Koopman Spectral Kernels for Comparing Complex Dynamics: Application to Multiagent Sport Plays

Keisuke Fujii[1(✉)], Yuki Inaba[2], and Yoshinobu Kawahara[1,3]

[1] Center for Advanced Intelligence Project, RIKEN, Osaka, Japan
`keisuke.fujii.zh@riken.jp`
[2] Japanese Institute of Sports Sciences, Tokyo, Japan
`yuki.inaba@jpnsport.go.jp`
[3] The Institute of Scientific and Industrial Research, Osaka University, Osaka, Japan
`ykawahara@sanken.osaka-u.ac.jp`

Abstract. Understanding the complex dynamics in the real-world such as in multi-agent behaviors is a challenge in numerous engineering and scientific fields. Spectral analysis using Koopman operators has been attracting attention as a way of obtaining a global modal description of a nonlinear dynamical system, without requiring explicit prior knowledge. However, when applying this to the comparison or classification of complex dynamics, it is necessary to incorporate the Koopman spectra of the dynamics into an appropriate metric. One way of implementing this is to design a kernel that reflects the dynamics via the spectra. In this paper, we introduced Koopman spectral kernels to compare the complex dynamics by generalizing the Binet-Cauchy kernel to nonlinear dynamical systems without specifying an underlying model. We applied this to strategic multiagent sport plays wherein the dynamics can be classified, e.g., by the success or failure of the shot. We mapped the latent dynamic characteristics of multiple attacker-defender distances to the feature space using our kernels and then evaluated the scorability of the play by using the features in different classification models.

1 Introduction

Groups of organisms competing and cooperating in nature are assumed to behave as complex and nonlinear dynamical systems, which currently elude formulation [7,9]. Understanding the complex dynamics of living organisms or artificial agents (and the component parts) is a challenging research area in biology [5], physics [7], and machine learning. In the field of physics, decomposition or spectral methods that factorize the dynamics into modes from the data are used such as proper orthogonal decomposition (POD) [1,25] or dynamic mode decomposition (DMD) [23,24]. The problem of learning dynamical systems in machine learning has been discussed such as in terms of Bayesian approaches [10] and

predictive state representation [19]. This topic is closely related to the decomposition technique in physics, aiming to estimate a prediction model by examining the obtained modes.

In this paper, we consider the following discrete-time nonlinear dynamical system:

$$\boldsymbol{x}_{t+1} = \boldsymbol{f}\left(\boldsymbol{x}_t\right) \tag{1}$$

where $\boldsymbol{x}_i$ is a state vector on the state space $\mathcal{M}$ (i.e., $x \in \mathcal{M} \subset \mathbb{R}^d$) and $\boldsymbol{f}$ is a state transition function that assumes the dynamical system to be nonlinear. A recent development is the use of Koopman spectral analysis with reproducing kernels (called kernel DMD). This defines a mode that can yield direct information about the nonlinear latent dynamics [16]. However, to compare or classify these complex dynamics, it is necessary to incorporate their Koopman spectrum into a metric appropriate for representing the similarity between the nonlinear dynamical systems.

Several works have applied approximation with a low-dimensional linear subspace to represent this similarity [12,30,33]. One approach has used the Binet-Cauchy (Riemannian) distance with a variety of kernels on a Grassman manifold [12], such as the kernel principal angle [33], and the trace and determinant kernel [30], which were designed for application in face recognition [33] and movie clustering [30]. The algorithm essentially calculates the Binet-Cauchy distance between two subspaces in the feature space, defined by the product of the canonical correlations. However, the main applications assumed a linear dynamical model [12,30,33] and thus generalization to nonlinear dynamics without specifying an underlying model remains to be addressed. In this paper, we map the latent dynamics to the feature space using the kernels, allowing binary classification to be applied to real-world complex dynamical systems.

Organized human group tasks such as navigation [13] or ballgame teams [8] provide excellent examples of complex dynamics and pose challenges in machine learning because of their switching and overlapping hierarchical subsystems [8], characterized by recursive shared intentionality [28]. Measurement systems have been developed that capture information regarding the position of a player in a ballgame, allowing analysis of particular shots [11]; however, plays involving collaboration between several teammates have not yet been addressed. In games such as basketball or football, coaches analyze team formations and players repeatedly practice moves that increase the probability of scoring ("scorability"). However, the selection of tactics is an ill-posed problem, and thus basically requires the implicit experience-based knowledge of the coach. An algorithm is needed that clarifies scorable moves involving multiple players in the team.

Previous research has classified team moves on a global scale by directly applying machine learning methods derived mainly from natural language processing. These include recursive neural networks (RNN) using optical flow images of the trajectories of all players [31] or the application of latent Dirichlet allocation (LDA) to the arrangement of individual trajectories [22]. However, the contribution of team movement to the success of a play remains unclear.

Previously, we reported that three maximum attacker-defender distances separately explained scorability [8], but the study addressed only the outcome of a play, rather than its time evolution and the interactions that it comprised. An algorithm is required that uses mapping to feature space to discriminate between successful and unsuccessful moves while accounting for these complex factors. In this paper, we map the latent dynamic characteristics of multiple attacker-defender distances [8] to the feature space using our kernels acquired by kernel DMD and then evaluated scorability.

The rest of the paper is organized as follows. Section 2 briefly reviews the background of Koopman spectral kernels, while Sect. 3 discusses methods for computing them. We then extended this to empirical example of actual human locomotion in Sect. 4. For application to multiple sporting agents, Sect. 5 reports our findings using the data on actual basketball games. Our approach proved capable of capturing complex team moves. Finally, Sect. 6 presents our discussion and conclusions.

