Independent component analysis for multivariate functional data

Joni Virta [a,b,*], Bing Li [c], Klaus Nordhausen [d], Hannu Oja [a]

[a] *University of Turku, Finland*
[b] *Aalto University, Finland*
[c] *Pennsylvania State University, PA, USA*
[d] *Vienna University of Technology, Austria*

A R T I C L E   I N F O

A B S T R A C T

We extend two methods of independent component analysis, fourth order blind identification and joint approximate diagonalization of eigen-matrices, to vector-valued functional data. Multivariate functional data occur naturally and frequently in modern applications, and extending independent component analysis to this setting allows us to distill important information from this type of data, going a step further than the functional principal component analysis. To allow the inversion of the covariance operator we make the assumption that the dependency between the component functions lies in a finite-dimensional subspace. In this subspace we define fourth cross-cumulant operators and use them to construct the two novel, Fisher consistent methods for solving the independent component problem for vector-valued functions. Both simulations and an application on a hand gesture data set show the usefulness and advantages of the proposed methods over functional principal component analysis.

© 2019 Elsevier Inc. All rights reserved.

1. Introduction

1.1. Independent component analysis

Independent component analysis is a classical problem in multivariate statistics and signal processing where one assumes that the observed independent and identically distributed random vectors are linear mixtures of latent random vectors having independent marginal distributions. At its simplest this corresponds to presuming that, given the observed random vector $\mathbf{x} \in \mathbb{R}^p$, there exists a non-singular unmixing matrix $\mathbf{\Gamma} \in \mathbb{R}^{p \times p}$ such that

$$\mathbf{\Gamma}\mathbf{x} = \mathbf{z}, \tag{1}$$

where the random vector $\mathbf{z} \in \mathbb{R}^p$ has independent marginals. In the independent component problem a random sample of $\mathbf{x}$ is observed and the objective is to estimate any matrix $\mathbf{\Gamma}$ such that (1) holds. We say any matrix as the formulation of the problem is clearly not well-defined, one can freely scale, permute and change the signs of the rows in (1) and the right-hand side still retains the independence of its components. As such the constraint $\mathrm{cov}(\mathbf{z}) = \mathbf{I}$ is usually introduced,

---

freeing us of the scale invariance. Further assuming that at most one of the components of $\mathbf{z}$ is normally distributed, one can show that the vector $\mathbf{z}$ can be estimated up to marginal signs, order and location [12].

Since its introduction in the 1980s a multitude of methods with varying approaches and assumptions have been proposed for solving the problem. These methods are generally based either on projection pursuit, decompositions of various matrices of cumulants or maximum likelihood. The most well-known example belonging to the first class is FastICA [18], a projection pursuit method that extracts the independent components either sequentially or simultaneously by maximizing some measure of non-Gaussianity. Several different variations of FastICA exist, see for example [24,32,33]. The second class includes not only classic methods like FOBI and JADE (see below) but also several newer ones such as [4,36]. For an example of likelihood-based methods, see, e.g., [41].

In this work we focus exclusively on two of the very first methods proposed for independent component analysis, fourth order blind identification (FOBI) [8] and joint approximate diagonalization of eigenmatrices (JADE) [9], which are simply based on the diagonalization of various moment-based matrices. As such FOBI and JADE offer an easy starting point for various extensions of independent component analysis into the realms of non-standard data structures, some examples including versions specially tailored for time series data [31], tensor-valued data [47,48] and univariate functional data [29].

Before describing our contribution we first briefly review the key steps behind FOBI and JADE, both to motivate our exposition and to contrast the constructions in the later sections. Namely, in both methods we assume that the zero-mean random vector $\mathbf{x} \in \mathbb{R}^p$ obeys the independent component model in (1), and additionally that the components of $\mathbf{z}$ have finite fourth moments, $\beta_i = \mathrm{E}(z_i^4) < \infty$, $i \in \{1, \ldots, p\}$. A basic result in independent component analysis then says that if $\boldsymbol{\Sigma}(\mathbf{x})^{-1/2}$ is the symmetric inverse square root of the covariance matrix $\boldsymbol{\Sigma}(\mathbf{x})$ of $\mathbf{x}$, then there exists an orthogonal matrix $\mathbf{U} \in \mathbb{R}^{p \times p}$ such that the standardized random vector satisfies $\mathbf{x}_{st} = \boldsymbol{\Sigma}(\mathbf{x})^{-1/2}\mathbf{x} = \mathbf{U}\mathbf{z}$.

For estimating the unknown matrix $\mathbf{U}$ both methods utilize fourth moments. Defining next the matrices

$$\mathbf{C}^{ij}(\mathbf{x}_{st}) = \mathrm{E}\left\{(\mathbf{x}_{st}^\top \mathbf{e}_i)(\mathbf{x}_{st}^\top \mathbf{e}_j)\mathbf{x}_{st}\mathbf{x}_{st}^\top\right\} - \delta_{ij}\mathbf{I} - \mathbf{e}_i\mathbf{e}_j^\top - \mathbf{e}_j\mathbf{e}_i^\top, \tag{2}$$

where $\mathbf{e}_i$ is the ith member of the canonical basis of $\mathbb{R}^p$ — that is, $\mathbf{e}_i$ has all components equal to 0 except its ith component, which is 1, and δ_{ij} is the Kronecker delta. The set $\mathcal{C} = \{\mathbf{C}^{ij}(\mathbf{x}_{st}) : i, j \in \{1, \ldots, p\}\}$ collects every fourth cross-cumulant of the standardized random vector $\mathbf{x}_{st}$. It can be shown that under the model the unknown orthogonal matrix $\mathbf{U}^\top$ diagonalizes all matrices in the set $\mathcal{C}$ and JADE estimates $\mathbf{U}^\top$ by simultaneously (approximately) diagonalizing these matrices. FOBI can be viewed as a lighter version of JADE in that it only diagonalizes the single matrix $\sum_{i=1}^{p} \mathbf{C}^{ii}(\mathbf{x}_{st}) = \mathrm{E}(\mathbf{x}_{st}\mathbf{x}_{st}^\top\mathbf{x}_{st}\mathbf{x}_{st}^\top) - (p+2)\mathbf{I}$, which is the sum of a subset of members of $\mathcal{C}$. By this heuristic it seems reasonable to speculate that JADE outperforms FOBI, which indeed is generally the case: see, for example, [35]. Additionally, for JADE to be Fisher consistent it is sufficient that at most one of the β_i's is zero, whereas for FOBI to be Fisher consistent we need the stronger condition that all β_i are distinct. However, JADE pays for its advantages by being computationally much heavier than FOBI, and when a quick application of an independent component analysis method is needed, FOBI is often the first choice.

1.2. Independent component analysis and functional data

As the main contribution of this work we further extend on the functional independent component analysis proposed in [29] by considering not real-valued functions but instead functions that take values in the p-dimensional Euclidean space. That is, for each of the n observational units we observe p functions, not necessarily residing in the same function space. Data of this form are increasingly common nowadays and some areas of application include: electroencephalography (EEG) data where p electrodes are put on each of the n patients' scalps to measure the changes in the electrical activities of their brains in time, thus producing p functional observations for each patient; socio-economic time series data with a total of p indices measured for a group of n areas/countries repeatedly over time; and three-dimensional location data ($p = 3$) measured for n observational units over time, meaning that for each unit we observe three curves tracking the x, y, and z locations of the unit in time. Note that ICA is regularly applied to such data but the approaches are most often based on vectorizing the data tensor and applying some standard forms of ICA to the resulting samples of vectors. Two examples of this in the context of EEG data are the spatial and temporal ICA, see, e.g., [6], where the vectorization implies completely ignoring either the spatial or temporal information in the data, respectively. Such loss of information is avoided by our proposed method as it does not break the functional nature of the observations.

Although univariate functional data analysis is currently exceedingly popular, its multivariate counterpart has received in comparison relatively little attention in the literature. Some previous contributions to the field include: [3,11,16,21,39] discussed multivariate functional principal component analysis; [22,45,50–52] developed multivariate functional clustering; [28] discussed sufficient dimension reduction methodology where both the predictor and the response can be multivariate functions; [14,17,42] used different measures of outlyingness to identify multivariate functional outliers.

Consider next the conceptual and theoretical differences between multivariate-functional and univariate-functional extensions of independent component analysis. The two key aspects of independent component analysis are statistical independence and the notion of marginals. In a sense, the multivariate functional extension considered here is conceptually easier than the univariate functional extension developed in [29]. As observed in that paper, unlike in the classical setting, the univariate functional data do not have natural marginal random variables on which to perform independent component analysis. [29] tackled this issue by using the coefficients in the Karhunen–Loeve expansion

as the marginal random variables to prompt the process. Independent components are then defined in terms of these coefficients, see also [15]. For multivariate functional data, however, we can take a more straightforward route of simply treating the component functions as the marginals. In this context the independent component problem has the intuitively appealing objective of, given an observed multivariate random function, trying to extract another multivariate random function with independent component functions. These independent component functions can then be various latent processes, such as vital signs in the context of EEG-data. A finite-dimensional analogue for our problem is the independent subspace analysis [7], where we try to divide a larger space into a collection of smaller, independent subspaces. To sum up, the independent components in [29] are random variables, but the independent components in this paper are random functions. From this perspective, this paper is not an extension of [29], but instead an extension of the classical independent component analysis into a different direction.

Another earlier work combining ICA with infinite-dimensional spaces is Kernel ICA [1] where a contrast function based on a generalized notion of correlation is maximized through the use of reproducing kernel Hilbert spaces. However, the two methods, Kernel ICA and ours, are aimed for strictly different types of data and in some sense use the two dimensionalities in opposite ways. Kernel ICA operates on finite-dimensional data coming from the model (1) and uses infinite dimensionality as a tool to estimate the unmixing matrix Γ, whereas our proposed method is aimed for naturally function-valued data but uses finite-dimensional representations of the observations to achieve its goals, see Section 4. Thus, the infiniteness of the space can be seen as an auxiliary property of Kernel ICA and an essential property of our proposed method, making the two strictly differing lines of work.

