Expanding the Family of Grassmannian Kernels: An Embedding Perspective

Mehrtash T. Harandi, Mathieu Salzmann, Sadeep Jayasumana,
Richard Hartley, and Hongdong Li

Australian National University, Canberra, ACT 0200, Australia
NICTA*, Locked Bag 8001, Canberra, ACT 2601, Australia

Abstract. Modeling videos and image-sets as linear subspaces has proven beneficial for many visual recognition tasks. However, it also incurs challenges arising from the fact that linear subspaces do not obey Euclidean geometry, but lie on a special type of Riemannian manifolds known as Grassmannian. To leverage the techniques developed for Euclidean spaces (*e.g.*, support vector machines) with subspaces, several recent studies have proposed to embed the Grassmannian into a Hilbert space by making use of a positive definite kernel. Unfortunately, only two Grassmannian kernels are known, none of which -as we will show- is *universal*, which limits their ability to approximate a target function arbitrarily well. Here, we introduce several positive definite Grassmannian kernels, including universal ones, and demonstrate their superiority over previously-known kernels in various tasks, such as classification, clustering, sparse coding and hashing.

Keywords: Grassmann manifolds, kernel methods, Plücker embedding

1 Introduction

This paper introduces a set of positive definite kernels to embed Grassmannians (*i.e.*, manifolds of linear subspaces that have a nonlinear Riemannian structure) into Hilbert spaces, which have a more familiar Euclidean structure. Nowadays, linear subspaces are a core representation of many visual recognition techniques. For example, several state-of-the-art video, or image-set, matching methods model the visual data as subspaces [5, 7, 22, 23, 10]. Linear subspaces have also proven a powerful representation for many other computer vision applications, such as chromatic noise filtering [21] and domain adaptation [4].

Despite their success, linear subspaces suffer from the drawback that they cannot be analyzed using Euclidean geometry. Indeed, subspaces lie on a special type of Riemannian manifolds, the Grassmann manifold, which has a nonlinear structure. As a consequence, popular techniques developed for Euclidean

---

spaces do not apply. Recently, this problem has been addressed by embedding the Grassmannian into a Hilbert space. This can be achieved either by tangent space approximation of the manifold, or by exploiting a positive definite kernel function to embed the manifold into a reproducing kernel Hilbert space (RKHS). In either case, any existing Euclidean technique can then be applied to the embedded data, since Hilbert spaces obey Euclidean geometry. Recent studies, however, report superior results with RKHS embedding over flattening the manifold using its tangent spaces [5, 23, 9]. Intuitively, this can be attributed to the fact that a tangent space is a first order approximation to the true geometry of the manifold, whereas, being higher-dimensional, an RKHS has the capacity of better capturing the nonlinearity of the manifold.

While RKHS embeddings therefore seem preferable, their applicability is limited by the fact that only very few positive definite Grassmannian kernels are known. Indeed, in the literature, only two kernels have been introduced to embed Grassmannians into RKHS: the Binet-Cauchy kernel [24] and the projection kernel [5]. The former is a homogeneous second order polynomial kernel, while the latter is a linear kernel. As simple (low-order) polynomial kernels, they are limited in their ability to closely approximate arbitrary functions. In contrast, universal kernels provide much better generalization power [19, 13].

In this paper, we introduce a set of new positive definite Grassmannian kernels, which, among others, includes universal Grassmannian kernels. To this end, we start from the perspective of the two embeddings from which the Binet-Cauchy and the projection kernels are derived: the Plücker embedding and the projection embedding. These two embeddings yield two distance functions. We then exploit the properties of these distances, in conjunction with several theorems analyzing the positive definiteness of kernels, to derive the ten new Grassmannian kernels summarized in Table 1.

Our experimental evaluation demonstrates the benefits of our Grassmannian kernels for classification, clustering, sparse coding and hashing. Our results show that our kernels outperform the Binet-Cauchy and projection ones for gender and gesture recognition, pose categorization and mouse behavior analysis.

2   Background Theory

In this section, we first review some notions of geometry of Grassmannians and then briefly discuss existing positive definite kernels and their properties. Throughout the paper, we use bold capital letters to denote matrices (*e.g.*, $\boldsymbol{X}$) and bold lower-case letters to denote column vectors (*e.g.*, $\boldsymbol{x}$). $\mathbf{I}_p$ is the $p \times p$ identity matrix. $\|\boldsymbol{X}\|_F = \sqrt{\mathrm{Tr}\left(\boldsymbol{X}^T \boldsymbol{X}\right)}$ indicates the Frobenius norm, with $\mathrm{Tr}(\cdot)$ the matrix trace.

2.1   Grassmannian Geometry

The space of p-dimensional linear subspaces of $\mathbb{R}^d$ for $0 < p < d$ is not a Euclidean space, but a Riemannian manifold known as the Grassmannian $\mathcal{G}(d, p)$ [1].

Table 1: The proposed Grassmannian kernels and their properties.

Kernel	Equation	Cond.	Properties
Polynomial	$k_{p,bc}(\boldsymbol{X}, \boldsymbol{Y}) = \left(\beta + \left|\det\left(\boldsymbol{X}^T\boldsymbol{Y}\right)\right|\right)^{\alpha}$	$\beta > 0$	pd
	$k_{p,p}(\boldsymbol{X}, \boldsymbol{Y}) = \left(\beta + \left\|\boldsymbol{X}^T\boldsymbol{Y}\right\|_F^2\right)^{\alpha}$	$\beta > 0$	pd
RBF	$k_{r,bc}(\boldsymbol{X}, \boldsymbol{Y}) = \exp\left(\beta\left|\det\left(\boldsymbol{X}^T\boldsymbol{Y}\right)\right|\right)$	$\beta > 0$	pd, universal
	$k_{r,p}(\boldsymbol{X}, \boldsymbol{Y}) = \exp\left(\beta\left\|\boldsymbol{X}^T\boldsymbol{Y}\right\|_F^2\right)$	$\beta > 0$	pd, universal
Laplace	$k_{l,bc}(\boldsymbol{X}, \boldsymbol{Y}) = \exp\left(-\beta\sqrt{1 - \left|\det(\boldsymbol{X}^T\boldsymbol{Y})\right|}\right)$	$\beta > 0$	pd, universal
	$k_{l,p}(\boldsymbol{X}, \boldsymbol{Y}) = \exp\left(-\beta\sqrt{p - \left\|\boldsymbol{X}^T\boldsymbol{Y}\right\|_F^2}\right)$	$\beta > 0$	pd, universal
Binomial	$k_{bi,bc}(\boldsymbol{X}, \boldsymbol{Y}) = \left(\beta - \left|\det\left(\boldsymbol{X}^T\boldsymbol{Y}\right)\right|\right)^{-\alpha}$	$\beta > 1$	pd, universal
	$k_{bi,p}(\boldsymbol{X}, \boldsymbol{Y}) = \left(\beta - \left\|\boldsymbol{X}^T\boldsymbol{Y}\right\|_F^2\right)^{-\alpha}$	$\beta > p$	pd, universal
Logarithm	$k_{log,bc}(\boldsymbol{X}, \boldsymbol{Y}) = -\log\left(2 - \left|\det\left(\boldsymbol{X}^T\boldsymbol{Y}\right)\right|\right)$	$-$	cpd
	$k_{log,p}(\boldsymbol{X}, \boldsymbol{Y}) = -\log\left(p + 1 - \left\|\boldsymbol{X}^T\boldsymbol{Y}\right\|_F^2\right)$	$-$	cpd