2   Background

2.1   Koopman Spectral Analysis and Dynamic Mode Decomposition

Spectral analysis (or decomposition) for analyzing dynamical systems is a popular approach aimed at extracting low-dimensional dynamics from the data. Common techniques include global eigenmodes for linearized dynamics, discrete Fourier transforms, and POD for nonlinear dynamics [25], as well as multiple variants of these techniques. DMD has recently attracted particular attention in areas of physics such as fluid mechanics [23] and several engineering fields [2,26] because of its ability to define a mode that can yield direct information even when applied to time series with nonlinear latent dynamics [23,24]. However, the original DMD has numerical disadvantages, related to the accuracy of the approximate expressions of the Koopman eigenfunctions derived from the data. A number of variants have been proposed to address this shortcoming, including exact DMD [29], optimized DMD [4], and baysian DMD [27]. Sparsity-promoting DMD [14] provides a framework for the approximation of the Koopman eigenfunctions with fewer bases. Extended DMD [32], which works on predetermined kernel basis functions, has also been proposed. These Koopman spectral analyses have been generalized to a reproducing kernel Hilbert space (RKHS) [16], an approach which is called *kernel DMD*.

In Koopman spectral analysis, the Koopman operator $\mathcal{K}$ [18] is an infinite dimensional linear operator acting on the scalar function $g_i : \mathcal{M} \to \mathbb{C}$. That is, it maps g_i to the new function $\mathcal{K}g_i$ as follows:

$$\left(\mathcal{K}g_i\right)(\boldsymbol{x}) = \left(g_i \circ \boldsymbol{f}\right)(\boldsymbol{x}),\tag{2}$$

where $\mathcal{K}$ denotes the composition of g_i with $\boldsymbol{f}$. We can see that $\mathcal{K}$ acts linearly on the function g_i. The dynamics defined by $\boldsymbol{f}$ may be nonlinear. Since $\mathcal{K}$ is a linear operator, it can generally perform eigenvalue decomposition:

$$\mathcal{K}\varphi_j\left(\boldsymbol{x}\right) = \lambda_j\varphi_j\left(\boldsymbol{x}\right), \tag{3}$$

where $\lambda_j \in \mathbb{C}$ is the jth eigenvalue (called the *Koopman eigenvalue*) and φ_j is the corresponding eigenfunction (called the *Koopman eigenfunction*). We denote the concatenation of g_j to $\boldsymbol{g} := [g_1, \ldots, g_p]^{\mathrm{T}}$. If each g_j lies within the space spanned by the eigenfunction φ_j, we can expand the vector-valued $\boldsymbol{g}$ in terms of these eigenfunctions as $\boldsymbol{g}(\boldsymbol{x}) = \sum_{j=1}^{\infty} \varphi_j(\boldsymbol{x})\boldsymbol{\psi}_j$, where $\boldsymbol{\psi}_j$ is a set of vector coefficients called *Koopman modes*. By iterative application of Eqs. (2) and (3), the following equation is obtained:

$$(\boldsymbol{g} \circ \boldsymbol{f})\left(\boldsymbol{x}\right) = \sum_{j=1}^{\infty} \lambda_j\varphi_j\left(\boldsymbol{x}\right)\boldsymbol{\psi}_j. \tag{4}$$

Therefore, λ_j characterizes the time evolution of the corresponding Koopman mode $\boldsymbol{\psi}_j$, i.e., the phase of λ_j determines its frequency and the magnitude determines the growth rate of its dynamics.

DMD is a popular approach for estimating the approximations of λ_j and $\boldsymbol{\psi}_j$ from a finite length observation data sequence $y_0, y_1, \ldots, y_\tau$ ($\in \mathbb{R}^p$), where $y_t := \boldsymbol{g}(\boldsymbol{x}_t)$. Let $\boldsymbol{A} = [y_0, y_1, \ldots, y_{\tau-1}]$ and $\boldsymbol{B} = [y_1, y_2, \ldots, y_\tau]$. Then, DMD basically approximates those by calculating the eigendecomposition of the least-squares solution to

$$\min_{P' \in \mathbb{R}^{p \times p}} (1/\tau) \sum_{t=0}^{\tau} \|y_{t+1} - \boldsymbol{P}'y_t\|^2, \tag{5}$$

i.e., $\boldsymbol{B}\boldsymbol{A}^\dagger(:= \boldsymbol{P})$ ($\bullet^\dagger$ is the pseudo-inverse of $\bullet$). Let the j-th right and left eigenvector of $\boldsymbol{P}$ be $\boldsymbol{\psi}_j$ and $\boldsymbol{\kappa}_j$, respectively, and assume that these are normalized so that $\boldsymbol{\kappa}_i^* \boldsymbol{\psi}_j = \delta_{ij}$ (δ_{ij} is the Kronecker's delta). Then, since any vector $\boldsymbol{b} \in \mathbb{C}^p$ can be written as $\boldsymbol{b} = \sum_{j=1}^{p} (\boldsymbol{\kappa}_i^* \boldsymbol{b})\boldsymbol{\psi}_j$, we have $\boldsymbol{g}(\boldsymbol{x}) = \sum_{j=1}^{p} \varphi_j(x)\boldsymbol{\psi}_j$ by applying it to $\boldsymbol{g}(\boldsymbol{x})$. Therefore, by applying $\mathcal{K}$ to both sides, we have

$$(\boldsymbol{g} \circ \boldsymbol{f})\left(\boldsymbol{x}\right) = \sum_{j=1}^{p} \lambda_j\varphi_j\left(\boldsymbol{x}\right)\boldsymbol{\psi}_j, \tag{6}$$

indicating a modal representation corresponding to Eq. (4) for the finite sum.