In Section 2 we go briefly through the basics of functional analysis. The section also introduces the Cartesian product space $\mathcal{H}$ where our observed functions will reside in and a natural subclass of linear operators therein. Section 3 equips the space $\mathcal{H}$ with a suitable probability structure and, having defined what we mean by a random multivariate function $X \in \mathcal{H}$, defines the covariance matrix operator of X. The proposed methods of functional independent component analysis are described in Section 4 along with a proof of their Fisher consistency. In Section 5 we derive the coordinate representations for the sample versions of the methods and in Section 6 use them in a simulation study and in an application on the *uWave* hand gesture data set [30]. Finally, we close in Section 7 with some discussion and prospective ideas. The simulation and real data example were conducted with R [38] using the packages fda [40], ggplot2 [49], JADE [34] and MASS [46].

2. Theoretical framework

2.1. The Hilbert space $\mathcal{H}$ of vector-valued functions

We next review the basics of functional analysis, see [13] for a standard treatment. Let $T \subset \mathbb{R}$ be an interval and $(\mathcal{H}_i, \langle \cdot, \cdot \rangle_i)$, $i \in \{1, \ldots, p\}$, be separable Hilbert spaces of functions from T to $\mathbb{R}$. Furthermore, let $\mathcal{B}_i$ be the Borel σ-field generated by the open sets in $\mathcal{H}_i$ with respect to the metric induced by $\langle \cdot, \cdot \rangle_i$. Let $\mathcal{H}$ be the direct sum of $\mathcal{H}_1, \ldots, \mathcal{H}_p$; that is, $\mathcal{H} = \times_{i=1}^{p} \mathcal{H}_i$ is the Cartesian product of the individual spaces and the inner product in $\mathcal{H}$ is defined by $\langle f, g \rangle_{\mathcal{H}} = \langle f_1, g_1 \rangle_1 + \cdots + \langle f_p, g_p \rangle_p$, for any $f = (f_1, \ldots, f_p) \in \mathcal{H}$ and $g = (g_1, \ldots, g_p) \in \mathcal{H}$. Denoting the norms induced by the inner products $\langle \cdot, \cdot \rangle_{\mathcal{H}}, \langle \cdot, \cdot \rangle_1, \ldots, \langle \cdot, \cdot \rangle_p$ by $\| \cdot \|_{\mathcal{H}}, \| \cdot \|_1, \ldots, \| \cdot \|_p$, respectively, the relation $\|f\|_{\mathcal{H}}^2 = \|f_1\|_1^2 + \cdots + \|f_p\|_p^2$ is easily seen to hold for any $f = (f_1, \ldots, f_p) \in \mathcal{H}$. Furthermore, a natural σ-field in $\mathcal{H}$ is the product σ-field $\mathcal{B} = \mathcal{B}_1 \times \cdots \times \mathcal{B}_p$ generated by all measurable rectangles $B_1 \times \cdots \times B_p$ where $B_i \in \mathcal{B}_i$, $i \in \{1, \ldots, p\}$.

Being separable, each $\mathcal{H}_i$ admits a countable orthonormal basis, $\mathcal{E}_i = \{e_{ik}\}_{k=1}^{\infty}$. Using the component bases we construct an orthonormal basis in $\mathcal{H}$ as follows. Let e_{ik}^{+} denote the p-dimensional vector of functions whose components are 0 except for the ith component, which is e_{ik}. Then $\{e_{ik}^{+} : i \in \{1, \ldots, p\}, k \in \{1, 2, \ldots\}\}$ is an orthonormal basis of $\mathcal{H}$. This construction implies that the product space $\mathcal{H}$ is also separable. Throughout the paper any vector $f \in \mathcal{H}$ which has exactly one non-zero component will be called canonical, in relation to such a vector's resemblance to the canonical basis vectors in the Euclidean spaces.

Let $\mathcal{L}(\mathcal{H}_j, \mathcal{H}_i)$ be the set of all bounded linear operators from $\mathcal{H}_j$ to $\mathcal{H}_i$. Then for any i, j, $(\mathcal{L}(\mathcal{H}_j, \mathcal{H}_i), \| \cdot \|_{\text{OP},ij})$ is a Banach space where the operator norm $\| \cdot \|_{\text{OP},ij}$ is defined as

$$\|L_{ij}\|_{\text{OP},ij} = \sup_{f_j \neq 0} \left(\frac{\|L_{ij}f_j\|_i}{\|f_j\|_j} \right).$$

In the following we will use the notation $\| \cdot \|_{\text{OP}}$ for all possible operator norms and the context will always make clear which operator norm we mean. Similarly, I will be used to denote the identity operator of all considered spaces, the context again making the intended use clear. Recall also that for all $L_{ij} \in \mathcal{L}(\mathcal{H}_j, \mathcal{H}_i)$, there exists the adjoint operator L_{ij}^{*}, defined as the unique member of $\mathcal{L}(\mathcal{H}_i, \mathcal{H}_j)$ that satisfies $\langle L_{ij}f_j, f_i \rangle_i = \langle f_j, L_{ij}^{*}f_i \rangle_j$, for all $f_i \in \mathcal{H}_i$ and $f_j \in \mathcal{H}_j$.

Finally, define the tensor product $f_i \otimes f_j$ of $f_i \in \mathcal{H}_i$ and $f_j \in \mathcal{H}_j$ as the linear operator from $\mathcal{H}_j$ to $\mathcal{H}_i$ having the action $g_j \mapsto \langle f_j, g_j \rangle_j f_i$. Equivalent properties to those listed for tensor product operators from $\mathcal{H}$ to $\mathcal{H}$ in Lemma A.1 of [29] can also be proven for the tensor product operators from $\mathcal{H}_j$ to $\mathcal{H}_i$.

2.2. Matrices of bounded linear operators in $\mathcal{H}$

We next consider a natural subset of the set of all bounded linear operators from $\mathcal{H}$ to $\mathcal{H}$, constructed using bounded linear operators from the component spaces to each other. For a set of operators $\{L_{ij} \in \mathcal{L}(\mathcal{H}_j, \mathcal{H}_i) : i, j \in \{1, \ldots, p\}\}$, let L be the operator

$$\mathcal{H} \to \mathcal{H}, \quad (f_1, \ldots, f_p) \in \mathcal{H} \mapsto \left(\sum_{j=1}^p L_{1j}f_j, \ldots, \sum_{j=1}^p L_{pj}f_j \right). \tag{3}$$

Intuitively, we can identify L with the matrix of bounded linear operators,

$$L \equiv \begin{pmatrix} L_{11} & \cdots & L_{1p} \\ \vdots & \ddots & \vdots \\ L_{p1} & \cdots & L_{pp} \end{pmatrix},$$

so that the map in (3) can be formally regarded as matrix multiplication. We denote the class of all such operators as $\mathcal{L}(\mathcal{H}) = \times_{i,j=1}^p \mathcal{L}(\mathcal{H}_j, \mathcal{H}_i)$. The same construction was used in [26]. See also [43] and [25].

Using (3), it is easy to check that an operator $L \in \mathcal{L}(\mathcal{H})$ is also linear. Furthermore, using the Cauchy–Schwarz inequality and the operator norm inequality one can show that, for all $L \in \mathcal{L}(\mathcal{H})$ and $f \in \mathcal{H}$, we have

$$\|Lf\|_{\mathcal{H}} \leq \left(\sum_{i,j=1}^p \|L_{ij}\|_{\mathrm{OP}}^2 \right)^{1/2} \|f\|_{\mathcal{H}}.$$

That is, an operator $L \in \mathcal{L}(\mathcal{H})$ is also bounded. Thus, an operator $L \in \mathcal{L}(\mathcal{H})$ inherits both linearity and boundedness from its component operators L_{ij}. Consequently, being a bounded linear operator, any $L \in \mathcal{L}(\mathcal{H})$ admits the adjoint operator L^*. Using some algebra it is easily seen that the elements of the adjoint satisfy $(L^*)_{ij} = L_{ji}^*$, drawing an analogy to the Hermitian adjoint of a matrix in $\mathbb{C}^{p \times p}$.

Two useful subsets of $\mathcal{L}(\mathcal{H})$ are now readily defined. Call a member $L \in \mathcal{L}(\mathcal{H})$ a diagonal matrix of operators (or simply diagonal) if $L_{ij} = 0$ whenever $i \neq j$ and $L_{ii}^* = L_{ii}$. The simplest diagonal operator is the identity operator for which $L_{ii} = I$, $i \in \{1, \ldots, p\}$. Diagonal operators play later a central role in estimating solutions to the functional independent component model and as one of our key results we prove in Section 4 a connection between diagonal operators and canonical vectors. Finally, an element $U \in \mathcal{L}(\mathcal{H})$ is called unitary if $U^*U = UU^* = I$. Using the component representation it is easily seen that a sufficient and necessary condition for U to be unitary is

$$\sum_{k=1}^p U_{ik}U_{jk}^* = \Delta_{ij}, \quad i, j \in \{1, \ldots, p\},$$

where Δ_{ij} is the zero operator if $i \neq j$, and is the identity operator from $\mathcal{H}_i$ to $\mathcal{H}_j$ if $i = j$. This is a clear analogy for the orthonormality of the rows of a unitary matrix in $\mathbb{C}^{p \times p}$.

3. Probability structure on $\mathcal{H}$

3.1. Random elements in $\mathcal{H}$

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A random element in $\mathcal{H}_i$ is a function $X_i : \Omega \to \mathcal{H}_i$ that is $\mathcal{F}/\mathcal{B}_i$-measurable, $i \in \{1, \ldots, p\}$. Similarly, a random element in $\mathcal{H}$ is a function $X : \Omega \to \mathcal{H}$ that is $\mathcal{F}/\mathcal{B}$-measurable. A random element X in $\mathcal{H}$ can thus be thought of as a random function $X(\cdot) = (X_1(\cdot), \ldots, X_p(\cdot))$, where X_i resides in $\mathcal{H}_i$, $i \in \{1, \ldots, p\}$. For the basic theory of random variables in function spaces see [5].

