf{x}_i) - \mathbf{m}_c^{\mathcal{K}}) (\psi(\mathbf{x}_i) - \mathbf{m}_c^{\mathcal{K}})^* \tag{16}$$

where $\mathbf{m}_c^{\mathcal{K}} = \frac{1}{N_c} \sum_{\mathbf{x}_i \in \mathcal{C}_c} \psi(\mathbf{x}_i)$ is the mean vector of each class.

In Hilbert spaces with pd kernels the main optimization problem for finding the discriminant projection is

- the one that maximizes the trace of the projected between class scatter matrix subject to having a projected orthogonal within-class scatter matrix [2, 9, 24]
- maximizes the trace of the projected between class scatter matrix subject to the useful null-space of within-class scatter matrix [2, 3, 25].

Using the theory developed in the previous Section, we formulate the optimization problems that find the discriminant projections with npf kernels in Krein spaces. That is, we aim at finding a set of projections $\mathbf{U}_1 = [\mathbf{u}_1^1 | \dots | \mathbf{u}_1^p]$ with every column $\mathbf{u}_1^j \in \mathcal{K}$

$$\mathbf{U}_1 = \max_{\mathbf{U}} \text{tr}[\mathbf{U}^* \mathbf{S}_b^{\mathcal{K}} \mathbf{U}] \text{ s.t } \mathbf{U}^* \mathbf{S}_w^{\mathcal{K}} \mathbf{U} = \mathbf{I}, \tag{17}$$

and $\mathbf{U}_2 = [\mathbf{u}_1^2 | \dots | \mathbf{u}_p^2]$ with every column $\mathbf{u}_j^2 \in \mathcal{K}$

$$\mathbf{U}_2 = \max_{\mathbf{U}} \text{tr} [\mathbf{U}^* \mathbf{S}_b^{\mathcal{K}} \mathbf{U}] \quad \text{s.t.} \quad \mathbf{U}^* \mathbf{S}_w^{\mathcal{K}} \mathbf{U} = \mathbf{0}. \quad (18)$$

The equivalent optimization problem (17) in Hilbert spaces was solved in [2, 9, 24], while approaches to solve optimization problem (18) were proposed in [2, 3, 25]. The Complete Kernel Fisher Discriminant (CKF) framework [2] solves the equivalent optimization problems (17) and (18) simultaneously by projecting the within-class scatter matrix onto the non-null space of total scatter matrix. The CKFD framework is not applicable in our case, since the developed theoretical framework in [2] can be only applied for the case of pd kernels. To alleviate this problem, in the following section, we propose the Complete Fisher Discriminant with Indefinite Kernels (CFDIK) in Krein spaces. To the best of our knowledge this is the first time discriminant subspace algorithms are proposed in Krein spaces with npd kernels.

5 Solving the Optimization Problems

In the following we will show how optimization problems (17) and (18) can be solved. Let us first define the block $\mathbf{M}_c \triangleq \frac{1}{N_c} \mathbf{1}_{N_c} \mathbf{1}_{N_c}^T$ and the block diagonal matrix \mathbf{M} as:

$$\mathbf{M} \triangleq \text{diag}[\mathbf{M}_1, \mathbf{M}_2, \dots, \mathbf{M}_C]. \quad (19)$$

The useful properties of \mathbf{M} are: (1) \mathbf{M} is idempotent, i.e. $\mathbf{M}^n = \mathbf{M}$ with $n \neq 0$, (2) $\mathbf{I} - \mathbf{M}$ is idempotent, (3) \mathbf{M} has C eigenvectors corresponding to C non-zero eigenvalues, (4) $\mathbf{I} - \mathbf{M}$ has $N - C$ eigenvectors corresponding to $N - C$ non-zero eigenvalues (5) for a full ranked symmetric matrix $\mathbf{A} \in \Re^{N \times N}$ matrices $\mathbf{A}\mathbf{M}\mathbf{A}$ and $\mathbf{A}(\mathbf{I} - \mathbf{M})\mathbf{A}^4$ have C and $N - C$ N eigenvectors corresponding to C and $N - C$ non-zero (positive) eigenvalues, respectively.