2.2   Kernels for Comparing Nonlinear Dynamical Systems

Selection of an appropriate representation of the data is a fundamental issue in pattern recognition. The important point is to design the features (i.e., kernels) that reflect structure of the data. Time series data is challenging to design the features because of the difficulty in reflecting the data structure (including time length). Researchers have developed alternative kernel methods, including the

use of graphs [15,17], subspaces [12,33] or trajectories [30]. In this paper, a kernel design applicable to dynamical systems was required. Several methods were proposed, based on the subspace angle with kernel methods such as for an auto-regressive moving average (ARMA) model [30]. These methodologies were previously reviewed [12], from the viewpoint of the Riemannian distance (or metric) on the Grassman manifold.

The Grassmann manifold $\mathcal{G}(m, D)$ is the set of m-dimensional linear subspaces of $\mathbb{R}^D$. Formally, the Riemannian distance between two subspaces is the geodesic distance on the Grassmann manifold. However, a more intuitive and computationally efficient way of defining the distances uses the principal angles [20]. A previous review [12] categorized the various Riemannian distances into the projection and Binet-Cauchy distance. The former has been used in applications such as face recognition [3,12], and the latter has been applied in video clustering [30] and face recognition [33], and has been generalized to (specific nonlinear) dynamical systems [30]. We then adopted the Binet-Cauchy distance when comparing complex systems.

The Binet-Cauchy distances were basically obtained with the product of canonical correlations using a variety of kernels [30]. However, the main applications assumed linear dynamical model [12,30,33] such as ARMA model. Thus, it is necessary to generalize to nonlinear dynamics without any specific underlying model, into which the Koopman spectrum of dynamics is incorporated. We called the kernels *Koopman spectral kernels.*

3   Design of Koopman Spectral Kernels

3.1   DMD with Reproducing Kernel

Conceptually, DMD can be considered as producing a local approximation of the Koopman eigenfunctions using a set of linear monomials of the observables as the basis functions. In practice, however, this is certainly not applicable to all systems (in particular, beyond the region of validity for local linearization). Then, DMD with reproducing kernels [16] approximates the Koopman eigenfunctions with richer basis functions.

Let $\mathcal{H}$ be the RKHS embedded with the dot product determined by a positive definite kernel k. Additionally, let $\phi : \mathcal{M} \to \mathcal{H}$ be a feature map, and an instance of ϕ with respect to $\boldsymbol{x}$ is denoted by $\phi_{\boldsymbol{x}}$ (i.e., $\phi_{\boldsymbol{x}} := \phi(\boldsymbol{x})$). Then, we define the Koopman operator $\mathcal{K}_{\mathcal{H}} : \mathcal{H} \to \mathcal{H}$ in the RKHS by

$$\mathcal{K}_{\mathcal{H}}\phi_{\boldsymbol{x}} = \phi_{\boldsymbol{x}} \circ \boldsymbol{f}. \tag{7}$$

Note that almost of the theoretical claims in this study do not necessarily require ϕ to be in the RKHS (it is sufficient to consider that ϕ stays within a Hilbert space), but this assumption should perform the calculation in practice.

In this paper, we robustify the kernel DMD by projecting data onto the direction of POD [4,16,29]. First, a centered Gram matrix is defined by $\bar{G} = \mathbf{H}G\mathbf{H}$, where G is a Gram matrix, $\mathbf{H} = \mathbf{I} - \mathbf{1}\tau$, $\mathbf{I}$ is a unit matrix, and $\mathbf{1}\tau$ is

a τ-by-τ matrix, for which each element takes the value $1/\tau$. The Gram matrix G_{xx} of the kernel $k(\boldsymbol{y}_i, \boldsymbol{y}_j)$ is defined at $\boldsymbol{y}_i$ and $\boldsymbol{y}_j$ (i and j dimensions) of the observation data matrix $\boldsymbol{A}$. Similarly, the Gram matrix G_{xy} of the kernel between $\boldsymbol{A}$ and $\boldsymbol{B}$ can be calculated. At this time, $G_{xx} = \mathcal{M}_\tau^* \mathcal{M}_\tau$ and $G_{xy} = \mathcal{M}_\tau^* \mathcal{M}_+$, where $\mathcal{M}_\tau^*$ indicates the Hermitian transpose of $\mathcal{M}_\tau$. Also, $\mathcal{M}_\tau := [\phi_{\boldsymbol{x}_0}, .., \phi_{\boldsymbol{x}_{\tau-1}}]$ and $\mathcal{M}_+ := [\phi_{\boldsymbol{x}_1}, .., \phi_{\boldsymbol{x}_\tau}]$, where $\phi_{\boldsymbol{x}_i}$ is considered as a feature map of $\boldsymbol{x}_i$ from the state space $\mathcal{M}$ to the RKHS $\mathcal{H}$.