In the following, we denote the set of all mth power integrable random elements in $\mathcal{H}$ by $\mathcal{X}^m(\mathcal{H})$, that is,

$$\mathcal{X}^m(\mathcal{H}) = \{(X : \Omega \to \mathcal{H}) : \mathrm{E}\left(\|X\|_{\mathcal{H}}^m\right) < \infty\}.$$

It is easily seen that requiring $X \in \mathcal{X}^2(\mathcal{H})$ or $X \in \mathcal{X}^4(\mathcal{H})$ is equivalent to requiring the component functions to respectively satisfy $X_i \in \mathcal{X}^2(\mathcal{H}_i)$ or $X_i \in \mathcal{X}^4(\mathcal{H}_i)$, $i \in \{1, \ldots, p\}$.

Next, define a random operator W_{ij} to be a mapping $W_{ij} : \Omega \to \mathcal{L}(\mathcal{H}_j, \mathcal{H}_i)$ that is $\mathcal{F}/\mathcal{B}_{\mathrm{OP}}$-measurable where $\mathcal{B}_{\mathrm{OP}}$ is the Borel σ-field generated by the open sets of $\mathcal{L}(\mathcal{H}_j, \mathcal{H}_i)$ with respect to the metric induced by the operator norm $\|\cdot\|_{\mathrm{OP}}$. If W_{ij} is a random operator with $\mathrm{E}\left(\|W_{ij}\|_{\mathrm{OP}}\right) < \infty$, then the bivariate map $(f_i, f_j) \mapsto \mathrm{E}\left(\langle f_i, W_{ij}f_j \rangle_i\right)$ is a bounded bilinear form and can be shown to induce a unique operator $A_{ij} \in \mathcal{L}(\mathcal{H}_j, \mathcal{H}_i)$ satisfying $\langle f_i, A_{ij}g_j \rangle_i = \mathrm{E}\langle f_i, W_{ij}g_j \rangle_i$ for all $f_i \in \mathcal{H}_i$ and $g_j \in \mathcal{H}_j$. We define the expected value of W_{ij} to be this operator, $\mathrm{E}\left(W_{ij}\right) = A_{ij}$

Using the previous we are now sufficiently equipped to define the first two moments, the mean function and the covariance matrix operator, of a random element $X \in \mathcal{H}$.

3.2. The covariance matrix operator Σ_{XX}

Assume next that $X \in \mathcal{X}^2(\mathcal{H})$. The expected values $\mathrm{E}(X_i) = \mu_i \in \mathcal{H}_i$, $i \in \{1, \ldots, p\}$, are readily defined as the Riesz representation of the bounded linear functional

$$\mathcal{H}_i \to \mathbb{R}, \quad f_i \mapsto \mathrm{E}\left(\langle f_i, X_i \rangle_i\right).$$

Using the component-wise expected values μ_i we further define the expected value of the random element X to be the function $\mu = (\mu_1, \ldots, \mu_p) \in \mathcal{H}$. As we can always center our observed data, it is not restricting to assume that $\mu = 0$, as we will do for the remainder of this work.

Consider then the random operator $(X_i \otimes X_j)$. Using the Cauchy–Schwarz inequality we have

$$\mathrm{E}\left(\|X_i \otimes X_j\|_{\mathrm{OP}}\right) \leq \left\{\mathrm{E}\left(\|X_i\|_i^2\right)\mathrm{E}\left(\|X_j\|_j^2\right)\right\}^{1/2},$$

the right-hand side of which is finite due to our assumption on square integrability. The random operator $(X_i \otimes X_j)$ thus induces the unique, bounded linear operator, $\Sigma_{X_i X_j} = \mathrm{E}(X_i \otimes X_j)$, the cross-covariance operator [2] between X_i and X_j. Using the definition of the expected value of a random operator one can further show that the adjoint operator of $\Sigma_{X_i X_j}$ is $\Sigma_{X_i X_j}^* = \Sigma_{X_j X_i}$.

Using the p^2 bounded linear operators $\Sigma_{X_i X_j}$ we next construct the covariance matrix operator $\Sigma_{XX} \in \mathcal{L}(\mathcal{H})$ as

$$\Sigma_{XX} \equiv \begin{pmatrix} \Sigma_{X_1 X_1} & \cdots & \Sigma_{X_1 X_p} \\ \vdots & \ddots & \vdots \\ \Sigma_{X_p X_1} & \cdots & \Sigma_{X_p X_p} \end{pmatrix},$$

It is easily seen that, for $f = (f_1, \ldots, f_p) \in \mathcal{H}$, we have the equality $f \otimes f = (f_i \otimes f_j)_{i,j=1}^p$ and the covariance matrix operator Σ_{XX} can then be written compactly as $\mathrm{E}(X \otimes X)$. This type of matrices of covariance operators were also used in [27] and [44].

Remark 1. For clarity we use two different notations for the covariance matrix operator of a random function $X \in \mathcal{X}^2(\mathcal{H})$: when it is understood as a bounded linear operator in $\mathcal{H}$ we use the notation Σ_{XX}; when it is understood as the mapping $\mathcal{X}^2(\mathcal{H}) \to \mathcal{L}(\mathcal{H})$, $X \mapsto \Sigma_{XX}$, we use the notation Σ.

Recall next four key properties of the ordinary covariance matrix $\mathrm{cov}(\mathbf{x})$ of a square-integrable random vector $\mathbf{x} = (x_1, \ldots, x_p)^\top$: (i) self-adjointness (symmetry), $\mathrm{cov}(\mathbf{x}) = \mathrm{cov}(\mathbf{x})^\top$, (ii) positive-semidefiniteness, for any $\mathbf{a} \in \mathbb{R}^p$ we have $\mathbf{a}^\top \mathrm{cov}(\mathbf{x})\mathbf{a} \geq 0$, (iii) affine equivariance, for any invertible matrix $\mathbf{A} \in \mathbb{R}^{p \times p}$ the covariance matrix transforms as $\mathrm{cov}(\mathbf{Ax}) = \mathbf{A}\mathrm{cov}(\mathbf{x})\mathbf{A}^\top$ and (iv) full independence property, if x_i and x_j are independent then $\mathrm{cov}(\mathbf{x})_{ij} = 0$. Not surprisingly, it turns out that all of these properties are shared also by the covariance matrix operator Σ_{XX}, as described in the following lemma.

Lemma 1. *Assuming $X \in \mathcal{X}^2(\mathcal{H})$, the covariance matrix operator $\Sigma_{XX} \in \mathcal{L}(\mathcal{H})$ has the following properties:*

(i) *It is a self-adjoint, non-negative, trace-class operator and as such admits a spectral decomposition with the associated orthonormal basis $\{\phi_k\}_{k=1}^\infty$.*
(ii) *As a mapping $\Sigma : \mathcal{X}^2(\mathcal{H}) \to \mathcal{L}(\mathcal{H})$, the covariance matrix operator is affine equivariant in the sense that $\Sigma(AX) = A\Sigma(X)A^*$ for any invertible bounded linear operator $A \in \mathcal{L}(\mathcal{H})$.*
(iii) *If X_i and X_j are independent, $\Sigma_{X_i X_j} = 0$.*

These properties were established in [29] for the case of univariate X.

Remark 2. A stronger version of the affine equivariance can be shown to hold. Let $A = (A_{ij})_{i=1,j=1}^{k,\ p}$ where A_{ij} is a linear operator from $\mathcal{H}_j$ to some suitable Hilbert space $\mathcal{G}_i$, and k is any positive integer. Then we still have $\Sigma(AX) = A\Sigma(X)A^*$, a property that is in $\mathbb{R}^p$ called full affine equivariance.

Part (i) of Lemma 1 guarantees the existence of the spectral decomposition of Σ_{XX} into a sum of rank-1 operators:

$$\Sigma_{XX} = \sum_{k=1}^\infty \lambda_k(\phi_k \otimes \phi_k),$$

where (ϕ_k, λ_k) are eigenvector–eigenvalue pairs, $\{\phi_k\}_{k=1}^\infty$ is an orthonormal basis of $\mathcal{H}$ and the eigenvalues satisfy $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$. This representation will be used next to define the independent component model in $\mathcal{H}$.

4. Independent component analysis in $\mathcal{H}$

4.1. Independent component model in $\mathcal{H}$

We say that $X \in \mathcal{X}^2(\mathcal{H})$ follows the $\mathcal{H}$-valued independent component model if there exists a matrix of operators $\Gamma = (\Gamma_{ij})_{i,j=1}^p \in \mathcal{L}(\mathcal{H})$ such that

$$\Gamma X = Z, \tag{4}$$

where $Z = (Z_1, \ldots, Z_p)$ is a random element in $\mathcal{H}$ having mutually independent component functions. Two random elements $X : \Omega \to \mathcal{G}_1$ and $Y : \Omega \to \mathcal{G}_2$, not necessarily having values in the same space, are defined to be independent if $\mathrm{E}\{q_1(X)q_2(Y)\} = \mathrm{E}\{q_1(X)\}\mathrm{E}\{q_2(Y)\}$ for all measurable functionals $q_1 : \mathcal{G}_1 \to \mathbb{R}$, $q_2 : \mathcal{G}_2 \to \mathbb{R}$. Actually, a weaker condition requiring the previous to hold only for all linear q_1, q_2 would also be enough for us as we use the condition only in the context of inner products.

While the case of identical component spaces, $\mathcal{H}_1 = \cdots = \mathcal{H}_p$, is most likely sufficient in practice, the following theoretical framework works equally well for an arbitrary choice of component spaces. However, we note that this freedom of choice can lead to mathematically nonintuitive scenarios. Consider, e.g., the following example pointed out by a reviewer: let $p = 2$ and let $\mathcal{H}_1$ be an L^2 space (on a suitable real interval T) and $\mathcal{H}_2$ be a Sobolev space of some order. Then the component operator Γ_{21} has the curious property of mapping possibly discontinuous functions into differentiable ones.