Using \mathbf{M} and the fact that $\mathbf{S}_w^{\mathcal{K}} = \mathbf{S}_t^{\mathcal{K}} - \mathbf{S}_b^{\mathcal{K}}$, $\mathbf{S}_w^{\mathcal{K}}$ we write

$$\mathbf{S}_b^{\mathcal{K}} = \tilde{\mathbf{X}}_{\psi} \mathbf{M} \tilde{\mathbf{X}}_{\psi}^*, \quad \mathbf{S}_w^{\mathcal{K}} = \tilde{\mathbf{X}}_{\psi} (\mathbf{I} - \mathbf{M}) \tilde{\mathbf{X}}_{\psi}^* \quad (20)$$

5.1 Solving the Optimization Problem (17)

In this section, we present how to diagonalize the within-class scatter matrix $\mathbf{S}_w^{\mathcal{K}}$ in the Krein feature space. Before proceeding we need the following Theorem I.

Theorem I: Define matrices \mathbf{A} and \mathbf{B} such that $\mathbf{A} = \Phi \Phi^*$ and $\mathbf{B} = \Phi^* \Phi$. Let \mathbf{U}_B be the eigenvectors corresponding to the non-zero eigenvalues Λ_B of \mathbf{B} . Then, $\mathbf{U}_A = \Phi \mathbf{U}_B |\Lambda_B|^{-1}$ diagonalizes $\Phi \Phi^*$,

The proof is omitted due to lack of space.

Using the fact that $\mathbf{I} - \mathbf{M}$ is idempotent, $\mathbf{S}_w^{\mathcal{H}}$ can be written as

$$\mathbf{S}_w^{\mathcal{K}} = \tilde{\mathbf{X}}_{\psi} (\mathbf{I} - \mathbf{M}) (\tilde{\mathbf{X}}_{\psi})^* = \left(\tilde{\mathbf{X}}_{\psi} (\mathbf{I} - \mathbf{M}) \right) \left(\tilde{\mathbf{X}}_{\psi} (\mathbf{I} - \mathbf{M}) \right)^*. \quad (21)$$

⁴ \forall symmetric real matrices \mathbf{A} matrices $\mathbf{A}\mathbf{M}\mathbf{A} = (\mathbf{M}\mathbf{A})^T \mathbf{M}\mathbf{A}$ and $\mathbf{A}(\mathbf{I} - \mathbf{M})\mathbf{A} = ((\mathbf{I} - \mathbf{M})\mathbf{A})^T (\mathbf{I} - \mathbf{M})\mathbf{A}$ are positive semi definite by construction.

By Theorem I, in order to diagonalize \mathbf{S}_w^K we need to apply eigen-analysis to \mathbf{K}_w

$$\mathbf{K}_w = \left(\tilde{\mathbf{X}}_\psi (\mathbf{I} - \mathbf{M}) \right)^* \left(\tilde{\mathbf{X}}_\psi (\mathbf{I} - \mathbf{M}) \right) = (\mathbf{I} - \mathbf{M}) \tilde{\mathbf{K}} (\mathbf{I} - \mathbf{M}). \quad (22)$$

Since, $\tilde{\mathbf{K}}$ is npd so is \mathbf{K}_w hence it admits an eigendecomposition as

$$\mathbf{K}_w = \mathbf{Q}_w |\Lambda_w|^{\frac{1}{2}} \mathbf{J} |\Lambda_w|^{\frac{1}{2}} \mathbf{Q}_w^T. \quad (23)$$

Now we seek an optimal solution that can be written as a linear combination of matrix $\tilde{\mathbf{X}}_\psi (\mathbf{I} - \mathbf{M}) \mathbf{Q}_w |\Lambda_w|^{-1}$ which diagonalizes \mathbf{S}_w^K , i.e.