We note that in the special case of $p = 1$, the Grassmann manifold becomes the projective space $\mathbb{P}^{d-1}$, which consists of all lines passing through the origin. A point on the Grassmann manifold $\mathcal{G}(p, d)$ may be specified by an arbitrary $d \times p$ matrix with orthogonal columns, i.e., $\boldsymbol{X} \in \mathcal{G}(d, p) \Rightarrow \boldsymbol{X}^T\boldsymbol{X} = \mathbf{I}_p$[1].

On a Riemannian manifold, points are connected via smooth curves. The distance between two points is defined as the length of the shortest curve connecting them on the manifold. The shortest curve and its length are called geodesic and geodesic distance, respectively. For the Grassmannian, the geodesic distance between two points $\boldsymbol{X}$ and $\boldsymbol{Y}$ is given by

$$\delta_g(\boldsymbol{X}, \boldsymbol{Y}) = \|\Theta\|_2 \,, \tag{1}$$

where Θ is the vector of principal angles between $\boldsymbol{X}$ and $\boldsymbol{Y}$.

Definition 1 (Principal Angles). *Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be two matrices of size $d \times p$ with orthonormal columns. The principal angles $0 \leq \theta_1 \leq \theta_2 \leq \cdots \leq \theta_p \leq \pi/2$ between two subspaces* span($\boldsymbol{X}$) *and* span($\boldsymbol{Y}$) *are defined recursively by*

---

[1] A point on the Grassmannian $\mathcal{G}(p, d)$ is a subspace spanned by the columns of a $d \times p$ full rank matrix and should therefore be denoted by span($\boldsymbol{X}$). With a slight abuse of notation, here we call $\boldsymbol{X}$ a Grassmannian point whenever it represents a basis for a subspace.

$$\cos(\theta_i) = \max_{\boldsymbol{u}_i \in \mathrm{span}(\boldsymbol{X})} \max_{\boldsymbol{v}_i \in \mathrm{span}(\boldsymbol{Y})} \boldsymbol{u}_i^T \boldsymbol{v}_i \tag{2}$$

$$\text{s.t.} \qquad \|\boldsymbol{u}_i\|_2 = \|\boldsymbol{v}_i\|_2 = 1$$
$$\boldsymbol{u}_i^T \boldsymbol{u}_j = 0, \ j = 1, 2, \cdots, i-1$$
$$\boldsymbol{v}_i^T \boldsymbol{v}_j = 0, \ j = 1, 2, \cdots, i-1$$

In other words, the first principal angle θ_1 is the smallest angle between any two unit vectors in the first and the second subspaces. The cosines of the principal angles correspond to the singular values of $\boldsymbol{X}^T \boldsymbol{Y}$ [1]. In addition to the geodesic distance, several other metrics can be employed to measure the similarity between Grassmannian points [5]. In Section 3, we will discuss two other metrics on the Grassmannian.

2.2 Positive Definite Kernels and Grassmannians

As mentioned earlier, a popular way to analyze problems defined on a Grassmannian is to embed the manifold into a Hilbert space using a valid Grassmannian kernel. Let us now formally define Grassmannian kernels:

Definition 2 (Real-valued Positive Definite Kernels). *Let $\mathcal{X}$ be a nonempty set. A symmetric function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is a positive definite (**pd**) kernel on $\mathcal{X}$ if and only if $\sum_{i,j=1}^{n} c_i c_j k(x_i, x_j) \geq 0$ for any $n \in \mathbb{N}$, $x_i \in \mathcal{X}$ and $c_i \in \mathbb{R}$.*

Definition 3 (Grassmannian Kernel). *A function $k : \mathcal{G}(p, d) \times \mathcal{G}(p, d) \to \mathbb{R}$ is a Grassmannian kernel if it is well-defined and pd. In our context, a function is well-defined if it is invariant to the choice of basis, i.e., $k(\boldsymbol{X}\boldsymbol{R}_1, \boldsymbol{Y}\boldsymbol{R}_2) = k(\boldsymbol{X}, \boldsymbol{Y})$, for all $\boldsymbol{X}, \boldsymbol{Y} \in \mathcal{G}(d, p)$ and $\boldsymbol{R}_1, \boldsymbol{R}_2 \in \mathrm{SO}(p)$, where $\mathrm{SO}(p)$ denotes the special orthogonal group.*

The most widely used kernel is arguably the Gaussian or radial basis function (RBF) kernel. It is therefore tempting to define a Radial Basis Grassmannian kernel by replacing the Euclidean distance with the geodesic distance. Unfortunately, although symmetric and well-defined, the function $\exp(-\beta\delta_g^2(\cdot, \cdot))$ is not *pd*. This can be verified by a counter-example using the following points on $\mathcal{G}(3, 2)^2$:

$$\boldsymbol{X}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}, \ \boldsymbol{X}_2 = \begin{bmatrix} -0.0996 & -0.3085 \\ -0.4967 & -0.8084 \\ -0.8622 & 0.5014 \end{bmatrix}, \ \boldsymbol{X}_3 = \begin{bmatrix} -0.9868 & 0.1259 \\ -0.1221 & -0.9916 \\ -0.1065 & -0.0293 \end{bmatrix}, \ \boldsymbol{X}_4 = \begin{bmatrix} 0.1736 & 0.0835 \\ 0.7116 & 0.6782 \\ 0.6808 & -0.7301 \end{bmatrix}.$$

The function $\exp(-\delta_g^2(\cdot, \cdot))$ for these points has a negative eigenvalue of -0.0038.

Nevertheless, two Grassmannian kernels, *i.e.*, the Binet-Cauchy kernel [24] and the projection kernel [5], have been proposed to embed Grassmann manifolds into RKHS. The Binet-Cauchy and projection kernels are defined as

$$k_{bc}^2(\boldsymbol{X}, \boldsymbol{Y}) = \det(\boldsymbol{X}^T \boldsymbol{Y} \boldsymbol{Y}^T \boldsymbol{X}), \tag{3}$$

$$k_p(\boldsymbol{X}, \boldsymbol{Y}) = \|\boldsymbol{X}^T \boldsymbol{Y}\|_F^2. \tag{4}$$

---

[2] Note that we rounded each value to its 4 most significant digits.