The objective in the $\mathcal{H}$-valued independent component analysis is to estimate some unmixing operator Γ such that ΓX has independent component functions. Like its vector-valued analogy in (1), the operator Γ is not uniquely defined. If one applies to both sides of (4) any diagonal operator $D \in \mathcal{L}(\mathcal{H})$, the right-hand side still retains independent component functions. This implies that, without further assumptions, we cannot find any unique functional form for the component functions. Indeed, as we show later in this section, our methods actually estimate $\{B_j Z_j\}_{j=1}^p$ where $B_j \in \mathcal{L}(\mathcal{H}_j), j \in \{1, \ldots, p\}$. However, this identifiability issue does not affect our goal of discovering independent components, as the resulting vector of functions has independent component functions regardless of the form of D.

We will next approach the problem by extending two methods of vector-valued independent component analysis, FOBI and JADE, to the case of vector-valued random functions.

4.2. Standardization of a random vector-valued function

The first step in vector-valued independent component analysis is the standardization of $\mathbf{x}$ by the inverse square root of the covariance matrix cov($\mathbf{x}$). This results in a standardized random vector which is then rotated in a specific way to obtain the independent components, see Section 1. The same strategy, however, does not work in the current context for the simple fact that no standardized elements, such as $\mathcal{N}(\mathbf{0}, \mathbf{I}_p)$, exist in an infinite-dimensional space. A further complication is that, as a bounded operator, the covariance operator has an unbounded inverse, and while a suitable regularization could be used to solve this issue, it would not help with the first one (non-existence of standardized elements). Instead, as in [29], we resort to an additional assumption. Let $\{\phi_k\}_{k=1}^\infty$ be the orthonormal basis of $\mathcal{H}$ consisting of the eigenvectors of Σ_{XX} in decreasing order according to the corresponding eigenvalues. For a fixed $d \in \mathbb{N}$, let $\mathcal{M}_d = \mathrm{span}(\{\phi_k\}_{k=1}^d)$ be the subspace of $\mathcal{H}$ spanned by the d first eigenvectors of Σ_{XX}. The simplifying assumption we make is the following.

Assumption 1. The component functions of X are dependent only along the d orthogonal directions $\{\phi_k\}_{k=1}^d$. That is, if $R_d = \sum_{k=d+1}^\infty \langle X, \phi_k \rangle_\mathcal{H} \phi_k$, then the p components of R_d are independent.

Note that Assumption 1 is not equivalent to assuming that the Hilbert space $\mathcal{H}$ is finite-dimensional. Rather, all we are assuming is that the components of interest are contained in a finite-dimensional space. In vector-valued independent component analysis this assumption is naturally always satisfied by picking simply $d = p$. A heuristic interpretation for the assumption in the current case is that the majority of the structure of the independent component functions is noise, meaning that the signal in the function Z is in some sense finite-dimensional and our objective is to separate it from the internal, infinite-dimensional noise space. Moreover, Assumption 1 is very reasonable in practice in the sense that higher order principal components (ϕ_k corresponding to large values of k) are typically of high frequency and difficult to estimate with any accuracy in finite samples. In this sense, assuming that the interesting components are contained in the space spanned by the first d eigenfunctions accounts to assuming that the interesting components are estimable with reasonable accuracy.

It would also in principle be possible to let $d = d_n$ depend on n and approach infinity as $n \to \infty$, and work out asymptotic results under this regime. This will, however, be pursued in a separate work.

Note that $\{\phi_k\}$ may not span the entire $\mathcal{H}$. However, by definition they are guaranteed to span $\overline{\mathrm{ran}}(\Sigma_{XX})$, the closure of the range space of Σ_{XX}. Because Σ_{XX} is self-adjoint, $\overline{\mathrm{ran}}(\Sigma_{XX})^\perp = \ker(\Sigma_{XX})$, the kernel space of Σ_{XX}. Meanwhile, for any $f \in \ker(\Sigma_{XX})$, we have

$$\langle f, \Sigma_{XX} f \rangle_\mathcal{H} = \mathrm{var}\{\langle f, X \rangle_\mathcal{H}\} = 0,$$

which implies that $\langle f, X \rangle_{\mathcal{H}} =$ constant almost surely. Since this holds for the special case $X(\omega) = 0$, we have $\langle f, X \rangle_{\mathcal{H}} = 0$ almost surely. This means f is orthogonal to the support of X. Since such functions are of no interest to us, we can, without loss of generality, reset $\mathcal{H}$ to be $\overline{\text{ran}}(\Sigma_{XX})$, as we will do for the rest of the paper.

For an arbitrary subspace $\mathcal{M} \subset \mathcal{H}$, let $P_{\mathcal{M}}$ and $Q_{\mathcal{M}}$ denote the orthogonal projections on to $\mathcal{M}$ and $\mathcal{M}^{\perp}$, respectively. Then Assumption 1 says that we can without loss of generality consider the projections $X^{(d)} = P_{\mathcal{M}_d} X$ instead of the original observations X. This simplifies the model (4) to the form

$$\Gamma_0 X^{(d)} = Z^{(d)}, \tag{5}$$

where $X^{(d)}, Z^{(d)}$ are random functions in $\mathcal{M}_d$, the component functions of $Z^{(d)}$ are independent and $\Gamma_0 \in \mathcal{L}(\mathcal{M}_d)$ is assumed to be invertible.

Remark 3. Later in this section the proposed methods are shown to be Fisher consistent, meaning that under the model (4) and Assumption 1 the final independent component scores are invariant to injective transformations $X \mapsto (P_{\mathcal{M}_d} A P_{\mathcal{M}_d} + Q_{\mathcal{M}_d}) X$, where $A \in \mathcal{L}(\mathcal{M}_d)$. However, as our estimation methods crucially depend on the existence of a random function Z it is not meaningful to speak of affine equivariance outside the model in the same general sense that holds for both vector-valued FOBI and JADE, see [35].

With this, we are now ready to present the first step towards the estimation of Z, an analogy for Lemma 6.4 in [29]. In the following, for a self-adjoint linear operator $A \in \mathcal{L}(\mathcal{M}_d)$, let $A^{-1/2}$ denote its self-adjoint inverse square root within $\mathcal{M}_d$, that is, $A^{-1/2} A A^{-1/2} = P_{\mathcal{M}_d}$.

Lemma 2. *Assume that $X \in \mathcal{X}^2(\mathcal{H})$ follows the model (5). Then*

$$\Sigma(X^{(d)})^{-1/2} X^{(d)} = U_0 \Sigma(Z^{(d)})^{-1/2} Z^{(d)},$$

for some unitary operator $U_0 \in \mathcal{L}(\mathcal{M}_d)$.

The standardized functions are in the following denoted by $\tilde{X} = \Sigma(X^{(d)})^{-1/2} X^{(d)}$ and $\tilde{Z} = \Sigma(Z^{(d)})^{-1/2} Z^{(d)}$ and naturally satisfy $\Sigma(\tilde{X}) = \Sigma(\tilde{Z}) = P_{\mathcal{M}_d}$. The next step towards finding Z is the estimation of the unknown unitary operator U_0 in Lemma 2. As described in the introduction both FOBI and JADE approach it via matrices of fourth cross-cumulants and before continuing we first define operatorial counterparts for them.

4.3. The fourth cross-cumulant operators $C^{ij}(X)$

In this section we assume that the zero-mean random function $X \in \mathcal{X}^4(\mathcal{M}_d)$ resides in the d-dimensional space $\mathcal{M}_d$ spanned by the fixed orthonormal basis $\{\phi_k\}_{k=1}^d$. We define the (i, j)th fourth cross-cumulant of X with respect to the basis $\{\phi_k\}_{k=1}^d$ to be

$$\begin{aligned}
C^{ij}(X) =& \mathrm{E}\left\{ \langle X, \phi_i \rangle_{\mathcal{H}} \langle X, \phi_j \rangle_{\mathcal{H}} (X \otimes X) \right\} - \mathrm{E}\left\{ \langle X', \phi_i \rangle_{\mathcal{H}} \langle X', \phi_j \rangle_{\mathcal{H}} (X \otimes X) \right\} \\
& - \mathrm{E}\left\{ \langle X', \phi_i \rangle_{\mathcal{H}} \langle X, \phi_j \rangle_{\mathcal{H}} (X' \otimes X) \right\} - \mathrm{E}\left\{ \langle X', \phi_i \rangle_{\mathcal{H}} \langle X, \phi_j \rangle_{\mathcal{H}} (X \otimes X') \right\},
\end{aligned} \tag{6}$$

where $i, j \in \{1, \dots, d\}$ and the random function X' is an independent copy of X. Repeated application of the Cauchy–Schwarz inequality shows that, for example, the first term in (6) satisfies

$$\mathrm{E}\left\{ \| \langle X, \phi_i \rangle_{\mathcal{H}} \langle X, \phi_j \rangle_{\mathcal{H}} (X \otimes X) \|_{\mathrm{OP}} \right\} \leq \mathrm{E}\left(\|X\|_{\mathcal{H}}^4 \right) < \infty,$$

implying that the first term of (6) exists as a uniquely defined bounded linear operator in $\mathcal{M}_d$. Similar considerations for the other terms show that the operator $C^{ij}(X) \in \mathcal{L}(\mathcal{M}_d)$ is then well-defined. Our main interest is in standardized random functions, $\Sigma(X) = P_{\mathcal{M}_d}$, and the following lemma provides a simplified form for (6) in that case.

Lemma 3. *Let the zero-mean random function $X \in \mathcal{X}^4(\mathcal{M}_d)$ satisfy $\Sigma(X) = P_{\mathcal{M}_d}$. Then we have*

$$C^{ij}(X) = \mathrm{E}\left\{ \langle X, \phi_i \rangle_{\mathcal{H}} \langle X, \phi_j \rangle_{\mathcal{H}} (X \otimes X) \right\} - \delta_{ij} P_{\mathcal{M}_d} - \phi_i \otimes \phi_j - \phi_j \otimes \phi_i.$$

The operator C^{ij} in Lemma 3 closely resembles the cross-cumulant matrix (2) for standardized random vectors and is next shown to serve similar purposes in constructing our versions of FOBI and JADE in $\mathcal{H}$.