$$\mathbf{U} = \tilde{\mathbf{X}}_\psi (\mathbf{I} - \mathbf{M}) \mathbf{Q}_w |\Lambda_w|^{-1} \mathbf{A}. \quad (24)$$

where $\mathbf{A} \in \mathfrak{R}^{(N-C) \times C}$. Using \mathbf{U} the objective matrix $\mathbf{U}^* \mathbf{S}_b^K \mathbf{U}$ is reformulated as

$$\begin{aligned} \mathbf{U}^* \mathbf{S}_b^K \mathbf{U} &= \mathbf{A}^T |\Lambda_w|^{-1} \mathbf{Q}_w^T (\mathbf{I} - \mathbf{M}) \tilde{\mathbf{X}}_\psi^* \tilde{\mathbf{X}}_\psi \mathbf{M} \tilde{\mathbf{X}}_\psi^* \tilde{\mathbf{X}}_\psi (\mathbf{I} - \mathbf{M}) \mathbf{Q}_w |\Lambda_w|^{-1} \mathbf{A} \\ &= \mathbf{A}^T |\Lambda_w|^{-1} \mathbf{Q}_w^T (\mathbf{I} - \mathbf{M}) \tilde{\mathbf{K}} \mathbf{M} \tilde{\mathbf{K}} (\mathbf{I} - \mathbf{M}) \mathbf{Q}_w |\Lambda_w|^{-1} \mathbf{A} \\ &= \mathbf{A}^T \left(\mathbf{M} \tilde{\mathbf{K}} (\mathbf{I} - \mathbf{M}) \mathbf{Q}_w |\Lambda_w|^{-1} \right)^T \left(\mathbf{M} \tilde{\mathbf{K}} (\mathbf{I} - \mathbf{M}) \mathbf{Q}_w |\Lambda_w|^{-1} \right) \mathbf{A} \end{aligned} \quad (25)$$

$\mathbf{K}_b = |\Lambda_w|^{-1} \mathbf{Q}_w^T (\mathbf{I} - \mathbf{M}) \tilde{\mathbf{K}} \mathbf{M} \tilde{\mathbf{K}} (\mathbf{I} - \mathbf{M}) \mathbf{Q}_w |\Lambda_w|^{-1}$ is positive semi-definite by definition. Then, optimization problem (17) is reformulated as

$$\mathbf{A}_o = \max_{\mathbf{A}} \operatorname{tr} [\mathbf{A}^T \mathbf{K}_b \mathbf{A}] \quad \text{s.t. } \mathbf{A}^T \mathbf{A} = \mathbf{I}, \quad (26)$$

which is solved by the choosing \mathbf{A}_o to contain as columns the $C - 1$ eigenvectors of \mathbf{K}_b that correspond to non-zero eigenvalues.

5.2 Solving Optimization Problem (18)

We cannot solve the optimization problem (18) by writing the solution \mathbf{U}_l as a linear combination of $\tilde{\mathbf{X}}_\psi (\mathbf{I} - \mathbf{M}) \mathbf{Q}_l$ where \mathbf{Q}_l is the complementary subspace of \mathbf{Q}_w (i.e., the eigenvectors of $(\mathbf{I} - \mathbf{M}) \tilde{\mathbf{K}} (\mathbf{I} - \mathbf{M})$ that correspond to null eigenvalues). Such a solution should be written as $\mathbf{U}_l = \tilde{\mathbf{X}}_\psi (\mathbf{I} - \mathbf{M}) \mathbf{Q}_l \mathbf{A}$. We have

$$\mathbf{U}_l^* \mathbf{U}_l = \mathbf{A}^T \mathbf{Q}_l^T (\mathbf{I} - \mathbf{M}) (\tilde{\mathbf{X}}_\psi)^* \tilde{\mathbf{X}}_\psi (\mathbf{I} - \mathbf{M}) \mathbf{Q}_l \mathbf{A} = \mathbf{0}, \quad (27)$$

which further gives $\mathbf{U}_l = \mathbf{0}$.