Property 1 (Relation to Principal Angles). Both k_p and k_{bc} are closely related to the principal angles between two subspaces. Let θ_i be the ith principal angle between $\boldsymbol{X}, \boldsymbol{Y} \in \mathcal{G}(p, d)$, *i.e.*, by SVD, $\boldsymbol{X}^T\boldsymbol{Y} = \boldsymbol{U}\boldsymbol{\Gamma}\boldsymbol{V}^T$, with Γ a diagonal matrix with elements $\cos\theta_i$. Then

$$k_p(\boldsymbol{X}, \boldsymbol{Y}) = \left\|\boldsymbol{X}^T\boldsymbol{Y}\right\|_F^2 = \mathrm{Tr}\left(\boldsymbol{U}\boldsymbol{\Gamma}\boldsymbol{V}^T\boldsymbol{V}\boldsymbol{\Gamma}\boldsymbol{U}^T\right) = \mathrm{Tr}\left(\Gamma^2\right) = \sum_{i=1}^{p} \cos^2(\theta_i) \ .$$

Similarly, one can show that $k_{bc}^2(\boldsymbol{X}, \boldsymbol{Y}) = \prod_{i=1}^{p} \cos^2(\theta_i)$.

3   Embedding Grassmannians to Hilbert Spaces

While k_p and k_{bc}^2 have been successfully employed to transform problems on Grassmannians to Hilbert spaces [5, 7, 23], the resulting Hilbert spaces themselves have received comparatively little attention. In this section, we aim to bridge this gap and study these two spaces, which can be explicitly computed. To this end, we discuss the two embeddings that define these Hilbert spaces, namely the Plücker embedding and the projection embedding. These embeddings, and their respective properties, will in turn help us devise our set of new Grassmannian kernels.

3.1   Plücker Embedding

To study the Plücker embedding, we first need to review some concepts of exterior algebra.

Definition 4 (Alternating Multilinear Map). *Let $\boldsymbol{V}$ and $\boldsymbol{W}$ be two vector spaces. A map $g : \underbrace{\boldsymbol{V} \times \cdots \times \boldsymbol{V}}_{k \ copies} \to \boldsymbol{W}$ is multilinear if it is linear in each slot, that is if*

$$g(\boldsymbol{v}_1, \cdots, \lambda\boldsymbol{v}_i + \lambda'\boldsymbol{v}_i', \cdots, \boldsymbol{v}_k) = \lambda g(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_i, \cdots, \boldsymbol{v}_k) + \lambda' g(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_i', \cdots, \boldsymbol{v}_k) \ .$$

Furthermore, the map g is alternating if, whenever two of the inputs to g are the same vector, the output is 0. That is, if $g(\cdots, \boldsymbol{v}, \cdots, \boldsymbol{v}, \cdots) = 0, \ \forall \boldsymbol{v}$.

Definition 5 (k^{th} Exterior Product). *Let $\boldsymbol{V}$ be a vector space. The kth exterior product of $\boldsymbol{V}$, denoted by $\bigwedge^k \boldsymbol{V}$ is a vector space, equipped with an alternating multilinear map $g : \underbrace{\boldsymbol{V} \times \cdots \times \boldsymbol{V}}_{k \ copies} \to \bigwedge^k \boldsymbol{V}$ of the form $g(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k) = \boldsymbol{v}_1 \wedge \cdots \wedge \boldsymbol{v}_k$, with $\wedge$ the wedge product.*

The wedge product is supercommutative and can be thought of as a generalization of the cross product in $\mathbb{R}^3$ to an arbitrary dimension. Importantly, note that the kth exterior product $\bigwedge^k \boldsymbol{V}$ is a vector space, that is

$$\bigwedge^k \boldsymbol{V} = \mathrm{span}\left(\{\boldsymbol{v}_1 \wedge \boldsymbol{v}_2 \wedge \cdots \wedge \boldsymbol{v}_k\}\right), \ \forall \boldsymbol{v}_i \in \boldsymbol{V} \ .$$

The Grassmannian $\mathcal{G}(d, p)$ can be embedded into the projective space $\mathbb{P}(\bigwedge^p \mathbb{R}^d)$ as follows. Let $\boldsymbol{X}$ be a point on $\mathcal{G}(p, d)$ described by the basis $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_p\}$, *i.e.*, $\boldsymbol{X} = \text{span}(\{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_p\})$. The Plücker map of $\boldsymbol{X}$ is given by:

Definition 6 (Plücker Embedding). *The Plücker embedding* $P : \mathcal{G}(p, d) \to \mathbb{P}(\bigwedge^p \mathbb{R}^d)$ *is defined as*

$$P(\boldsymbol{X}) = [\boldsymbol{x}_1 \wedge \boldsymbol{x}_2 \wedge \cdots \wedge \boldsymbol{x}_p], \tag{5}$$

where $\boldsymbol{X}$ *is the subspace spanned by* $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_p\}$.

Example 1. Consider the space of two-dimensional planes in $\mathbb{R}^4$, *i.e.*, $\mathcal{G}(2, 4)$. In this space, an arbitrary subspace is described by its basis $\mathbf{B} = [\boldsymbol{w}_1 | \boldsymbol{w}_2]$. Let $\boldsymbol{e}_i$ be the unit vector along the ith axis. We can write $\boldsymbol{w}_j = \sum_{i=1}^4 a_{j,i} \boldsymbol{e}_i$. Then

$$
\begin{aligned}
P(\mathbf{B}) &= \Big(\sum_{i=1}^4 a_{1,i} \boldsymbol{e}_i \Big) \wedge \Big(\sum_{i=1}^4 a_{2,j} \boldsymbol{e}_j \Big) \\
&= (a_{1,1} a_{2,2} - a_{1,2} a_{2,1})(\boldsymbol{e}_1 \wedge \boldsymbol{e}_2) + (a_{1,1} a_{2,3} - a_{1,3} a_{2,1})(\boldsymbol{e}_1 \wedge \boldsymbol{e}_3) \\
&+ (a_{1,1} a_{2,4} - a_{1,4} a_{2,1})(\boldsymbol{e}_1 \wedge \boldsymbol{e}_4) + (a_{1,2} a_{2,3} - a_{1,3} a_{2,2})(\boldsymbol{e}_2 \wedge \boldsymbol{e}_3) \\
&+ (a_{1,2} a_{2,4} - a_{1,4} a_{2,2})(\boldsymbol{e}_2 \wedge \boldsymbol{e}_4) + (a_{1,3} a_{2,4} - a_{1,4} a_{2,3})(\boldsymbol{e}_3 \wedge \boldsymbol{e}_4) .
\end{aligned}
$$

Hence, the Plücker embedding of $\mathcal{G}(2, 4)$ is a 6-dimensional space spanned by $\{\boldsymbol{e}_1 \wedge \boldsymbol{e}_2, \boldsymbol{e}_1 \wedge \boldsymbol{e}_3, \cdots, \boldsymbol{e}_3 \wedge \boldsymbol{e}_4\}$. A closer look at the coordinates of the embedded subspace reveals that they are indeed the minors of all possible 2×2 submatrices of $\mathbf{B}$. This can be shown to hold for any d and p.