Theorem 1. *Assume that $Z \in \mathcal{X}^4(\mathcal{M}_d)$ has independent component functions and that $\Sigma(Z) = P_{\mathcal{M}_d}$. Then we have for any unitary matrix of operators $U = (U_{k\ell})_{k,\ell=1}^p \in \mathcal{L}(\mathcal{M}_d)$ and for any $i, j \in \{1, \dots, d\}$:*

$$C^{ij}(UZ) = U D^{ij} U^*,$$

where $D^{ij} = D^{ij}(U, Z)$ is a diagonal matrix of operators with the diagonal operators

$$D_{kk}^{ij} = \mathrm{E}\left\{ (Z_k \otimes Z_k)(\xi_{ik} \otimes \xi_{jk})(Z_k \otimes Z_k) \right\} - \langle \xi_{ik}, \xi_{jk} \rangle_k P_k - (\xi_{ik} \otimes \xi_{jk}) - (\xi_{jk} \otimes \xi_{ik}), \tag{7}$$

for $k \in \{1, \dots, p\}$, where $\xi_i = (\xi_{i1}, \dots, \xi_{ip}) = U^* \phi_i$ and P_k is the projection operator from the kth component space of $\mathcal{H}$ to the kth component space of $\mathcal{M}_d$.

Theorem 1 essentially says that U^* diagonalizes (as in a diagonal operator) the operator $C^{ij}(UZ)$ for every choice of $i, j \in \{1, \dots, d\}$ and these decompositions provide us a mean of finding the missing unitary operator U. Our version of JADE will later utilize all p^2 of these operators and for FOBI we use just a subset of them, captured by the FOBI-operator $C(X) \in \mathcal{L}(\mathcal{M}_d)$,

$$C(X) = \sum_{i=1}^{d} C^{ii}(X). \tag{8}$$

The next theorem gives some useful properties of this operator.

Theorem 2. *Assume that $Z \in \mathcal{X}^4(\mathcal{M}_d)$ has independent component functions and that $\Sigma(Z) = P_{\mathcal{M}_d}$. Then, for any unitary matrix of operators $U = (U_{k\ell})_{k,\ell=1}^{p} \in \mathcal{L}(\mathcal{M}_d)$, the FOBI-operator (8) satisfies*

$$C(UZ) = UC(Z)U^* = UDU^*,$$

where $D = C(Z)$ is a diagonal matrix of operators with the diagonal entries

$$D_{kk} = \mathrm{E}\left\{(Z_k \otimes Z_k)^2\right\} - (d_k + 2)P_k,$$

where $d_k = \dim\{span(\{\phi_{mk}\}_{m=1}^{d})\}$, the dimension of the kth component space, and P_k is the projection operator from the kth component space of $\mathcal{H}$ to the kth component space of $\mathcal{M}_d$.

The first equality in Theorem 2 does not need the independence of the component functions of Z but actually holds for all standardized $Z \in \mathcal{X}^4(\mathcal{M}_d)$, as long as the operator U is unitary. Thus the functional $C : \mathcal{X}^4(\mathcal{M}_d) \to \mathcal{L}(\mathcal{M}_d)$ is unitary equivariant.

Recall from the introduction that in FOBI we diagonalize a single matrix and in JADE multiple matrices simultaneously. The functional analogy for the former is the spectral decomposition of $C(X)$ and for the latter we define next the joint diagonalization of a set of operators. Namely, define the joint diagonalizer of a finite set of operators, $\mathcal{S} = \{S_i \mid S_i \in \mathcal{L}(\mathcal{M}_d), i \in \{1, \dots, I\}\}$, to be the orthonormal basis $\{\psi_k\}_{k=1}^{d}$ of $\mathcal{M}_d$ that maximizes the objective function

$$w(\psi_1, \dots, \psi_d) = \sum_{i=1}^{I} \sum_{k=1}^{d} \langle \psi_k, S_i \psi_k \rangle_{\mathcal{H}}^2. \tag{9}$$

In the previous paragraphs we have discussed two kinds of diagonality, the diagonality in the sense of diagonal operators in Theorems 1 and 2 and the diagonality in the sense of the spectral decomposition. The final tool we need for the estimation of the independent functions is a connection between these two concepts. Recall that by a canonical vector we mean any element of $\mathcal{H}$ which has at most one non-zero component. The needed connection is now provided by the next pair of lemmas which show that the spectral decomposition and joint diagonalization of diagonal operators mimic the eigendecomposition and joint diagonalization of diagonal real matrices in the sense that the spectral decompositions and the joint diagonalizer of a set of diagonal operators consist entirely of canonical vectors.

Lemma 4. *Let $D \in \mathcal{L}(\mathcal{H})$ be a diagonal matrix of operators with finite rank d and let its spectral decomposition be*

$$D = \sum_{k=1}^{d} \tau_k(\psi_k \otimes \psi_k)$$

where the eigenvalues $\{\tau_k\}_{k=1}^{d}$ are distinct. Then the eigenvectors $\{\psi_k\}_{k=1}^{d}$ are canonical.

Lemma 5. *Let $\mathcal{S} = \{S_i\}_{i=1}^{I}$ be a finite collection of bounded linear operators in $\mathcal{M}_d$ and let $\{\psi_k\}_{k=1}^{d}$ be an orthonormal basis of $\mathcal{M}_d$. Then we have*

$$w(\psi_1, \dots, \psi_d) \leq \sum_{i=1}^{I} \|S_i\|_{HS}^2,$$

where $\| \cdot \|_{HS}$ is the Hilbert–Schmidt norm and an equality is reached if and only if each ψ_k is an eigenvector of each S_i, $k \in \{1, \dots, d\}, i \in \{1, \dots, I\}$. In particular, if all operators in $\mathcal{S}$ are diagonal and share an eigenbasis then the elements of the joint diagonalizer are canonical.

4.4. Finding the unitary transformation U_0

Using the previously defined fourth cross-cumulant operators we next formulate the functional counterparts for the steps taken in vector-valued FOBI and JADE to estimate the orthogonal matrix **U**.

Definition 1. Let $X \in \mathcal{X}^4(\mathcal{H})$ follow the model (5). Then we define

(i) FOBI-basis of X is the set $\{\psi_k^F\}_{k=1}^d$ of eigenfunctions of the FOBI-operator $C(\tilde{X})$,
(ii) JADE-basis of X is the joint diagonalizer $\{\psi_k^J\}_{k=1}^d$ of the set of operators $\mathcal{C} = \{C^{ij}(\tilde{X})\}_{i,j=1}^d$.

In the next theorem Lemmas 4 and 5 are applied respectively to the FOBI-basis and JADE-basis to find U_0. However, to guarantee consistency we need to make some additional assumptions which guarantee that the eigenbases are unique up to signs and order. For the FOBI-solution we need the following.

Assumption 2. The eigenvalues of $C(\tilde{Z})$ are distinct.

One consequence of Assumption 2 is that FOBI cannot estimate two latent functions having the same distribution. For JADE the corresponding assumption is much more relaxed but to use Lemma 5 we first need the additional assumption that all the diagonal operators in Theorem 1 share a common eigenbasis.

Assumption 3. The operators $D^{ij}(U_0, \tilde{Z})$, $i, j \in \{1, \ldots, p\}$, have a common eigenbasis.

While this sounds somewhat stringent, in Section 5.2 we explain that Assumption 3 is in fact not that strict, and is satisfied under some general conditions and choices of d. The need for the next assumption guaranteeing the uniqueness of the eigenbasis for JADE now follows directly from the equality condition in Lemma 5.

Assumption 4. For each pair (ψ_k^J, ψ_ℓ^J), $k, \ell \in \{1, \ldots, d\}$, there exists a pair (i, j), $i, j \in \{1, \ldots, d\}$, such that the eigenvalues of $D^{ij}(U_0, \tilde{Z})$ related to ψ_k^J and ψ_ℓ^J are distinct.

The next theorem is our main result and shows how the FOBI-basis and JADE-basis can be used to estimate the independent component functions.

Theorem 3. *Let $X \in \mathcal{X}^4(\mathcal{H})$ follow the model (5) and let $\{\psi_k^F\}_{k=1}^d$ and $\{\psi_k^J\}_{k=1}^d$ be the FOBI-basis and JADE-basis of X, respectively. Assume further that either Assumption 2 (FOBI) or Assumptions 3 and 4 (JADE) are satisfied. Then the FOBI and JADE estimators of the latent functions are respectively the d elements of $\mathcal{X}^4(\mathcal{H})$ given as*

$$\hat{Z}_k^F = (\psi_k^F \otimes \psi_k^F)\Sigma(X^{(d)})^{-1/2}X^{(d)}, \quad k \in \{1, \ldots, d\}, \qquad \hat{Z}_k^J = (\psi_k^J \otimes \psi_k^J)\Sigma(X^{(d)})^{-1/2}X^{(d)}, \quad k \in \{1, \ldots, d\},$$

where each estimator $\hat{Z}_k^\bullet$ corresponds to exactly one latent function Z_j.

Let F be the distribution of an arbitrary $X \in \mathcal{X}^4(\mathcal{H})$ and let $\Gamma_k^J(F) := (\psi_k^J \otimes \psi_k^J)\Sigma(X^{(d)})^{-1/2}$ denote the kth JADE "unmixing operator" used in Theorem 3, treated as a functional of the distribution F. Then, Theorem 3 states that the collection of estimators $\Gamma_1^J, \ldots, \Gamma_d^J$ is Fisher consistent in the following sense: For $X \in \mathcal{X}^4(\mathcal{H})$ coming from the model (5) and satisfying Assumptions 3 and 4, the population-level transformations $\Gamma_1^J(F)X^{(d)}, \ldots, \Gamma_d^J(F)X^{(d)}$ yield the functions $\hat{Z}_1^J, \ldots, \hat{Z}_d^J$ which serve as "estimators" for the latent components Z_j. By an "estimator" we mean that each recovered function $\hat{Z}_k^J$ corresponds to a single latent component Z_j, implying that the method succeeds in eliminating the dependencies between the component functions of X and recovers independent functions, as desired. An analogous reasoning holds for FOBI.