Here, suppose that the eigenvalues and eigenvectors can be truncated based on eigenvalue magnitude. In other words, $\bar{G} \approx \bar{B}\bar{G}\bar{B}^*$ where p ($\leq \tau$) eigenvalues are adopted. Then, a principal orthogonal direction in the feature space is given by

$$\nu_j = \mathcal{M}_\tau \mathbf{H} \bar{S}_{jj}^{-1/2} \boldsymbol{\beta}_j, \tag{8}$$

where $\boldsymbol{\beta}_j$ is the jth row of $\bar{B}$. Let $\mathcal{U} = [\nu_1, \ldots, \nu_j] = \mathcal{M}_\tau \mathbf{H} \bar{B} \bar{S}^{-1/2}$. Since $\mathcal{M}_+ = \mathcal{K}_\mathcal{H} \mathcal{M}_\tau$, the projection of $\mathcal{K}_\mathcal{H}$ onto the space spanned by ν_j is given as follows:

$$\hat{F} = \mathcal{U} \mathcal{K}_\mathcal{H} \mathcal{U} = \bar{S}^{-1/2} \bar{B}^* \mathbf{H} (\mathcal{M}_\tau \mathcal{M}_+) \mathbf{H} \bar{B} \bar{S}^{-1/2}. \tag{9}$$

Note that $G_{xy} = \mathcal{M}_\tau^* \mathcal{M}_+$. Then, if we let $\hat{F} = \hat{T}^{-1} \hat{\Lambda} \hat{T}$ be the eigendecomposition of $\hat{F}$, we obtain the centered DMD mode $\bar{\varphi}_j = \mathcal{U} \boldsymbol{b}_j = \mathcal{M}_\tau \mathbf{H} \bar{B} \bar{S}^{-1/2} \boldsymbol{b}_j$, where b_j is the jth row of $\hat{T}^{-1}$. The diagonal matrix $\hat{\Lambda}$ comprising the eigenvalues represents the temporal evolution of the mode.

3.2   Koopman Spectral Kernels

For calculating the similarity between the dynamical systems DS_i and DS_j, we compute Koopman spectral kernels based on the idea of Binet-Cauchy kernels. The Binet-Cauchy kernels are basically calculated from the traces of compound matrices [30] defined as follows. Let M be a matrix in $\mathbb{R}^{m \times n}$. For $q \leq min(m,n)$, define $I_q^n = \{\boldsymbol{i} = i_1, \cdots, i_q : 1 \leq i_1 < ... < i_q \leq n, i_i \in \mathbb{N}\}$, and likewise I_q^m. We denote by $C_q(M)$ the qth compound matrix, that is, the $\binom{m}{q} \times \binom{n}{q}$ matrix whose elements are the minors $det((M_{k,l})_{k \neq \boldsymbol{i}, l \neq \boldsymbol{j}})$, where $\boldsymbol{i} \in I_q^n$ and $\boldsymbol{j} \in I_q^m$ are assumed to be arranged in lexicographical order. In the unifying viewpoint [30], Binet-Cauchy kernels is a general representation including various kernels [6,15, 17,21], divided into two strategies. One is the trace kernel obtained by setting $q = 1$ (i.e., $C_1(M) = M$), which directly reflects the property of temporal evolution of the dynamical systems, including diffusion kernel [17] and graph kernel [15]. Second is the determinant kernel obtained by setting order q to be equal to the order of the dynamical systems n (i.e., $C_n(M) = det(M)$), which extracts coefficients of dynamical systems, including the Martin distance [21] and the distance based on the subspace angle [6].

We expand the kernels to applying Koopman spectral analysis, which are called the *Koopman trace kernel* and *Koopman determinant kernel*, respectively. Both kernels reflect the Koopman eigenvalue, the eigenfunction, and the mode

(i.e., system trajectory including the initial condition). However, richer information of system trajectory does not necessarily increase expressiveness such as in classification with real-world data. Therefore, we also expanded the kernel of principal angle [33] to applying Koopman spectral analysis, which is called *Koopman kernel of principal angle*. The kernel principal angle is theoretically a simple case of the trace kernel [30], which is defined as the inner product of linear subspaces in this feature space. In this paper, for a simple comparison, we compute the kernel with the inner product of the Koopman modes (i.e. not the trajectory and independent of initial condition).

Koopman Trace Kernel and Determinant Kernel. First, for the trace kernel, we generalize the kernel assmuing the ARMA model [30], to nonlinear dynamical systems without specifying an underlying model. The trace kernel of DS_i and DS_j can be theoretically defined as follows:

$$k\left(DS_i, DS_j\right) := \sum_{t=0}^{\infty} \left(e^{-\kappa t} \boldsymbol{g}_i\left(\boldsymbol{x}_{i,t}\right)^{\mathrm{T}} \boldsymbol{W} \boldsymbol{g}_j\left(\boldsymbol{x}_{j,t}\right)\right), \qquad (10)$$

where $\boldsymbol{g}_i$ and $\boldsymbol{g}_j$ is the observation function and $\boldsymbol{W}$ is an arbitrary semidefinite matrix (here, $\boldsymbol{W} = \boldsymbol{1}$). Moreover, for converging the above equation, we suppose the exponential discount $\mu(t) = e^{-\kappa t}(\kappa > 0)$. In this paper, noises in observation and latent dynamics are not considered. Koopman trace kernel can be computed using the modal representation given by the kernel DMD as follows:

$$k\left(DS_i, DS_j\right) = \boldsymbol{\varphi}_i\left(\boldsymbol{x}_{i,0}\right)^{\mathrm{T}} \sum_{t=0}^{\infty} \left(e^{-\kappa t} \boldsymbol{\Lambda}_i^t \left(\boldsymbol{\Psi}_i^{\mathrm{T}} \boldsymbol{W} \boldsymbol{\Psi}_j\right) \boldsymbol{\Lambda}_j^t\right) \boldsymbol{\varphi}_j\left(\boldsymbol{x}_{j,0}\right), \qquad (11)$$