Proposition 1. *The Plücker coordinates of* $\boldsymbol{X} \in \mathcal{G}(d, p)$ *are the* $p \times p$ *minors of the matrix* $\boldsymbol{X}$ *obtained by taking* p *rows out of the* d *possible ones.*

Remark 1. The space induced by the Plücker map of $\mathcal{G}(p, d)$ is $\binom{p}{d}$-dimensional.

To be able to exploit the Plücker embedding to design new kernels, we need to define an inner product over $\mathbb{P}(\bigwedge^p \mathbb{R}^d)$. Importantly, to be meaningful, this inner product needs to be invariant to the specific realization of a point on $\mathcal{G}(p, d)$ (recall that, *e.g.*, swapping two columns of a specific realization $\boldsymbol{X} \in \mathcal{G}(d, p)$ still corresponds to the same point on $\mathcal{G}(d, p)$). Furthermore, we would also like this inner product to be efficient to evaluate, thus avoiding the need to explicitly compute the high-dimensional embedding. Note in particular that, for vision applications, the dimensionality of $\mathbb{P}(\bigwedge^p \mathbb{R}^d)$ becomes overwhelming and hence explicitly computing the embedding is impractical. To achieve these goals, we rely on the following definition and theorem:

Definition 7 (Compound Matrices). *Given a* $d \times p$ *matrix* $\boldsymbol{A}$, *the matrix whose elements are the minors of* $\boldsymbol{A}$ *of order* q *arranged in a lexicographic order is called the* q^{th} *compound of* $\boldsymbol{A}$, *and is denoted by* $\boldsymbol{C}_q(\boldsymbol{A})$.

Theorem 1 (Binet-Cauchy Theorem). *Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be two rectangular matrices of size $d \times p_1$ and $d \times p_2$, respectively. Then, $\boldsymbol{C}_q(\boldsymbol{A}^T\boldsymbol{B}) = \boldsymbol{C}_q(\boldsymbol{A})^T\boldsymbol{C}_q(\boldsymbol{B})$.*

Therefore, for $\boldsymbol{X}, \boldsymbol{Y} \in \mathbb{R}^{d \times p}$, we have $\mathrm{Tr}\left(\boldsymbol{C}_p(\boldsymbol{X})^T\boldsymbol{C}_p(\boldsymbol{Y})\right) = \mathrm{Tr}\left(\boldsymbol{C}_p(\boldsymbol{X}^T\boldsymbol{Y})\right) = \det\left(\boldsymbol{X}^T\boldsymbol{Y}\right)$.

Since, for $\boldsymbol{X} \in \mathcal{G}(p,d)$, $\boldsymbol{C}_p(\boldsymbol{X})$ stores all $p \times p$ minors and hence conveys the Plücker coordinates of $\boldsymbol{X}$, this would suggest defining the inner product for the Plücker embedding as $\det(\boldsymbol{X}^T\boldsymbol{Y})$. This is indeed what was proposed in [5, 24] where $\det(\cdot)$ was used as a linear kernel. However, while $\det(\boldsymbol{X}^T\boldsymbol{Y})$ is invariant to the action of $\mathrm{SO}(p)$, it is not invariant to the specific realization of a subspace. This can be simply verified by permuting the columns of $\boldsymbol{X}$, which does not change the subspace, but may change the sign of $\det(\cdot)$. Note that this sign issue was also observed by Wolf *et al.* [24]. However, this problem was circumvented by considering the second-order polynomial kernel k_{bc}^2.

In contrast, here, we focus on designing a valid inner product that satisfies this invariance condition. To this end, we define the inner product in $\mathbb{P}(\bigwedge^p \mathbb{R}^d)$ as $\langle \boldsymbol{X}, \boldsymbol{Y} \rangle_P = |P(\boldsymbol{X})^T P(\boldsymbol{Y})| = \left|\det\left(\boldsymbol{X}^T\boldsymbol{Y}\right)\right|$. This inner product induces the distance

$$\delta_{bc}^2(\boldsymbol{X}, \boldsymbol{Y}) = \|P(\boldsymbol{X}) - P(\boldsymbol{Y})\|^2 = 2 - 2\left|\det\left(\boldsymbol{X}^T\boldsymbol{Y}\right)\right|. \tag{6}$$

Clearly, if $\{\theta_i\}_{i=1}^p$ is the set of principal angles between two Grassmannian points $\boldsymbol{X}$ and $\boldsymbol{Y}$, then $\langle \boldsymbol{X}, \boldsymbol{Y} \rangle_P = \prod_{i=1}^p \cos(\theta_i)$, which is invariant to the specific realization of a subspace since $0 \leq \theta_i \leq \pi/2$.

In the following, we show that the Plücker embedding has the nice property of being closely related to the true geometry of the corresponding Grassmannian:

Theorem 2 (Curve Length Equivalence). *The length of any given curve is the same under δ_{bc} and δ_g up to a scale of $\sqrt{2}$.*

Proof. Given in supplementary material. □

3.2 Projection Embedding

We now turn to the case of the projection embedding. Note that this embedding has been better studied than the Plücker one [8].

Definition 8 (Projection Embedding). *The projection embedding $\Pi : \mathcal{G}(p,d) \to \mathrm{Sym(d)}$ is defined as*

$$\Pi(\boldsymbol{X}) = \boldsymbol{X}\boldsymbol{X}^T. \tag{7}$$

The projection embedding $\Pi(\cdot)$ is a diffeomorphism from a Grassmann manifold onto the idempotent symmetric matrices of rank p, *i.e.*, it is a one-to-one, continuous, differentiable mapping with a continuous, differentiable inverse [3]. The space induced by this embedding is a smooth, compact submanifold of $\mathrm{Sym(d)}$ of dimension $d(d-p)$. Since $\Pi(\boldsymbol{X})$ is a symmetric $d \times d$ matrix, a natural choice of inner product is $\langle \boldsymbol{X}, \boldsymbol{Y} \rangle_\Pi = \mathrm{Tr}\left(\Pi(\boldsymbol{X})^T \Pi(\boldsymbol{Y})\right) = \left\|\boldsymbol{X}\boldsymbol{Y}\right\|_F^2$.