More specifically, for each $k \in \{1, \ldots, d\}$ the procedure actually recovers the p-variate function $\hat{Z}_k^\bullet = \langle \psi_k^\bullet, \tilde{X} \rangle_{\mathcal{H}} \psi_k^\bullet = \langle h_k^\bullet, \tilde{Z} \rangle_{\mathcal{H}} \psi_k^\bullet$ where the only dependency on the latent function Z is through the inner product $\langle \psi_k^\bullet, \tilde{X} \rangle_{\mathcal{H}} = \langle h_k^\bullet, \tilde{Z} \rangle_{\mathcal{H}}$. Furthermore, every $h_k^\bullet$ is canonical, meaning that each of the estimates $\hat{Z}_k^\bullet$ contains information on exactly one latent component Z_j and this information is entirely contained in the single inner product, $\langle \psi_k^\bullet, \tilde{X} \rangle_{\mathcal{H}}$. In the following we will refer to these inner products as the independent component scores. As more than one score can be related to a single latent function Z_j, the d-vector of independent component scores can further be divided into m mutually independent subvectors, $\mathbf{Z}_{(\ell)} \in \mathbb{R}^{d_\ell}$, $\sum_{\ell=1}^m d_\ell = d$, so that each subvector corresponds to a single latent function Z_j.

5. The methods in practice

5.1. Sample versions of the methods

For deriving the sample version of the proposed method we make the simplifying assumption that the component spaces are the same, $\mathcal{H}_1 = \cdots = \mathcal{H}_p$. The generalization to the case of different component spaces follows easily.

Let $X^1, \ldots, X^n$ be a random sample of X. Here, we use superscript to represent the position in a sample, to differentiate from the subscript in X_i which represents the ith component of X. Furthermore, let X_{ij} represent the jth component of X^i. Although our theory is based on infinite-dimensional spaces, our observations are always finite-dimensional and so let $X_{ij}(t_{m,ij})$ denote the value of the jth component function of the ith observation at the time point $t_{m,ij}$, $m \in \{1, \ldots, M_{ij}\}$, $j \in \{1, \ldots, p\}$, $i \in \{1, \ldots, n\}$. We thus allow the measurement times and the numbers of measurements to differ across both observations and components. The underlying assumption in functional data analysis is that the observed

values $X_{ij}(t_{m,ij})$ correspond to latent (smooth) functions that we observe only at the discrete times $t_{m,ij}$. The first step in implementing the method is thus to express all the observations as functions using some suitable basis.

For approximating the space $\mathcal{H}$, fix a K-element basis $\mathcal{G}_0 = \{g_k\}_{k=1}^K$, the span of which we denote as $\mathcal{M}_0$. The functional approximations $\hat{x}_{ij}(t) = \sum_{k=1}^K \hat{c}_{ijk} g_k(t)$ of the observed curves in $\mathcal{M}_0$ can be found as

$$(\hat{c}_{ij1}, \ldots, \hat{c}_{ijK})^\top = \operatorname{argmin} \sum_{m=1}^{M_{ij}} \left\{ X_{ij}(t_{m,ij}) - \sum_{k=1}^K c_{ijk} g_k(t_{m,ij}) \right\}^2,$$

which is a least-squares type problem. Having estimated the coordinates $\hat{c}_{ijk}$ we denote in the following the coordinate vector of X_{ij} in the basis $\mathcal{G}_0$ as $[X_{ij}]_{\mathcal{G}_0} = (c_{ij1}, \ldots, c_{ijK})^\top \in \mathbb{R}^K$. Consider then the pK-dimensional product space $\mathcal{M} = \mathcal{M}_0 \times \cdots \times \mathcal{M}_0$. The space $\mathcal{M}$ then has the natural direct sum basis $\mathcal{G} = \mathcal{G}_0 \oplus \cdots \oplus \mathcal{G}_0$. The stacked vector of the coordinates of all p component functions of X^i in the basis $\mathcal{G}$ is denoted by $[X^i]_{\mathcal{G}} = ([X_{i1}]_{\mathcal{G}_0}^\top, \ldots, [X_{ip}]_{\mathcal{G}_0}^\top)^\top \in \mathbb{R}^{pK}$ and the matrix of all coordinates of all observations by $[X]_{\mathcal{G}} = ([X^1]_{\mathcal{G}}, \ldots, [X^n]_{\mathcal{G}})^\top \in \mathbb{R}^{n \times pK}$. We assume without loss of generality that the coordinate representations of the observations are centered, $\sum_{i=1}^n [X_{ij}]_{\mathcal{G}_0} = \mathbf{0}, j \in \{1, \ldots, p\}$.

Let $\mathbf{G}_{\mathcal{G}} = (\langle g_k, g_{k'} \rangle_{\mathcal{H}})_{k,k'=1}^K$ denote the Gram matrix of a basis $\mathcal{G} = \{g_k\}_{k=1}^K$. For orthonormal bases the Gram matrix equals the identity matrix and if $\mathcal{G}$ is a direct sum basis $\mathcal{G} = \mathcal{G}_0 \oplus \cdots \oplus \mathcal{G}_0$ then clearly $\mathbf{G}_{\mathcal{G}} = \operatorname{diag}(\mathbf{G}_{\mathcal{G}_0}, \ldots, \mathbf{G}_{\mathcal{G}_0}) = (\mathbf{I}_p \otimes \mathbf{G}_{\mathcal{G}_0})$, where $\mathbf{G}_{\mathcal{G}_0}$ is the Gram matrix of the basis $\mathcal{G}_0$ and $\otimes$ is the Kronecker product between matrices. The next theorem now describes how the coordinate representations can be used to carry out the proposed methods in practice.

Theorem 4. *Let $[\hat{\Phi}]_{\mathcal{G}} \in \mathbb{R}^{pK \times d}$ contain the d first eigenvectors of the matrix $(1/n)[X]_{\mathcal{G}}^\top [X]_{\mathcal{G}} (\mathbf{I}_p \otimes \mathbf{G}_{\mathcal{G}_0})$ and let the diagonal matrix $\hat{\Lambda}_d \in \mathbb{R}^{d \times d}$ hold the corresponding eigenvalues as its diagonal elements. Then, let $[\tilde{X}^i]_{\mathcal{V}} = \hat{\Lambda}_d^{-1/2} [\hat{\Phi}]_{\mathcal{G}}^\top (\mathbf{I}_p \otimes \mathbf{G}_{\mathcal{G}_0})[X^i]_{\mathcal{G}} \in \mathbb{R}^d, i \in \{1, \ldots, n\}$, contain the coordinates of the standardized observations in the eigenbasis. Finally, let*

(i) the columns of $[\hat{\Psi}^F]_{\mathcal{V}} \in \mathbb{R}^{d \times d}$ be the eigenvectors of the matrix

$$\frac{1}{n} \sum_{i=1}^n [\tilde{X}^i]_{\mathcal{V}}^\top [\tilde{X}^i]_{\mathcal{V}} \cdot [\tilde{X}^i]_{\mathcal{V}} [\tilde{X}^i]_{\mathcal{V}}^\top - (d+2)\mathbf{I}_d,$$

(ii) the columns of $[\hat{\Psi}^J]_{\mathcal{V}} = ([\hat{\psi}_1^J]_{\mathcal{V}}, \ldots, [\hat{\psi}_d^J]_{\mathcal{V}}) \in \mathbb{R}^{d \times d}$ be the orthonormal set of vectors satisfying

$$[\hat{\Psi}^J]_{\mathcal{V}} = \operatorname*{argmax}_{[\hat{\psi}^J]_{\mathcal{V}}^\top [\hat{\psi}^J]_{\mathcal{V}} = \mathbf{I}} \sum_{k=1}^d \sum_{\ell=1}^d \sum_{m=1}^d \left\{ [\hat{\psi}_m^J]_{\mathcal{V}}^\top (_{\mathcal{V}}[\hat{C}^{k\ell}(\tilde{X})]_{\mathcal{V}})[\hat{\psi}_m^J]_{\mathcal{V}} \right\}^2,$$

where $_{\mathcal{V}}[\hat{C}^{k\ell}(\tilde{X})]_{\mathcal{V}} = (1/n) \sum_{i=1}^n ([\tilde{X}^i]_{\mathcal{V}}^\top \mathbf{e}_k)([\tilde{X}^i]_{\mathcal{V}}^\top \mathbf{e}_\ell) \cdot [\tilde{X}^i]_{\mathcal{V}} [\tilde{X}^i]_{\mathcal{V}}^\top - \delta_{k\ell}\mathbf{I}_d - \mathbf{e}_k \mathbf{e}_\ell^\top - \mathbf{e}_\ell \mathbf{e}_k^\top.$

Then, choosing either the FOBI-solution $[\hat{\Psi}^F]_{\mathcal{V}}$ or the JADE-solution $[\hat{\Psi}^J]_{\mathcal{V}}$ the independent component scores are given by

$$\hat{\mathbf{Z}}^i = [\hat{\Psi}^\bullet]_{\mathcal{V}}^\top \hat{\Lambda}_d^{-1/2} [\hat{\Phi}]_{\mathcal{G}}^\top (\mathbf{I}_p \otimes \mathbf{G}_{\mathcal{G}_0})[X^i]_{\mathcal{G}}.$$

We note that the procedure in Theorem 4 is more general than the sequential approach of applying ordinary ICA to the basis function expansion of the data. To conduct the latter, one would need to substitute in Theorem 4 the ordinary covariance matrix $(1/n)[X]_{\mathcal{G}}^\top [X]_{\mathcal{G}}$ for the matrix $(1/n)[X]_{\mathcal{G}}^\top [X]_{\mathcal{G}} (\mathbf{I}_p \otimes \mathbf{G}_{\mathcal{G}_0})$ and replace the standardized observations $\hat{\Lambda}_d^{-1/2}[\hat{\Phi}]_{\mathcal{G}}^\top (\mathbf{I}_p \otimes \mathbf{G}_{\mathcal{G}_0})[X^i]_{\mathcal{G}}$ with $\hat{\Lambda}_d^{-1/2}[\hat{\Phi}]_{\mathcal{G}}^\top [X^i]_{\mathcal{G}}$. That is, the difference between the two approaches lies in the use of the matrices $(\mathbf{I}_p \otimes \mathbf{G}_{\mathcal{G}_0})$ which allow the proposed method to take into account both the functional nature (through the Gram matrix $\mathbf{G}_{\mathcal{G}_0}$ of the used basis) and the multivariate nature (through the Kronecker product) of the data. However, in the special case of an orthonormal basis we have $\mathbf{G}_{\mathcal{G}_0} = \mathbf{I}_K$, making the two approaches then equivalent.