To find \mathbf{U}_l , we write $\mathbf{U}_l = \tilde{\mathbf{X}}_\psi \Xi_l \mathbf{A}$, $\Xi_l \in \mathfrak{R}^{N \times C}$, $\mathbf{A} \in \mathfrak{R}^{C \times (C-1)}$. Additionally, \mathbf{U}_l satisfies

$$\mathbf{U}_l^* \mathbf{S}_w^K \mathbf{U}_l = \mathbf{A}^T \Xi_l^T \tilde{\mathbf{K}} (\mathbf{I} - \mathbf{M}) \tilde{\mathbf{K}} \Xi_l \mathbf{A} = \mathbf{0}. \quad (28)$$

From the properties of matrices $\mathbf{I} - \mathbf{M}$, $\tilde{\mathbf{K}}$, ($\tilde{\mathbf{K}}^T = \tilde{\mathbf{K}}$) has $N - C$ non-zero eigenvalues. The constraint (28) can be satisfied by choosing Ξ_l from performing eigenanalysis of $\tilde{\mathbf{K}} (\mathbf{I} - \mathbf{M}) \tilde{\mathbf{K}}$ and keeping the C eigenvectors which correspond to the zero eigenvalues.

Using \mathbf{U}_l , $\mathbf{U}_l^* \mathbf{S}_b^{\mathcal{K}} \mathbf{U}_l$ is reformulated as

$$\begin{aligned} \mathbf{U}_l^* \mathbf{S}_b^{\mathcal{K}} \mathbf{U}_l &= \mathbf{A}^T \Xi_l^T \tilde{\mathbf{X}}_\psi^* \tilde{\mathbf{X}}_\psi \mathbf{M} \tilde{\mathbf{X}}_\psi^* \tilde{\mathbf{X}}_\psi \Xi_l \mathbf{A} = \mathbf{A}^T \Xi_l^T \tilde{\mathbf{K}} \mathbf{M} \tilde{\mathbf{K}} \Xi_l \mathbf{A} \\ &= \mathbf{A}^T \left(\mathbf{M} \tilde{\mathbf{K}} \Xi_l \right)^T \left(\mathbf{M} \tilde{\mathbf{K}} \Xi_l \right) \mathbf{A} \end{aligned} \tag{29}$$

$\mathbf{K}_{b,2} = \Xi_l^T \tilde{\mathbf{K}} \mathbf{M} \tilde{\mathbf{K}} \Xi_l$ is positive semi-definite by construction. Using the above equation optimization problem (18) is reformulated as

$$\mathbf{A}_o = \max_{\mathbf{A}} \text{tr} \left[\mathbf{A}^T \mathbf{K}_{b,2} \mathbf{A} \right], \text{ s.t. } \mathbf{A}^T \mathbf{A} = \mathbf{I}, \tag{30}$$

which is solved by the eigenanalysis of \mathbf{K}_b and keeping the $C - 1$ eigenvectors that correspond to the $C - 1$ non-zero eigenvalues. Finally we prove that the projections \mathbf{U}_w and \mathbf{U}_l derived from the optimization problems (17) and (18) are orthogonal ($\mathbf{U}_l^* \mathbf{U}_w = \mathbf{0}$) (the proof is omitted due to lack of space).

6 Comparison with the Methods in [18],[21],[26],[27]

The literature regarding learning with indefinite kernels mainly revolves around the design of classifiers [18, 21, 27]. In particular in [21] an geometrical interpretation of learning a large margin classifier with indefinite kernels has been given. The most closely related works are the classification frameworks proposed in [18] and [21] for two class problems.