This inner product can be shown to be invariant to the specific realization of a subspace, and induces the distance

$$\delta_p^2(\boldsymbol{X}, \boldsymbol{Y}) = \left\| \Pi(\boldsymbol{X}) - \Pi(\boldsymbol{Y}) \right\|_F^2 = 2p - 2\left\| \boldsymbol{X}^T \boldsymbol{Y} \right\|_F^2 .$$

Due to space limitation, we do not discuss the properties of the projection embedding, such as isometry [3] and length of curves [6]. We refer the reader to [8] for a more thorough discussion of the projection embedding.

4   Grassmannian Kernels

From the discussion in Section 3, k_{bc}^2 and k_p, defined in Eq. 4 and Eq. 3, can be seen to correspond to a homogeneous second order polynomial kernel in the space induced by the Plücker embedding and to a linear kernel in the space induced by the projection embedding, respectively. In this section, we show that the inner products that we defined in Section 3 for the Plücker and projection embeddings can actually be exploited to derive many new Grassmannian kernels, including universal kernels and conditionally positive definite kernels. In the following, we denote by $k_{\cdot,bc}$ and $k_{\cdot,p}$ kernels derived from the Plücker embedding (Binet-Cauchy kernels) and from the projection embedding, respectively.

4.1   Polynomial Kernels

Given an inner product, which itself defines a valid linear kernel, the most straightforward way to create new kernels is to consider higher degree polynomials. Such polynomial kernels are known to be *pd*. Therefore, we can readily define polynomial kernels on the Grassmannian as

$$k_{p,bc}(\boldsymbol{X}, \boldsymbol{Y}) = \left(\beta + \left| \det \left(\boldsymbol{X}^T \boldsymbol{Y} \right) \right| \right)^\alpha , \quad \beta > 0 , \tag{8}$$

$$k_{p,p}(\boldsymbol{X}, \boldsymbol{Y}) = \left(\beta + \left\| \boldsymbol{X}^T \boldsymbol{Y} \right\|_F^2 \right)^\alpha , \quad \beta > 0 . \tag{9}$$

Note that the kernel used in [24] is indeed the homogeneous second order $k_{p,bc}$ with $\alpha = 2$ and $\beta = 0$.

4.2   Universal Grassmannian Kernels

Although often used in practice, polynomial kernels are known not to be universal [19]. This can have a crucial impact on their representation power for a specific task. Indeed, from the *Representer Theorem* [16], we have that, for a given set of training data $\{x_j\}$, $j \in \mathbb{N}_n$, $\mathbb{N}_n = \{1, 2, \cdots, n\}$ and a *pd* kernel k, the function learned by any algorithm can be expressed as

$$\hat{f}(x_*) = \sum_{j \in \mathbb{N}_n} c_j k(x_*, x_j) . \tag{10}$$

Importantly, only *universal kernels* have the property of being able to approximate any target function f_t arbitrarily well given sufficiently many training samples. Therefore, k_p and k_{bc}^2 may not generalize sufficiently well for certain problems. In the following, we develop several universal Grassmannian kernels. To this end, we make use of negative definite kernels and of their relation to *pd* ones. Let us first formally define negative definite kernels.

Definition 9 (Real-valued Negative Definite Kernels). *Let $\mathcal{X}$ be a nonempty set. A symmetric function $\psi : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is a negative definite (**nd**) kernel on $\mathcal{X}$ if and only if $\sum_{i,j=1}^{n} c_i c_j k(x_i, x_j) \leq 0$ for any $n \in \mathbb{N}$, $x_i \in \mathcal{X}$ and $c_i \in \mathbb{R}$ with $\sum_{i=1}^{n} c_i = 0$.*

Note that, in contrast to positive definite kernels, an additional constraint of the form $\sum c_i = 0$ is required in the negative definite case.

The most important example of *nd* kernels is the distance function defined on a Hilbert space. More specifically:

Theorem 3 ([9]). *Let $\mathcal{X}$ be a nonempty set, $\mathcal{H}$ be an inner product space, and $\psi : \mathcal{X} \to \mathcal{H}$ be a function. Then $f : (\mathcal{X} \times \mathcal{X}) \to \mathbb{R}$ defined by $f(x_i, x_j) = \|\psi(x_i) - \psi(x_j)\|_{\mathcal{H}}^2$ is negative definite.*

Therefore, being distances in Hilbert spaces, both δ_{bc}^2 and δ_p^2 are *nd* kernels. We now state an important theorem which establishes the relation between *pd* and *nd* kernels.

Theorem 4 (Theorem 2.3 in Chapter 3 of [2]). *Let μ be a probability measure on the half line $\mathbb{R}_+$ and $0 < \int_0^\infty t\,d\mu(t) < \infty$. Let $\mathcal{L}_\mu$ be the Laplace transform of μ, i.e., $\mathcal{L}_\mu(s) = \int_0^\infty e^{-ts}\,d\mu(t)$, $s \in \mathbb{C}_+$. Then, $\mathcal{L}_\mu(\beta f)$ is positive definite for all $\beta > 0$ if and only if $f : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_+$ is negative definite.*

The problem of designing a *pd* kernel on the Grassmannian can now be cast as that of finding an appropriate probability measure μ. Below, we show that this lets us reformulate popular kernels in Euclidean space as Grassmannian kernels.

RBF Kernels. Grassmannian RBF kernels can be obtained by choosing $\mu(t) = \delta(t-1)$ in Theorem 4, where $\delta(t)$ is the Dirac delta function. This choice yields the Grassmannian RBF kernels (after discarding scalar constants)

$$k_{r,bc}(\boldsymbol{X}, \boldsymbol{Y}) = \exp\left(\beta \left| \det\left(\boldsymbol{X}^T \boldsymbol{Y}\right)\right|\right), \quad \beta > 0, \tag{11}$$

$$k_{r,p}(\boldsymbol{X}, \boldsymbol{Y}) = \exp\left(\beta \left\|\boldsymbol{X}^T \boldsymbol{Y}\right\|_F^2\right), \quad \beta > 0. \tag{12}$$

Note that the RBF kernel obtained from the projection embedding, *i.e.* $k_{r,p}$, was also used by Vemulapalli *et al.* [23]. However, the positive definiteness of this kernel was neither proven nor discussed.

Laplace Kernels. The Laplace kernel is another widely used Euclidean kernel, defined as $k(\boldsymbol{x}, \boldsymbol{y}) = \exp(-\beta\|\boldsymbol{x} - \boldsymbol{y}\|)$. To obtain Laplace kernels on the Grassmannian, we make use of the following theorem for nd kernels.