where, $\boldsymbol{\Lambda_i}$ is a diagonal matrix consisting of Koopman eigenvalues, $\boldsymbol{\Psi_i}$ is the Koopman mode, and $\boldsymbol{\varphi}_i$ is the Koopman eigenfunction (also for j). Although the equation includes an infinite sum, we can efficiently compute the matrix $\boldsymbol{M} := \sum_{t=0}^{\infty}\left(e^{-\kappa t}\boldsymbol{\Lambda}_i^t(\boldsymbol{\Psi}_i^{\mathrm{T}}\boldsymbol{W}\boldsymbol{\Psi}_j)\boldsymbol{\Lambda}_j^t\right)$ using the following Sylvester equation $\boldsymbol{M} = e^{-\kappa}\boldsymbol{\Lambda}_i^{\mathrm{T}}\boldsymbol{M}\boldsymbol{\Lambda}_j + \boldsymbol{\Psi}_i^{\mathrm{T}}\boldsymbol{W}\boldsymbol{\Psi}_j$, where the Koopman mode $\boldsymbol{\Psi} = \mathcal{U}^*\boldsymbol{H}\mathcal{M}_\tau\boldsymbol{H}\mathcal{U}\hat{T}^{-1}$ for i and j. For creating a trace kernel independent of the initial conditions [30], we take expectation over $\boldsymbol{x}_{i,0}$ and $\boldsymbol{x}_{j,0}$ in the trace kernel, yielding

$$k\left(DS_i, DS_j\right) = tr\left(\boldsymbol{\Sigma}_{\boldsymbol{\varphi}_i(\boldsymbol{x}_{i,0}), \boldsymbol{\varphi}_j(\boldsymbol{x}_{j,0})}\boldsymbol{M}\right), \qquad (12)$$

where the initial Koopman eigenvalue $\boldsymbol{\varphi}(x_0) = \boldsymbol{a}^*(\mathcal{M}_\tau\boldsymbol{H}\mathcal{U})^*\mathcal{M}_{\tau,0}$ for i and j [16]. Here, $\boldsymbol{a}$ is the left eigenvector of $\hat{F}$ and $\mathcal{M}_{\tau,0}$ is a vector indicating the first single column of $\mathcal{M}_\tau$. $\Sigma_{\boldsymbol{\varphi}_i(\boldsymbol{x}_{i,0}), \boldsymbol{\varphi}_j(\boldsymbol{x}_{j,0})} \in \mathbb{C}^{p \times p}$ is the covariance of all initial values $\boldsymbol{\varphi}_n\left(\boldsymbol{x}_0\right) \in \mathbb{C}^{p \times n}$ of DS_i for each index $1, ... p$ of eigenvalues (p was fixed for all i). Similarly, the determinant kernel using the representation given by kernel DMD can be computed:

$$k\left(DS_i, DS_j\right) = det\left(\boldsymbol{\Psi}_i \boldsymbol{M} \boldsymbol{\Psi}_j^{\mathrm{T}}\right), \qquad (13)$$

where $\boldsymbol{M} = e^{-\kappa}\boldsymbol{\Lambda}_i^{\mathrm{T}}\boldsymbol{M}\boldsymbol{\Lambda}_j + \boldsymbol{\varphi}_i(\boldsymbol{x}_{i,0})\boldsymbol{\varphi}_j(\boldsymbol{x}_{j,0})^{\mathrm{T}}$. Determinant kernels independent of the initial condition can only be computed for a single output system [30].

Koopman Kernel of Principal Angle. The kernel of principal angle can be computed using the Koopman modes given by kernel DMD. With respect to DS_i, we define the kernel of principal angles as the inner product of the Koopman modes in the feature space: $A^*A = \hat{T}_i^{-1}\mathcal{U}_i^*\mathbf{H}G_{xxi}\mathbf{H}\mathcal{U}_i\hat{T}_i$. If the rank of $\hat{F}$ is r_i, A^*A is a r_i-order square matrix. Also for DS_j, we create a similar matrix B^*B. Furthermore, we define the inner product of the linear subspaces between DS_i and DS_j as $A^*B = \hat{T}_i^{-1}\mathcal{U}_i^*\mathbf{H}G_{xxij}\mathbf{H}\mathcal{U}_j\hat{T}_j$. G_{xxij} is a $n_i \times n_j$ matrix obtained by picking up the upper-right part of the centered Gram matrix obtained by connecting $\boldsymbol{A}_i$ and $\boldsymbol{A}_j$ in series (n_i and n_j are the lengths of the time series). Then, using these matrices, we solve the following generalized eigenvalue problem:

$$\begin{pmatrix} 0 & (A^*B)^* \\ A^*B & 0 \end{pmatrix} \boldsymbol{V} = \boldsymbol{\lambda}_{ij} \begin{pmatrix} B^*B & 0 \\ 0 & A^*A \end{pmatrix} \boldsymbol{V}, \qquad (14)$$

where the size of $\boldsymbol{\lambda}_{ij}$ is finally adjusted to $r_{ij} = \min(r_i, r_j)$ in descending order, and $\boldsymbol{V}$ is a generalized eigenvector. The eigenvalue $\boldsymbol{\lambda}_{ij}$ is the kernel of principal angle.