In this problem we have two classes \mathcal{C}_1 and \mathcal{C}_2 . We define matrices $\mathbf{S}_b^{\mathcal{K}}$ and $\mathbf{S}_w^{\mathcal{K}}$ as in (15) and in (16), respectively, for the two class problem. Then, the methods in [18, 21] find a vector $\mathbf{w} \in \mathcal{K}$ and a scalar b such that

$$\mathbf{w}_o = \arg \max_{\mathbf{w}} \frac{\mathbf{w}^* \mathbf{S}_b^{\mathcal{K}} \mathbf{w}}{\mathbf{w}^* \mathbf{S}_w^{\mathcal{K}} \mathbf{w}}. \tag{31}$$

In order to solve the above optimization problem \mathbf{w} was written as a linear combination of the training samples as $\mathbf{w} = \sum_{i=1}^n \mathbf{a}_i \psi(\mathbf{x}_i) = \mathbf{X}_\psi \mathbf{a}$ then optimization problem (31) can be written as

$$\mathbf{a}_o = \arg \max_{\mathbf{a}} \frac{\mathbf{a}^T \mathbf{K} \mathbf{M} \mathbf{K} \mathbf{a}}{\mathbf{a}^T \mathbf{K} (\mathbf{I} - \mathbf{M}) \mathbf{K} \mathbf{a}}, \tag{32}$$

since matrices $\mathbf{N} = \mathbf{K} \mathbf{M} \mathbf{K}$ and $\mathbf{C} = \mathbf{K} (\mathbf{I} - \mathbf{M}) \mathbf{K}$ are positive semi-definite by construction, the solution is given by keeping the eigenvector that corresponds to the largest eigenvalues of $\mathbf{C}^{-1} \mathbf{N}$. The matrix \mathbf{C} is not invertible since it contains only one eigenvector that corresponds to non-zero eigenvalues. In [18, 21] a standard heuristic approach was applied, i.e. \mathbf{a} was found by performing eigenanalysis to $(\mathbf{C} + \beta \mathbf{I})^{-1} \mathbf{N}$ where β is a small positive constant arbitrarily chosen. Unfortunately, this is not the solution to the optimization problem (31). Since, for two-class data both optimization problems (31) and (17) we can readily find the optimal \mathbf{w} that optimizes both (31) by applying the methodology proposed in Section 5.

7 Experimental Results

We tested the proposed IKPCA and CKFDA approaches in the face recognition problem. The first indefinite kernel we used is the CS kernel used in [16] (also widely referred to as a mollifier)

$$k(\mathbf{x}, \mathbf{y}) = \exp\left(\frac{1}{\|\mathbf{x}-\mathbf{y}\|^2-\gamma}\right) H(\alpha - \|\mathbf{x}-\mathbf{y}\|^2) \quad (33)$$

where $H(r)$ is the usual heaviside function. The CS kernels are of great importance in robust statistics since they are less influenced by outliers. In our experiments we used as distance $\|\mathbf{x}-\mathbf{y}\|^2$ the weighted distance $\|\mathbf{x}-\mathbf{y}\|_{\mathbf{W}}^2 = (\mathbf{x}-\mathbf{y})^T \mathbf{W}(\mathbf{x}-\mathbf{y})$ proposed in [28] where \mathbf{W} is sum of the power spectrum of the filters used and \mathbf{x} and \mathbf{y} are the vectorized Fourier responses of the images. As filters we used a Gabor filter bank of 8 orientations and 5 scales.

The second class of indefinite kernels we used are defined as the minimum of the correlation surface of image registration algorithms [29]

$$k(I_i, I_j) = \min(\min \text{cor}(I_i, I_j), \min \text{cor}(I_j, I_i)) \quad (34)$$

where $\text{cor}(I_i, I_j)$ is the correlation surface between two images I_i and I_j of the matching algorithm [29]. We chose this particular kernel in order to illustrate the applicability of the proposed feature extraction methods in fully automatic face recognition schemes.