The optimization problem required by the JADE-solution is easily solved with standard joint diagonalization techniques, e.g., the Jacobi angle algorithm, see [10]. Theorem 4 shows that the resulting vector of independent component scores is a linear transformation of the original vector of coordinates, $\hat{\mathbf{Z}}^i = \mathbf{A}[X^i]_{\mathcal{G}}$ for some $d \times pK$ matrix $\mathbf{A}$. Consequently, we can get interpretations for the independent component scores by considering the elements of $\mathbf{A}$ and observing which of the original coordinates most influence each of the obtained scores. The same procedure is used in the standard principal component analysis where the elements of the matrix $\mathbf{A}$ are called loadings. An example of such an interpretation will be given in the real data example in Section 6.

5.2. Choosing the value of d

We next give some rough guidelines on choosing an appropriate reduced dimension d. Naturally, we can estimate independent component scores corresponding to each latent function Z_j only if $d \geq p$. Moreover, even if we put $d = p$ it could still happen that some of the component functions have too low variation and cannot fit amongst the d eigenvectors of Σ_{XX} with the highest eigenvalues. From this point of view it would thus make sense to increase d further to make sure we capture all the latent functions. However, doing this also increases the odds of introducing more and more of the non-dependent part of the model (noise) to the estimation.

Further complication is brought in by Assumption 3 which in the sample version requires that all the diagonal matrices in the JADE-decomposition share a single eigenbasis. It can be shown that a sufficient condition for this is that each of the subvectors $\mathbf{Z}_{(\ell)}$, $\ell \in \{1, \ldots, m\}$ has either length one or an elliptical distribution. This condition is more likely to be fulfilled for small values of d and since $d = p$ is a natural meeting point for all these rules, allowing us to estimate all p latent functions in the best case, we advocate the use of the value $d = p$ in practice. This rule of thumb will be used in the examples of the next section.

6. Examples

6.1. Simulation study

In this simulation study we compare the two proposed methods to the alternative of applying only the principal component analysis part of the algorithm, that is, only projecting the data onto the space spanned by the first d eigenfunctions of Σ_{XX}.

For our setting we used $p = 4$ and considered for all four component functions the same 11-element Fourier basis $\mathcal{G}_0$. The first coefficients in the 11-dimensional coordinate vectors of the component functions were considered as "signal" and generated either as u_1, g_1, χ_1, e_1 (Setting 1) or as u_1, u_2, u_3, u_4 (Setting 2) where $u_1, u_2, u_3, u_4 \sim \mathcal{U}(0, 1)$, $g_1 \sim \Gamma(3, \sqrt{3})$, $\chi_1 \sim \chi_3^2$, $e_1 \sim \mathcal{E}(1)$ and all the previous random variables were independent and standardized to have zero means and unit variances. The remaining 40 coordinates were thought of as "noise" and sampled as independent standard normal random variables. The kurtosis of the signal components were thus distinct in the first setting and identical in the second. We generated samples of sizes $n = 1000, 2000, 4000, 8000, 16000, 32000, 64000$ and mixed the individual generated functions, $Z_i = (Z_{i1}, Z_{i2}, Z_{i3}, Z_{i4})^\top$, as $[Z^i]_\mathcal{G} \mapsto [X^i]_\mathcal{G} = \Omega[Z^i]_\mathcal{G}$ with a mixing matrix $\Omega \in \mathbb{R}^{44 \times 44}$. For simplicity, we considered estimation only in the true case $d = 4$.

To obtain Ω we first generated the matrix $\Omega_0 = \text{diag}\left\{(\mathbf{B}_4)^{1/2}, \mathbf{I}_{40}\right\}$, where $\mathbf{B}_4 = \mathbf{A}\mathbf{A}^\top + \lambda\mathbf{I}_4$, the matrix $\mathbf{A} \in \mathbb{R}^{4 \times 4}$ has independent standard normal elements and $\lambda = 0.5, 1.0, 1.5, 2.0, 2.5$ is a tuning parameter that controls how separated the spectra of the mixed and unmixed parts are. The mixing matrix Ω is now obtained by permuting the rows and columns of Ω_0 so that only the leading coefficients of the component functions are going to be mixed in the transformations $[Z^i]_\mathcal{G} \mapsto \Omega[Z^i]_\mathcal{G}$. This unorthodox procedure goes to ensure that the dependency between the four functions exists only in the directions given by the eigenvectors of Σ_{XX} with the eigenvalues $a_j^2 + \lambda$, $j \in \{1, \ldots, 4\}$, where a_j are the singular values of $\mathbf{A}$. Thus if $\lambda > 1$, the four largest eigenvalues always (on the population level) correspond to the directions of interest, meaning that the assumptions of our model are fulfilled and we always pick the correct four eigenvectors. A similar mixing scheme was used also in [29]. A total of 1000 replications of each combination of λ and sample size were generated for the study.

Subjecting the data to our proposed independent component methods, both of them estimate a matrix

$$\mathbf{W} = [\hat{\Psi}^\bullet]_\mathcal{V}^\top \hat{\Lambda}_d^{-1/2} [\hat{\Phi}]_\mathcal{G}^\top (\mathbf{I}_p \otimes \mathbf{G}_{\mathcal{G}_0}) \in \mathbb{R}^{d \times pK} = \mathbb{R}^{4 \times 44},$$

see Section 5, while the principal component analysis uses only the matrix $\mathbf{W} = [\hat{\Phi}]_\mathcal{G}^\top (\mathbf{I}_p \otimes \mathbf{G}_{\mathcal{G}_0}) \in \mathbb{R}^{4 \times 44}$. The independent/principal component scores are then $\mathbf{W}[X^i]_\mathcal{G} = \mathbf{W}\Omega[Z^i]_\mathcal{G}$ and for the methods to successfully separate the independent component functions each row of the gain matrix $\mathbf{W}\Omega$ should pick from $[Z^i]_\mathcal{G}$ coefficients relating only to a single component function. For assessing the performance of a single replication we first squared the elements of the estimated gain matrix and then summed row-wise over each block of size 4×11, resulting into a 4×4 matrix $\mathbf{R}$. The closer the matrix $\mathbf{R}$ is to the set $\mathcal{P}$ of matrices with a single non-zero element in each row and column, the better the result of estimation. To quantify this we use the minimum distance index [19], $D(\mathbf{R}) \in [0, 1]$, which has the value zero if and only if the separation is perfect, $\mathbf{R} \in \mathcal{P}$.

From the results we expect that the principal component analysis fails to estimate the sources under all settings, as the orthogonal transformation found by it is not enough to undo our mixing by the general matrix $\mathbf{B}_4^{1/2}$. The theory behind standard FOBI, on which our coordinate representation was seen to be based, says that FOBI cannot estimate components with matching kurtosis values [8] as is the case with the identical uniform distributions in our Setting 2. On the other hand, both FOBI and JADE should be able to find the solution in Setting 1 with differing, non-zero kurtosis values, the latter most likely outmatching the former. The resulting mean minimum distance indices across 1000 replications for different settings and parameter values are shown in Fig. 1 and distinctly verify our preconceptions. As discussed earlier, the separation fails on average if $\lambda \leq 1$ and we further see that the success of the separation is not particularly dependent on the value of λ, as long as we have $\lambda > 1$.

6.2. Real data example

We consider the *uWave* gesture data set available from http://www.ruf.rice.edu/~mobile/project_uwave.html [30]. At each day of the study the eight participants did ten repetitions of each of the eight gesture patterns in the Nokia
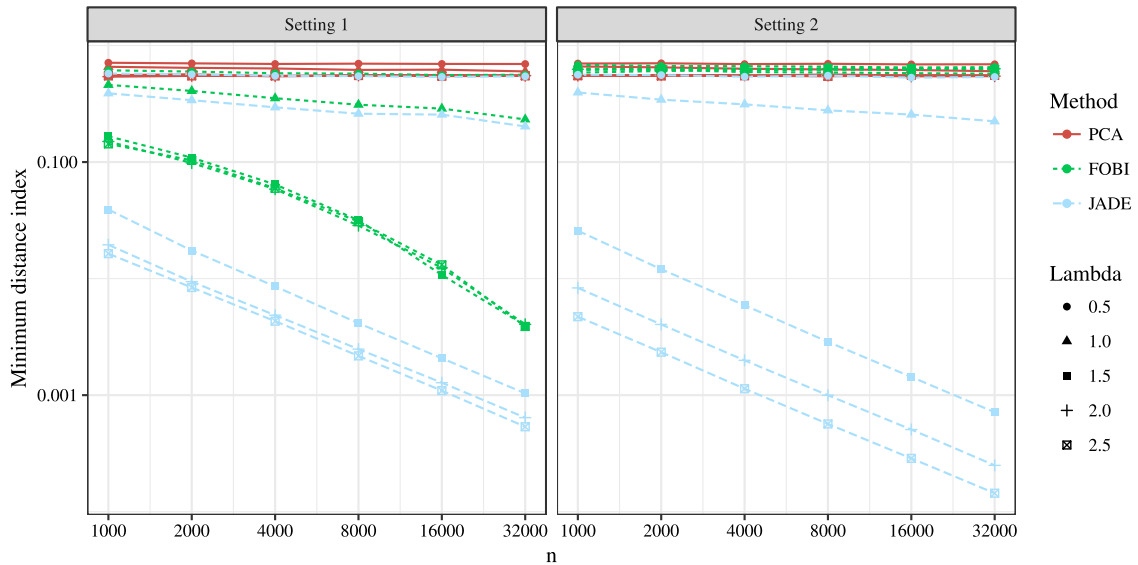
Fig. 1. Results of the simulation comparison between the proposed methods (FOBI, JADE) and applying only the principal component analysis part of the algorithm (PCA). Samples of 4-variate random functions were simulated for different sample sizes n and separation parameters λ. The lines show the average minimum distance indices of the different combinations of methods and settings over 1000 replications. The minimum distance index measures the accuracy of the separation, with lower values indicating better separation. The scales of both axes are logarithmic.

gesture vocabulary [23] using a Wii® remote measuring the 3D-acceleration of the gesture. Each participant had a total of seven study days making the total number of observed samples 4480. Of these we discarded two samples which had a measurement only for a single time point. Of the observed 3-variate curves (x, y and z-acceleration) we further took the subset corresponding to the three visually most similar gestures, a square, a clockwise circle and a counterclockwise circle, making our data a sample of multivariate functional data with $n = 1679$ and $p = 3$. A standard Fourier basis of 11 functions was fitted to all observations of each component function.