4   Embedding and Classification of Dynamics

A direct but important application of this analysis is the embedding and classification of dynamics using extracted features. A set of Koopman spectra estimated from the analysis can be used as the basis for a low-dimensional subspace representing the dynamics. The classification of dynamics can be performed using feature vectors determined by the Koopman spectral kernels. We used the Gaussian kernel, with the kernel width set as the median of the distances from a data matrix.

Before applying our approach to multiagent sports data, an experiment was conducted using open-source real-world data. In this case, human locomotion data were taken from the CMU Graphics Lab Motion Capture Database (available at http://mocap.cs.cmu.edu). To verify the classification performance, we computed the trace kernel of an auto-regressive (AR) model, representing a conventional linear dynamical model. For embedding of the distance matrix with our kernels, components of the distance matrix between DS_i and DS_j in the feature space were obtained using $dist(DS_i, DS_j) = k(\boldsymbol{A}_i, \boldsymbol{A}_i) + k(A_j, A_j) - 2k(A_i, A_j)$. Figure 1a–c shows the embedding of the sequences using multidimensional scaling (MDS) with the distance matrix, computed with the Koopman kernel principal angle, Koopman determinant kernel, and trace kernel of the AR model, respectively. Classification of performances into jumping, running, and walking was computed using the k-nearest neighbor algorithm. Error rates of the test data were small in this order: the Koopman kernel of principal angle (0.261), Koopman determinant kernel (0.348), trace kernel of the AR model (0.522), and Koopman trace kernel (0.601). Two Koopman spectral kernels performed better in classification than the kernel of the linear dynamical model.
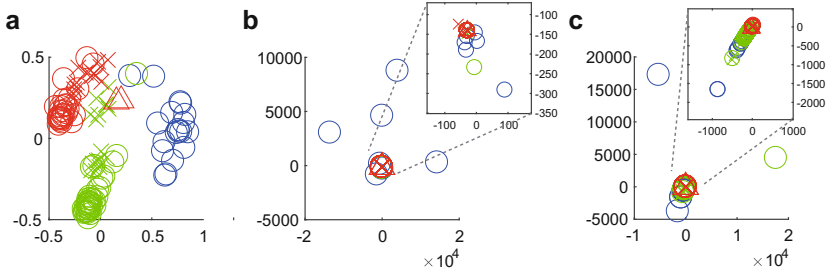
Fig. 1. MDS embedding of (**a**) Koopman kernel of principal angle, (**b**) Koopman determinant kernel, and (**c**) trace kernel of AR model. Blue, red, and green indicate jump, run, and walk, respectively (x and triangle show the movements with turn and stop, respectively). (Color figure online)

## 5   Application to Multiagent Sport Plays

We used player-tracking data from two international basketball games in 2015 collected by the STATS SportVU system. The total playing time was 80 min, and the total score of the two teams was 276. Positional data comprised the xy position of every player and the ball on the court, recorded at 25 frames per second. We eliminated transitions in attack to automatically extract the time periods to be analyzed (called an *attack-segment*). We defined an attack-segment as the period from all players on the attacking side court entry to 1 s before a shot was made. We analyzed a total of 192 attack-segments, 77 of which ended in a successful shot.

Next, we calculated effective attacker-defender distances to predict the success or failure of the shot (details were given by [8]), which were temporally and spatially corrected (Fig. 2a). Although all of the distances were 25 dimensions (five attackers and defenders), we previously reduced to four dimensions [8]: (1) ball-mark distance, (2) ball-help distance, (3) pass-mark distance, and (4) pass-help distance (Fig. 2b–c). These distances were used to create seven input vector series: (i) a one-dimensional distance (1), (ii) a two-dimensional distance comprising (1) and (2), and (iii–iv) three- and four-dimensional (1–3, 1–4) important distances, respectively. For verification, (v) total 25 distances and (vi) 25-dimensional Euclidean distances without spatiotemporal correction were calculated. We also used (vii) the xy position (total 20 dimensions) of all the ten players.

When predicting the outcome of a team-attack movement, it is preferable to compute the posterior probability rather than the outcome identification of the shot accuracy itself. We used a naive Bayes classifier and a related vector machine (RVM) for classification. Figure 3a shows the result of applying the naive Bayes classifier. The horizontal axis shows the seven input vector series and the vertical axis the classification error. The Koopman kernel principal angles derived by inputting four important distances demonstrated minimum error of 35.9%. The result of applying the RVM is shown in Fig. 3b, using the same axes.
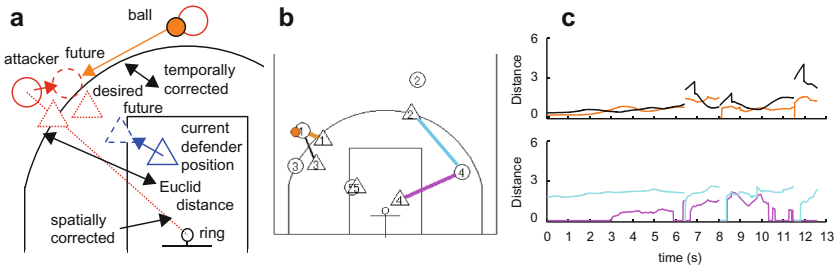
Fig. 2. Diagrams and examples of attacker-defender distance. (a) Diagram of attacker-defender distance with spatiotemporal correction. (b) Examples of four important distances. Orange, black, pink and light blue indicate the ball-mark, ball-help, pass-mark, and pass-help distance, respectively. (c) Example of time series in the same four important attacker-defender distances. (Color figure online)
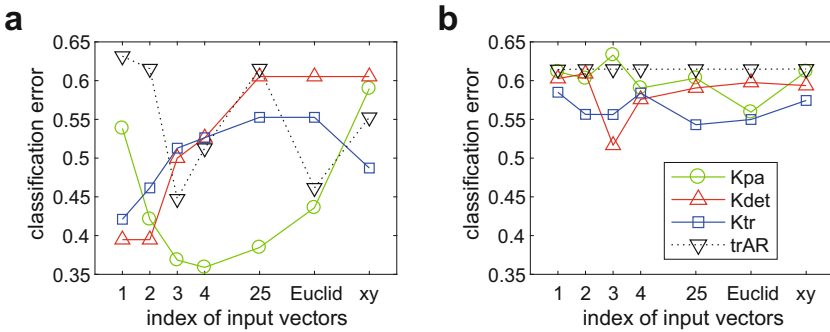
Fig. 3. Results from applying (a) the naive Bayes classifier and (b) the relevant vector machine. Kpa, Kdet, Ktr, and trAR are Koopman kernel of principal angle, Koopman determinant kernel and trace kernel, and trace kernel with AR model, respectively.