All the reported results were acquired using $C-1$ features produced from the optimization problem (17) and $C-1$ features from the optimization problem (18). For the IKPCA algorithm the reported results were acquired using $N-1$ features produced by the algorithm presented in Section 3. The classifier used was a simple nearest neighbor classifier using as distance the normalized correlation or the projected features.

7.1 Face Recognition Experiments in Yale B Database

The extended Yale B database [30] contains 16128 images of 38 subjects under 9 poses and 64 illumination conditions. We used a subset that consists of 64 near frontal images for each subject. For training, we randomly selected a subset with 5, 10 and 20 images per subject. The training set was also further split into training and validation to find the optimal parameters of the kernels used (i.e., γ and α). For testing, we used the remaining images. Finally, we performed 20 different random realizations of the training/test sets.

For comparison reasons we used the the pd Gaussian RBF (GRBF) kernel using the same distances $\|\mathbf{x}-\mathbf{y}\|_{\mathbf{W}}^2$ as $k(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{1}{\sigma^2}\|\mathbf{x}-\mathbf{y}\|_{\mathbf{W}}^2\right)$. Using the GRBF kernel the IKPCA and the proposed CIKFDA framework collapse to the KPCA and CKFDA [2] frameworks, respectively. Table I summarizes the obtained results. As we can see the proposed IKPCA and CIKFDA with the proposed npd kernel outperforms all other algorithms.

Table 1. Average recognition rates and standard deviations on the Extended YALE B database

5—10—20	Proposed Kernel (33)			GRBF		
IKPCA	80.5(1.12)	93.5(0.89)	96.6(0.25)	78.8(1.02)	90.8(0.83)	93.4(0.75)
CIKFDA (17)	77.2(1.12)	93.1(0.82)	97.8 (0.25)	76.8(1.67)	90.4(1.01)	95.9(0.88)
CIKFDA (18)	74.6(1.8)	92.7(0.76)	97.1 (0.3)	75.4(1.63)	89.1(1.21)	95.0(0.73)

Face Recognition Experiments in a Subset of FERET. In order to simulate results acquired using a fully automatic system, we used directly the faces returned from a face detector (both training and testing)⁵. The kernel in (34) was used for matching the faces provided by the face detector. To the best of our knowledge there are very few works reporting results in such a difficult setting with the most recent one the work published in [31].

In this experiment, we attempted to combine the experimental setting suggested in [2, 4] with facial images obtained directly from the face detector. We did so in order to show the power of the proposed CIKFDA when more than one images are available for training. In particular, we randomly selected 600 facial images corresponding to 200 subjects, such that each subject has three images (taken from FA, FB, DupI and DupII). We randomly chose two out of three images for training and then used the third image for testing.

Table 2 summarizes the recognition rates. We also report recently proposed state-of-the-art methods for face recognition [33] in manually aligned facial images (aligned according to the eye coordinates), for comparison reasons. We also compared our method with LBPs using both manually aligned images and detector extracted images. We achieved a recognition rate of 95% which demonstrates that the proposed scheme can be efficiently combined with fully automatic methods for face detection and matching. We also significantly outperform (by 7%) state-of-the-art methods that used **manually** aligned data.

Table 2. Recognition rates in the subset of FERET. SRC represents the results acquired using the method in [33] with manual alignment. LBP-d represents the results of Local Binary Patterns using detector extracted images and LBP-m represents the results with images after manual alignment.

Methods	SRC [33]	LBP-d	LBP-m	IKPCA	CIKFDA (17)	CIKFDA (17)
RR	87	35	89	90	95	93.5

7.2 Face Recognition with the XM2VTS Database

We carried out face verification experiments on the test set of Configuration I of the XM2VTS database. The training set contained 200 subjects with three images per subject which enabled us to apply our kernel combined with the

⁵ In particular we used the publicly available face detector implemented in OpenCV

proposed discriminant analysis. The evaluation set contained three images per client for genuine claims and 25 evaluation impostors with eight images per impostor. The testing set contained two images per client and 70 impostors with eight images per impostor. For additional details on the XM2VTS database and the protocol used, the interested reader may refer to [34].