Theorem 5 (Corollary 2.10 in Chapter 3 of [2]). *If $\psi : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is negative definite and satisfies $\psi(\boldsymbol{x}, \boldsymbol{x}) \geqq 0$ then so is ψ^α for $0 < \alpha < 1$.*

As a result, both $\delta_p(\cdot, \cdot)$ and $\delta_{bc}(\cdot, \cdot)$ are nd by choosing $\alpha = 1/2$ in Theorem 5. By employing either $\delta_p(\cdot, \cdot)$ or $\delta_{bc}(\cdot, \cdot)$ along with $\mu(t) = \delta(t - 1)$ in Theorem 4, we obtain the Grassmannian Laplace kernels

$$k_{l,bc}(\boldsymbol{X}, \boldsymbol{Y}) = \exp\left(-\beta\sqrt{1 - |\det(\boldsymbol{X}^T\boldsymbol{Y})|} \right), \quad \beta > 0, \tag{13}$$

$$k_{l,p}(\boldsymbol{X}, \boldsymbol{Y}) = \exp\left(-\beta\sqrt{p - \|\boldsymbol{X}^T\boldsymbol{Y}\|_F^2} \right), \quad \beta > 0. \tag{14}$$

As shown in [19], the RBF and Laplace kernels are universal for $\mathbb{R}^d, d > 0$. Since the Plücker and projection embeddings map to Euclidean spaces, this property clearly extends to the Grassmannian RBF and Laplace kernels.

Binomial Kernels. By choosing $\mu(t) = \exp(-\beta_0 t)u(t)$, where $u(t)$ is the unit (or Heaviside) step function, *i.e.*, $u(t) = \int_{-\infty}^t \delta(x)\mathrm{d}x$, we obtain the Grassmannian binomial kernels

$$k_{b,bc}(\boldsymbol{X}, \boldsymbol{Y}) = \frac{1}{\beta - |\det(\boldsymbol{X}^T\boldsymbol{Y})|}, \quad \beta > 1, \tag{15}$$

$$k_{b,p}(\boldsymbol{X}, \boldsymbol{Y}) = \frac{1}{\beta - \|\boldsymbol{X}^T\boldsymbol{Y}\|_F^2}, \quad \beta > p. \tag{16}$$

Note that the generating function μ is a valid measure only for $\beta_0 > 0$. This translates into the constraints on β given in Eq. 15 and Eq. 16.

A more general form of binomial kernels can be derived by noting that, if $k(\cdot, \cdot) : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_+$ is pd, then so is $k^\alpha(\cdot, \cdot)$, $\alpha > 0$ (see Proposition 2.7 in Chapter 3 of [2]). This lets us define the Grassmannian kernels

$$k_{bi,bc}(\boldsymbol{X}, \boldsymbol{Y}) = \left(\beta - |\det(\boldsymbol{X}^T\boldsymbol{Y})| \right)^{-\alpha}, \quad \beta > 1, \ \alpha > 0, \tag{17}$$

$$k_{bi,p}(\boldsymbol{X}, \boldsymbol{Y}) = \left(\beta - \|\boldsymbol{X}^T\boldsymbol{Y}\|_F^2 \right)^{-\alpha}, \quad \beta > p, \ \alpha > 0. \tag{18}$$

To show that the binomial kernels are universal, we note that

$$(1 - t)^{-\alpha} = \sum_{j=0}^{\infty} \binom{-\alpha}{j}(-1)^j t^j, \quad \text{with} \quad \binom{\alpha}{j} = \prod_{i=1}^{j}(\alpha - i + 1)/i.$$

It can be seen that $\binom{-\alpha}{j}(-1)^j > 0$, which implies that both $k_{bi,bc}$ and $k_{bi,p}$ have non-negative and full Taylor series. This, as was shown in Corollary 4.57 of [19], is a necessary and sufficient condition for a kernel to be universal.

4.3   Conditionally Positive Kernels

Another important class of kernels is the so-called conditionally positive definite kernels [2]. Formally:

Definition 10 (Conditionally Positive Definite Kernels). *Let $\mathcal{X}$ be a nonempty set. A symmetric function $\psi : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is a conditionally positive definite (**cpd**) kernel on $\mathcal{X}$ if and only if $\sum_{i,j=1}^{n} c_i c_j k(x_i, x_j) \geq 0$ for any $n \in \mathbb{N}$, $x_i \in \mathcal{X}$ and $c_i \in \mathbb{R}$ with $\sum_{i=1}^{n} c_i = 0$.*

The relations between *cpd* kernels and *pd* ones were studied by Berg *et al.* [2] and Schölkopf [15] among others. Before introducing *cpd* kernels on the Grassmannian, we state an important property of *cpd* kernels.

Proposition 2 ([15]). *For a kernel algorithm that is translation invariant, one can equally use* cpd *kernels instead of* pd *ones.*

This property relaxes the requirement of having *pd* kernels for certain types of kernel algorithms. A kernel algorithm is translation invariant if it is independent of the position of the origin. For example, in SVMs, maximizing the margin of the separating hyperplane between two classes is independent of the position of the origin. As a result, one can seamlessly use a *cpd* kernel instead of a *pd* kernel in SVMs. To introduce *cpd* kernels on Grassmannians, we rely on the following proposition:

Proposition 3 ([2]). *If $f : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_+$ is* nd *then* $-\log(1 + f)$ *is* cpd.

This lets us derive the Grassmannian *cpd* kernels

$$k_{log,bc}(\boldsymbol{X}, \boldsymbol{Y}) = -\log\left(2 - \det\left(\boldsymbol{X}^T\boldsymbol{Y}\right)\right), \tag{19}$$

$$k_{log,p}(\boldsymbol{X}, \boldsymbol{Y}) = -\log\left(p + 1 - \left\|\boldsymbol{X}^T\boldsymbol{Y}\right\|_F^2\right). \tag{20}$$

The ten new kernels derived in this section are summarized in Table 1. Note that given the linear *Plücker* and *projection* kernels, *i.e.*, $k_{lin,bc}(\boldsymbol{X}, \boldsymbol{Y}) = |\det(\boldsymbol{X}^T\boldsymbol{Y})|$ and $k_{lin,p}(\boldsymbol{X}, \boldsymbol{Y}) = \|\boldsymbol{X}^T\boldsymbol{Y}\|_F^2$, it is possible to obtain the polynomial and Gaussian extensions via standard kernel construction rules [17]. However, our approach lets us derive many other kernels in a principled manner by, *e.g.*, exploiting different measures in Theorem 4. Nonetheless, here, we confined ourselves to deriving kernels corresponding to the most popular ones in Euclidean space, and leave the study of additional kernels as future work.