The performance of the naive Bayes classifier was superior to that of the RVM. In both cases, the Koopman spectral kernels produced better classification than the kernel of the linear dynamical model.

Figure 4a–c show embedding via MDS with the distance matrix of the Koopman kernel of principal angle countered by frequencies of success and failure of the shot. For example, the best case of the four important attacker-defender distances (Fig. 4a) showed the expressiveness in scorability due to wide distribution across the plot. In contrast, they were less widely distributed when only single distance (Fig. 4b) or the xy coordinates of all players (Fig. 4c) were used.
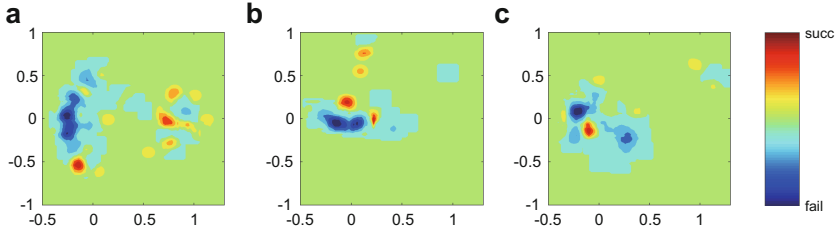
Fig. 4. MDS embedding of Koopman kernel of principal angle with three input vector series. The series consisted of (a) four important distances, (b) single important distance, and (c) xy coordinates of all players. Red and blue indicates success and failure of the shot, respectively. (Color figure online)

6   Discussion and Conclusion

The results of the two empirical examples showed that the best performances of the Koopman spectral kernels (Koopman determinant kernel and kernel of principal angle) are superior to that of the AR model assuming a linear dynamical model. Our proposed kernels can be computed in a closed form; but practically, the values of the Koopman determinant kernel were too large and the performance of the Koopman trace kernel was no better than that of the others. In contrast, the Koopman kernel of principal angle showed effective expressiveness only using Koopman modes.

When applied to multiagent sports data, the highest performance was provided by the classifier using the four important distances. This vector series reflects four characteristics: the scorability of a player in the current and future (i) shot, (ii) dribble, and (iii) pass, and (iv) the scorability of a dribbler after the pass. The proposed kernel reflected the time series of all interactions between players and was more effective for the classification than the kernel based on the information only on the shot itself. Well-trained teams aim to create scoring opportunities by continuously selecting tactical passes and dribbles or by improvising when no shooting opportunity is available.

However, even the best classification was not high (64.1% accuracy) when applied to real multiagent sports data. Two factors may have been neglected by our framework. The first is the existence of local interactions between players, such as local competitive and cooperative play by the attackers and defenders [8] when seen in higher spatial resolution than was available in this study. The approach needs to reflect the hierarchical characteristics of global dynamics and local dynamics. The second is the limitation of the input vector series to the attacker-defender distances. To achieve more accurate classifiers, not only the most important factor (i.e., distance) but also further hand-made time-series input vector series (e.g., Cartesian coordinates or specific movement parameters) should be used.

Overall, we developed Koopman spectral kernels that can be computed in closed form and used to compare multiple nonlinear dynamical systems. In competitive sports, coaches spend considerable amounts of time analyzing videos of their own team and the opposing team. Application of a system such as the one presented here may save time and create tactical plans that can currently be generated only by experienced coaches. More generally, the algorithm can be applied to the analysis of the complex dynamics of groups of living organisms or artificial agents, which currently elude formulation.