A face detector was also used to provide the faces. The applied kernel combined with the proposed kernel discriminant analysis on the samples of the face detector achieved a TER (Total Error Rate)⁶ equal to 1.92%. Table 3 summarizes the best results of each competition in fully automatic facial image registration scenarios, as well as, the performance of some recent algorithms tested under automatic alignment by using the eye coordinates. Our method, which is applied directly on the results of the detector, achieved a TER which is the best reported for the XM2VTS database according to the best of our knowledge. The results reported with the SRC method in [33] were achieved using manually aligned data.

Table 3. Best Results in XM2VTS database under automatic image alignment

Methods	Best of [36]	Best of [37]	Best of [38]	[39]	SRC [33]	Proposed Approach
TER%	13.10	3.86	2.14	2.3	4	1.92

8 Conclusions

In this paper we presented a theoretical framework for discriminant feature extraction in Krein spaces. In particular we proposed a Complete Indefinite Kernel Fisher Discriminant Analysis (CIKFDA) which discovers discriminant projections both in the range and null spaces of the within-class-scatter matrix in the Krein space. We demonstrated the superiority of the proposed approach in fully automatic face recognition scenarios where state-of-the-art results were achieved using the output images acquired from a face detector.

References

1. Scholkopf, B., Smola, A., Müller, K.R.: Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation* 10, 1299–1319 (1998)
2. Yang, J., Frangi, A.F., Yang, J., Zhang, D., Jin, Z.: KPCA plus LDA: A complete kernel Fisher discriminant framework for feature extraction and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27, 230–244 (2005)
3. Cevikalp, H., Neamtu, M., Wilkes, M.: Discriminative common vector method with kernels. *IEEE Transactions on Neural Networks* 17 (2006)
4. Chengjun, L.: Capitalize on dimensionality increasing techniques for improving face recognition grand challenge performance. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28, 725–737 (2006)

⁶ For more details regarding the TER refer to [34].

5. Ong, C.S., Mary, X., Canu, S., Smola, A.J.: Learning with non-positive kernels. In: ICML, pp. 81–88 (2004)
6. Chen, Y., Gupta, M.R., Recht, B.: Learning kernels from indefinite similarities. In: ICML, pp. 145–152 (2009)
7. Lu, J., Plataniotis, K.N., Venetsanopoulos, A.N.: Face recognition using kernel direct discriminant analysis algorithms. *IEEE Transactions on Neural Networks* 14, 117–126 (2003)
8. Yang, M.: Kernel eigenfaces vs. kernel fisherfaces: Face recognition using kernel methods. In: FG, pp. 215–220 (2002)
9. Baudat, G., Anouar, F.: Generalized discriminant analysis using a kernel approach. *Neural Computation* 12, 2385–2404 (2000)
10. Roweis, S., Saul, L.: Nonlinear dimensionality reduction by locally linear embedding. *Science* 290 (2000)
11. Tenenbaum, J.B., De Silva, V., Langford, J.C.: A global geometric framework for nonlinear dimensionality reduction. *Science* 290 (2000)
12. Scholkopf, B., Smola, A.: *Learning with Kernels*. MIT Press, Cambridge (2002)
13. Simard, P., LeCun, Y., Denker, J., Victorri, B.: Transformation invariance in pattern recognition: Tangent distance and propagation. *IJIST* 11, 181–197 (2000)
14. Haasdonk, B., Burkhardt, H.: Invariant kernel functions for pattern analysis and machine learning. *Machine Learning* 68, 35–61 (2007)
15. Jacobs, D., Weinshall, D., Gdalyahu, Y.: Class representation and image retrieval with non-metric distances. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22 (2000)
16. Jamshidi, A., Kirby, M.: Examples of compactly supported functions for radial basis approximations. In: ICML, pp. 155–160 (2006)
17. Dubuisson, M., Jain, A.: A modified hausdorff distance for object matching. In: ICPR, vol. 1, pp. 566–568 (1994)
18. Pekalska, E., Haasdonk, B.: Kernel discriminant analysis for positive definite and indefinite kernels. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31 (2009)
19. Pekalska, E., Paclik, P., Duin, R.P.W.: A generalized kernel approach to dissimilarity-based classification. *Journal of Machine Learning Research* 2, 175–211 (2002)
20. Woznica, A., Kalousis, A., Hilario, M.: Distances and (indefinite) kernels for sets of objects. In: ICDM, pp. 1151–1156 (2006)
21. Haasdonk, B., Pekalska, E.: Indefinite kernel fisher discriminant. In: ICPR (2008)
22. Hassibi, B., Sayed, A.H., Kailath, T.: Linear estimation in Krein spaces. I. Theory. *IEEE Transactions on Automatic Control* 41, 18–33 (1996)
23. Pekalska, E., Duin, R.: *The dissimilarity representation for pattern recognition: foundations and applications*. World Scientific Pub. Co. Inc. (2005)
24. Liu, C.: Capitalize on dimensionality increasing techniques for improving face recognition grand challenge performance. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28, 725–737 (2006)
25. Cevikalp, H., Neamtu, M., Wilkes, M., Barkana, A.: Discriminative common vectors for face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27, 4–13 (2005)
26. Haasdonk, B., Pekalska, E.: Indefinite kernel discriminant analysis. In: COMPSTAT (2010)
27. Haasdonk, B.: Feature space interpretation of SVMs with indefinite kernels. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27, 482–492 (2005)