5   Experimental Evaluation

In this section, we compare our new kernels with the baseline kernels k_{bc}^2 and k_p using three different kernel-based algorithms on Grassmannians: kernel SVM, kernel k-means and kernelized Locality Sensitive Hashing (kLSH). Additional
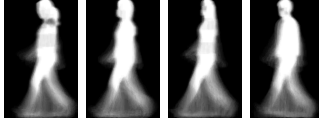
Fig. 1: GEI samples from CASIA [26].

Table 2: **Gender recognition.** Accuracies on the CASIA gait dataset [26].

kernel	k_{bc}^2	$k_{p,bc}$	$k_{r,bc}$	$k_{l,bc}$	$k_{bi,bc}$	$k_{log,bc}$
SVM	$76.8\% \pm 9.1$	$84.1\% \pm 7.2$	$85.8\% \pm 4.6$	$84.5\% \pm 4.5$	$\mathbf{86.4\%} \pm \mathbf{4.4}$	$82.7\% \pm 7.4$
GGDA[7]	$83.7\% \pm 3.7$.	$89.0\% \pm 3.7$	$88.3\% \pm 3.6$	$88.0\% \pm 3.6$	$\mathbf{89.4\%} \pm \mathbf{3.1}$	$84.9\% \pm 3.5$

kernel	k_p	$k_{p,p}$	$k_{r,p}$	$k_{l,p}$	$k_{bi,p}$	$k_{log,p}$
SVM	$83.7\% \pm 4.3$	$\mathbf{89.3\%} \pm \mathbf{5.8}$	$88.2\% \pm 5.8$	$87.6\% \pm 5.5$	$88.7\% \pm 5.1$	$85.8\% \pm 8.3$
GGDA[7]	$90.3\% \pm 4.7$	$\mathbf{93.5\%} \pm \mathbf{2.7}$	$91.3\% \pm 3.8$	$91.0\% \pm 3.8$	$91.1\% \pm 3.1$	$89.7\% \pm 3.6$

results using kernel sparse coding are given in supplementary material. In our experiments, unless stated otherwise, we obtained the kernel parameters (*i.e.*, β for all kernels except the logarithm ones and α for the polynomial and binomial cases) by cross-validation.

5.1   Gender Recognition from Gait

We first demonstrate the benefits of our kernels on a binary classification problem on the Grassmannian using SVM and the Grassmannian Graph-embedding Discriminant Analysis (GGDA) proposed in [7]. To this end, we consider the task of gender recognition from gait (*i.e.*, videos of people walking). We used Dataset-B of the CASIA gait database [26], which comprises 124 individuals (93 males and 31 females). The gait of each subject was captured from 11 viewpoints. Every video is represented by a gait energy image (GEI) of size 32×32 (see Fig. 1), which has proven effective for gender recognition [25].

In our experiment, we used the videos captured with normal clothes and created a subspace of order 3 from the 11 GEIs corresponding to the different viewpoints. This resulted in 731 points on $\mathcal{G}(3, 1024)$. We then randomly selected 20 individuals (10 male, 10 female) as training set and used the remaining individuals for testing. In Table 2, we report the average accuracies over 10 random partitions. Note that for the SVM classifier, all new kernels derived from the Plücker embedding outperform k_{bc}^2, with highest accuracy obtained with the binomial kernel. Similarly, all new projection kernels outperform k_p, and the polynomial kernel achieves the overall highest accuracy of 89.3%. For GGDA, bar the case of $k_{log,p}$, all new kernels also outperform previously-known ones.

5.2   Pose Categorization

As a second experiment, we evaluate the performance of our kernels on the task of clustering on the Grassmannian using kernel k-means. To this end, we used

Fig. 2: Sample images from CMU-PIE.

Fig. 3: Sample images from the mouse behavior dataset [11].

Table 3: **Pose categorization.** Clustering accuracies on the CMU-PIE dataset.

kernel	k_{bc}^2	$k_{p,bc}$	$k_{r,bc}$	$k_{l,bc}$	$k_{bi,bc}$	$k_{log,bc}$
accuracy	70.3%	72.2%	**78.9%**	74.8%	78.5%	72.2%
NMI	0.763	0.779	**0.803**	0.786	0.798	0.772
kernel	k_p	$k_{p,p}$	$k_{r,p}$	$k_{l,p}$	$k_{bi,p}$	$k_{log,p}$
accuracy	77.1%	79.9%	80.9%	79.8%	**82.9%**	74.4%
NMI	0.810%	0.817	0.847	0.843	**0.853**	0.812

the CMU-PIE face dataset [18], which contains images of 67 subjects with 13 different poses and 21 different illuminations (see Fig. 2 for examples). From each image, we computed a 2×2 spatial pyramid of LBP [14] histograms and concatenated them to form a 232 dimensional descriptor. For each subject, we collected the images acquired with the same pose, but different illuminations, in an image set, which we then represented as a linear subspace of order 3. This resulted in a total of $67 \times 13 = 871$ Grassmannian points on $\mathcal{G}(3, 232)$. We used 10 samples from each pose to compute the kernel parameters.

The goal here is to cluster together image sets representing the same pose. To evaluate the quality of the clusters, we report both the clustering accuracy and the Normalized Mutual Information (NMI) [20], which measures the amount of statistical information shared by random variables representing the cluster distribution and the underlying class distribution of the data points. From the results given in Table 3, we can see that, with the exception of $k_{log,p}$, the new kernels in each embedding outperform their respective baseline, k_p or k_{bc}^2. For the Binet-Cauchy kernels, the maximum accuracy (and NMI score) is reached by the RBF kernel. The overall maximum accuracy of 82.9% is achieved by the projection-based binomial kernel.

We also evaluated the intrinsic k-means algorithm of [22]. This algorithm achieved 67.7% accuracy and an NMI score of 0.75. Furthermore, intrinsic k-means required $9766s$ to perform clustering on an i7 machine using Matlab. On the same machine, the runtimes for kernel k-means using $k_{r,bc}$ and $k_{bi,p}$ (which achieve the highest accuracies in Table 3) were $3.1s$ and $2.5s$, respectively. This clearly demonstrates the benefits of RKHS embedding to tackle clustering problems on the Grassmannian.
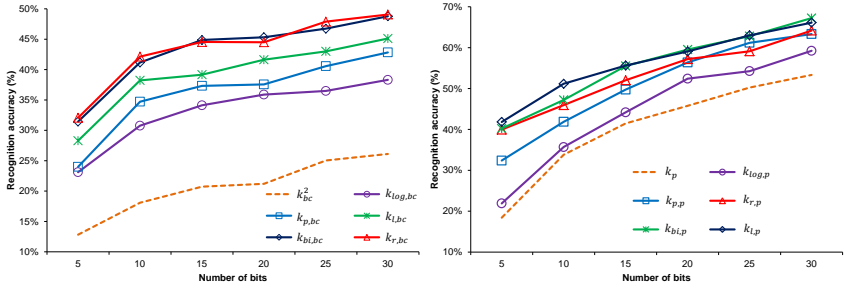
Fig. 4: **Video hashing.** Approximate nearest-neighbor accuracies for kernels derived from **(left)** the Plücker and **(right)** the projection embeddings.