References

1. Bonnet, J., Cole, D., Delville, J., Glauser, M., Ukeiley, L.: Stochastic estimation and proper orthogonal decomposition: complementary techniques for identifying structure. Exp. Fluids **17**(5), 307–314 (1994)
2. Brunton, B.W., Johnson, L.A., Ojemann, J.G., Kutz, J.N.: Extracting spatial-temporal coherent patterns in large-scale neural recordings using dynamic mode decomposition. J. Neurosci. Methods **258**, 1–15 (2016)
3. Chang, J.M., Beveridge, J.R., Draper, B.A., Kirby, M., Kley, H., Peterson, C.: Illumination face spaces are idiosyncratic. In: Proceedings of International Conference on Image Processing, Computer Vision, & Pattern Recognition, vol. 2, pp. 390–396 (2006)
4. Chen, K.K., Tu, J.H., Rowley, C.W.: Variants of dynamic mode decomposition: boundary condition, Koopman, and Fourier analyses. J. Nonlinear Sci. **22**(6), 887–915 (2012)
5. Couzin, I.D., Krause, J., Franks, N.R., Levin, S.A.: Effective leadership and decision-making in animal groups on the move. Nature **433**(7025), 513–516 (2005)
6. De Cock, K., De Moor, B.: Subspace angles between ARMA models. Syst. Control Lett. **46**(4), 265–270 (2002)
7. Fodor, E., Nardini, C., Cates, M.E., Tailleur, J., Visco, P., van Wijland, F.: How far from equilibrium is active matter? Phys. Rev. Lett. **117**(3), 038103 (2016)
8. Fujii, K., Yokoyama, K., Koyama, T., Rikukawa, A., Yamada, H., Yamamoto, Y.: Resilient help to switch and overlap hierarchical subsystems in a small human group. Scientific Reports 6 (2016)
9. Fujii, K., Isaka, T., Kouzaki, M., Yamamoto, Y.: Mutual and asynchronous anticipation and action in sports as globally competitive and locally coordinative dynamics. Scientific Reports 5 (2015)
10. Ghahramani, Z., Roweis, S.T.: Learning nonlinear dynamical systems using an EM algorithm. In: Advances in Neural Information Processing Systems, pp. 431–437 (1999)
11. Goldman, M., Rao, J.M.: Live by the three, die by the three? The price of risk in the NBA. In: Proceedings of MIT Sloan Sports Analytics Conference (2013)
12. Hamm, J., Lee, D.D.: Grassmann discriminant analysis: a unifying view on subspace-based learning. In: Proceedings of International Conference on Machine Learning, pp. 376–383 (2008)

13. Hutchins, E.: The technology of team navigation. In: Intellectual Teamwork: Social and Technological Foundations of Cooperative Work, vol. 1, pp. 191–220 (1990)
14. Jovanović, M.R., Schmid, P.J., Nichols, J.W.: Sparsity-promoting dynamic mode decomposition. Phys. Fluids **26**(2), 024103 (2014)
15. Kashima, H., Tsuda, K., Inokuchi, A.: Kernels for graphs. Kernel Methods Comput. Biol. **39**(1), 101–113 (2004)
16. Kawahara, Y.: Dynamic mode decomposition with reproducing kernels for Koopman spectral analysis. In: Proceedings of Advances in Neural Information Processing Systems, pp. 911–919 (2016)
17. Kondor, R.I., Lafferty, J.: Diffusion kernels on graphs and other discrete input spaces. In: Proceedings of International Conference on Machine Learning, vol. 2, pp. 315–322 (2002)
18. Koopman, B.O.: Hamiltonian systems and transformation in Hilbert space. Proc. Natl. Acad. Sci. **17**(5), 315–318 (1931)
19. Kulesza, A., Jiang, N., Singh, S.P.: Spectral learning of predictive state representations with insufficient statistics. In: Proceedings of Association for the Advancement of Artificial Intelligence, pp. 2715–2721 (2015)
20. Loan, C.V., Golub, G.: Matrix Computations, 3rd edn. Johns Hopkins University Press, Baltimore (1996)
21. Martin, R.J.: A metric for ARMA processes. IEEE Trans. Signal Process. **48**(4), 1164–1170 (2000)
22. Miller, A.C., Bornn, L.: Possession sketches: mapping NBA strategies (2017)
23. Rowley, C.W., Mezić, I., Bagheri, S., Schlatter, P., Henningson, D.S.: Spectral analysis of nonlinear flows. J. Fluid Mech. **641**, 115–127 (2009)
24. Schmid, P.J.: Dynamic mode decomposition of numerical and experimental data. J. Fluid Mech. **656**, 5–28 (2010)
25. Sirovich, L.: Turbulence and the dynamics of coherent structures. I. Coherent structures. Q. Appl. Math. **45**(3), 561–571 (1987)
26. Susuki, Y., Mezić, I.: Nonlinear koopman modes and power system stability assessment without models. IEEE Trans. Power Syst. **29**(2), 899–907 (2014)
27. Takeishi, N., Kawahara, Y., Tabei, Y., Yairi, T.: Bayesian dynamic mode decomposition. In: Proceedings of the International Joint Conference on Artificial Intelligence (2017)
28. Tomasello, M., Carpenter, M.: Shared intentionality. Dev. Sci. **10**(1), 121–125 (2007)
29. Tu, J.H., Rowley, C.W., Luchtenburg, D.M., Brunton, S.L., Kutz, J.N.: On dynamic mode decomposition: theory and applications. J. Comput. Dyn. **1**(2), 391–421 (2014)
30. Vishwanathan, S., Smola, A.J., Vidal, R.: Binet-Cauchy kernels on dynamical systems and its application to the analysis of dynamic scenes. Int. J. Comput. Vis. **73**(1), 95–119 (2007)
31. Wang, K.C., Zemel, R.: Classifying NBA offensive plays using neural networks. In: Proceedings of MIT Sloan Sports Analytics Conference (2016)
32. Williams, M.O., Kevrekidis, I.G., Rowley, C.W.: A data-driven approximation of the Koopman operator: extending dynamic mode decomposition. J. Nonlinear Sci. **25**(6), 1307–1346 (2015)
33. Wolf, L., Shashua, A.: Learning over sets using kernel principal angles. J. Mach. Learn. Res. **4**, 913–931 (2003)