28. Ashraf, A., Lucey, S., Chen, T.: Re-interpreting the application of gabor filters as a manipulation of the margin in linear support vector machines. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1335–1341 (2010)
29. Tzimiropoulos, G., Argyriou, V., Zafeiriou, S., Stathaki, T.: Robust fft-based scale-invariant image registration with image gradients. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32, 1899–1906 (2010)
30. Lee, K., Ho, J., Kriegman, D.J.: Acquiring linear subspaces for face recognition under variable lighting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 684–698 (2005)
31. Wagner, A., Wright, J., Ganesh, A., Zhou, Z., Ma, Y.: Toward a practical face recognition system: Robust alignment and illumination by sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34, 372–386 (2012)
32. Phillips, P.J., Moon, H., Rizvi, S.A., Rauss, P.J.: The FERET evaluation methodology for face recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22, 1090–1104 (2000)
33. Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S., Ma, Y.: Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 210–227 (2009)
34. Messer, K., Matas, J., Kittler, J., Luettin, J., Maitre, G.: XM2VTSDB: The extended M2VTS database. In: AVBPA, pp. 72–77 (1999)
35. Ong, C.S.: *Kernels: Regularization and Optimization*. PhD thesis, The Australian National University (2005)
36. Matas, J., et al.: Comparison of face verification results on the XM2VTS database. In: ICPR, pp. 858–863 (2000)
37. Messer, K., et al.: Face Verification Competition on the XM2VTS Database. In: Kittler, J., Nixon, M.S. (eds.) AVBPA 2003. LNCS, vol. 2688, pp. 964–974. Springer, Heidelberg (2003)
38. Messer, K., Kittler, J., Short, J., Heusch, G., Cardinaux, F., Marcel, S., Rodriguez, Y., Shan, S., Su, Y., Gao, W., Chen, X.: Performance Characterisation of Face Recognition Algorithms and Their Sensitivity to Severe Illumination Changes. In: Zhang, D., Jain, A.K. (eds.) ICB 2005. LNCS, vol. 3832, pp. 1–11. Springer, Heidelberg (2005)
39. González-Jiménez, D., Alba-Castro, J.: Shape-driven Gabor jets for face description and authentication. *IEEE Transactions on Information Forensics and Security* 2, 769–780 (2007)