5.3   Mouse Behavior Analysis

Finally, we utilized kernelized Locality-Sensitive Hashing (kLHS) [12] to perform recognition on the 2000 videos of the mice behavior dataset [11]. The basic idea of kLSH is to search for a projection from an RKHS to a low-dimensional Hamming space, where each sample is encoded with a b-bit vector called the hash key. The approximate nearest-neighbor to a query can then be found efficiently in time sublinear in the number of training samples.

The mice dataset [11] contains 8 behaviors (*i.e.*, *drinking*, *eating*, *grooming*, *hanging*, *rearing*, *walking*, *resting* and *micro-movement of head*) of several mice with different coating colors, sizes and genders (see Fig. 3 for examples). In each video, we estimated the background to extract the region containing the mouse in each frame. These regions were then resized to 48×48, and the video represented with an order 6 subspace, thus yielding points on $\mathcal{G}(6, 2304)$. We randomly chose 1000 videos for training and used the remaining 1000 videos for testing. We report the average recognition accuracy over 10 random partitions.

Fig. 4 depicts the recognition accuracies of the new and baseline kernels as a function of the number of bits b. For the Plücker embedding kernels, the gap between our RBF kernel and k_{bc}^2 reaches 23% for a hash key of size 30. For the same hash key size, the projection-based heat kernel outperforms k_p by more than 14%, and thus reaches the overall highest accuracy of 67.2%.

6   Conclusions and Future Work

We have introduced a set of new positive definite kernels to embed Grassmannian into Hilbert spaces, which have a more familiar Euclidean structure. This set includes, among others, universal Grassmannian kernels, which have the ability to approximate general functions. Our experiments have demonstrated the superiority of such kernels over previously-known Grassmannian kernels, *i.e.*, the Binet-Cauchy kernel [24] and the projection kernel [5]. It is important to keep in mind, however, that choosing the right kernel for the data at hand remains an open problem. In the future, we intend to study if searching for the best probability measure in Theroem 4 could give a partial answer to this question.

References

1. Absil, P.A., Mahony, R., Sepulchre, R.: Optimization Algorithms on Matrix Manifolds. Princeton University Press, Princeton, NJ, USA (2008)
2. Berg, C., Christensen, J.P.R., Ressel, P.: Harmonic Analysis on Semigroups. Springer (1984)
3. Chikuse, Y.: Statistics on Special Manifolds. Springer (2003)
4. Gopalan, R., Li, R., Chellappa, R.: Unsupervised adaptation across domain shifts by generating intermediate data representations. IEEE Trans. Pattern Analysis and Machine Intelligence (2014)
5. Hamm, J., Lee, D.D.: Grassmann discriminant analysis: a unifying view on subspace-based learning. In: Proc. Int. Conference on Machine Learning (ICML). pp. 376–383 (2008)
6. Harandi, M., Sanderson, C., Shen, C., Lovell, B.C.: Dictionary learning and sparse coding on grassmann manifolds: An extrinsic solution. In: Proc. Int. Conference on Computer Vision (ICCV) (December 2013)
7. Harandi, M.T., Sanderson, C., Shirazi, S., Lovell, B.C.: Graph embedding discriminant analysis on Grassmannian manifolds for improved image set matching. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2705–2712 (2011)
8. Helmke, U., Hüper, K., Trumpf, J.: Newtons's method on Grassmann manifolds. Preprint: [arXiv:0709.2205] (2007)
9. Jayasumana, S., Hartley, R., Salzmann, M., Li, H., Harandi, M.: Kernel methods on the Riemannian manifold of symmetric positive definite matrices. In: CVPR. pp. 73–80 (2013)
10. Jayasumana, S., Hartley, R., Salzmann, M., Li, H., Harandi, M.: Optimizing over radial kernels on compact manifolds. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2014)
11. Jhuang, H., Garrote, E., Yu, X., Khilnani, V., Poggio, T., Steele, A.D., Serre, T.: Automated home-cage behavioural phenotyping of mice. Nature communications 1, 68 (2010)
12. Kulis, B., Grauman, K.: Kernelized locality-sensitive hashing. IEEE Trans. Pattern Analysis and Machine Intelligence 34(6), 1092–1104 (2012)
13. Micchelli, C.A., Xu, Y., Zhang, H.: Universal kernels. Journal of Machine Learning Research 7, 2651–2667 (2006)
14. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Analysis and Machine Intelligence 24(7), 971–987 (2002)
15. Scholkopf, B.: The kernel trick for distances. In: Proc. Advances in Neural Information Processing Systems (NIPS). pp. 301–307 (2001)
16. Schölkopf, B., Herbrich, R., Smola, A.J.: A generalized representer theorem. In: Computational learning theory. pp. 416–426. Springer (2001)
17. Shawe-Taylor, J., Cristianini, N.: Kernel Methods for Pattern Analysis. Cambridge University Press (2004)
18. Sim, T., Baker, S., Bsat, M.: The cmu pose, illumination, and expression database. IEEE Trans. Pattern Analysis and Machine Intelligence 25(12), 1615–1618 (2003)
19. Steinwart, I., Christmann, A.: Support vector machines. Springer (2008)
20. Strehl, A., Ghosh, J., Mooney, R.: Impact of similarity measures on web-page clustering. In: AAAI Workshop on Artificial Intelligence for Web Search. pp. 58–64 (2000)

21. Subbarao, R., Meer, P.: Nonlinear mean shift over Riemannian manifolds. Int. Journal of Computer Vision 84(1), 1–20 (2009)
22. Turaga, P., Veeraraghavan, A., Srivastava, A., Chellappa, R.: Statistical computations on Grassmann and Stiefel manifolds for image and video-based recognition. IEEE Trans. Pattern Analysis and Machine Intelligence 33(11), 2273–2286 (2011)
23. Vemulapalli, R., Pillai, J.K., Chellappa, R.: Kernel learning for extrinsic classification of manifold features. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1782–1789 (2013)
24. Wolf, L., Shashua, A.: Learning over sets using kernel principal angles. Journal of Machine Learning Research 4, 913–931 (2003)
25. Yu, S., Tan, T., Huang, K., Jia, K., Wu, X.: A study on gait-based gender classification. IEEE Trans. Image Processing (TIP) 18(8), 1905–1910 (2009)
26. Zheng, S., Zhang, J., Huang, K., He, R., Tan, T.: Robust view transformation model for gait recognition. In: International Conference on Image Processing (ICIP). pp. 2073–2076 (2011)