In pre-processing data, latent groups are most easily visually recognized from bivariate scatter plots and our objective is thus to extract from the data a pair of components that best reveal the latent group memberships. To evaluate the methods' capabilities for this we used the following scheme. For each of the 1000 replications we randomly partitioned the data into a training set of 400 observations and a test set of 1279 observations. Next, for each value of $d \in \{2, \ldots, 10\}$, the training set was subjected to either principal component analysis (conducted as in the previous example), FOBI or JADE. As low kurtosis is often an indicator of a multimodal distribution, for the independent component analysis methods we chose from the resulting independent component scores the two having the lowest fourth moments and for principal component analysis we considered two rules, taking the two scores with highest variances or taking the two scores with lowest fourth moments. Each chosen pair of scores was then used in quadratic discriminant analysis to create a classification rule and, finally, the proportion of correct classifications in the test set was computed for each rule.

The results are shown in Fig. 2 where the y-axis was cut from 0.7 downwards to allow more accurate representation of the interesting part of the plot. The curve for principal component analysis using kurtosis as a criterion continued descending until hitting the y-value of around 0.5 at $d = 6$. The main points of interest include the following. All methods perform equally well when $d = 2$ as then the chosen two components necessarily span the same space. Principal component analysis using variance as the criterion always chooses by definition the two first principal components regardless of the value of d, yielding a constant curve, and principal component analysis using kurtosis as the criterion clearly cannot find the relevant information at all. For $d = 3, 4$, FOBI and JADE are superior to principal component analysis in extracting the two components containing the classification information. Thus our heuristic suggestion of setting $d = p$ proved to be useful in this context.

Examples of the scatter plots of the pairs of components extracted from the training data by the three methods for $d = 3$ (with the same strategy for selecting the two components out of the three as earlier) are given in Fig. 3 where the principal components have been scaled to better show the details. The figure shows that of the two components found by principal component analysis only the first one provides information on the separation of the group locations while for FOBI and JADE both components carry information. Moreover, the overlapping of the groups follows a more irregular pattern for PCA than for the ICA-methods, making it more difficult for quadratic discriminant analysis to classify the observations based on principal components than based on independent components, as evidenced by Fig. 2.
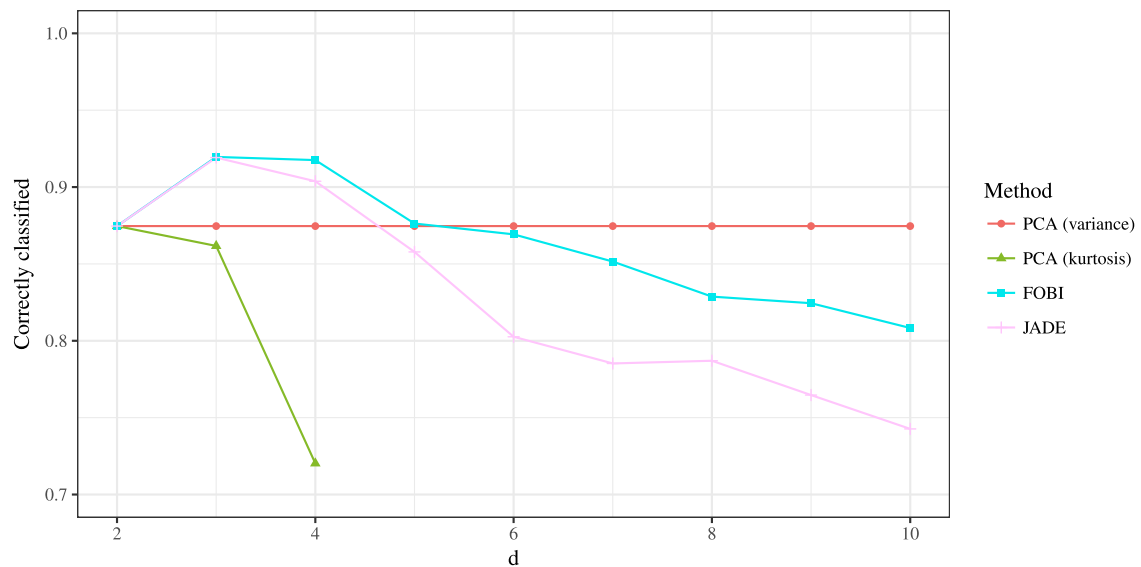
Fig. 2. Results of the comparison between the proposed methods (FOBI, JADE) and applying only the principal component analysis part of the algorithm (PCA) in the 3-variate *uWave* data set. For each number d of extracted components, the two with the lowest kurtosis, or, alternatively, highest variances in case of PCA, in the training set (400 observations) were used to create a classification rule. The plot shows the proportions of correctly classified cases in the test set (1279 observations) for the different combinations of methods and reduced dimension d. Each point in the plot represents an average over 1000 replications.
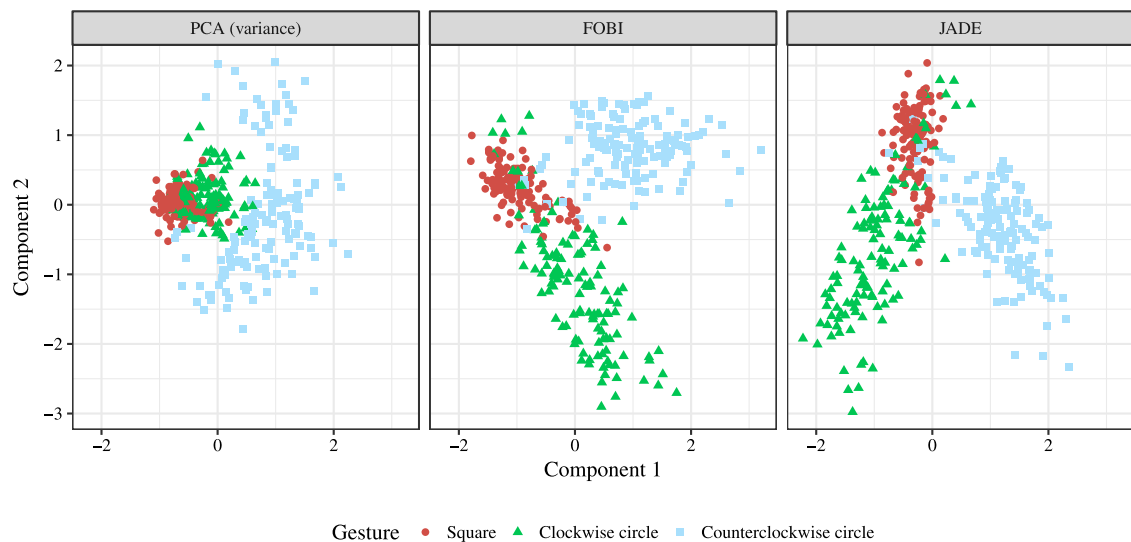

Fig. 3. Typical examples of pairs of classifying components found by the three methods, PCA (with variance as criterion), FOBI and JADE, from the *uWave* data set. The reduced dimension $d = 3$ was used in each case. The scatter plots show that all three methods succeed in capturing classifying information, with FOBI and JADE achieving a clearer group separation than PCA.

Interpretations for the FOBI independent component scores can now be obtained by examining the loading matrix reproduced in Table 1 where any loadings with absolute value greater than 0.6 have been shaded. For example, the final element of the second row tells the contribution of the 11th basis vector of the second observed function X_2 to the first estimated score $\hat{Z}_1$. We can now make two main observations. First, no separation information is carried by the basis elements of order six or higher. Since the higher index functions in Fourier bases control the finer, high-frequency properties of the resulting functions this reveals that most of the classification information is expectedly contained in the large-scale properties of the movements and accelerations. Secondly, the *y*-acceleration hardly contributes to any of the scores, showing that only the *x* and *z* directions are relevant in the classification. Also this makes sense, assuming that

Table 1
The loadings of the FOBI estimate for the *uWave* data set. For each of the $d = 3$ estimated independent components $\hat{Z}_j$, we obtain loadings corresponding to each basis vector of all three original functions X_k. The loadings with absolute value greater than 0.6 have been shaded. The locations of the shaded loadings indicate that the three independent components are comprised mostly of the first five basis elements of the first and third original functions (*x*- and *z*-acceleration).

$\hat{Z}_j$	X_k	1	2	3	4	5	6	7	8	9	10	11
1	1	−0.90	0.36	−1.21	−0.13	0.79	0.17	0.19	−0.01	−0.03	−0.06	0.04
1	2	−0.33	−0.07	−0.52	−0.32	0.59	−0.02	0.07	−0.00	−0.05	−0.08	0.01
1	3	−0.26	0.92	−0.40	−1.16	0.59	−0.32	0.32	0.00	−0.02	−0.05	−0.01
2	1	0.49	1.49	−0.08	−1.67	−0.10	−0.45	−0.07	−0.21	0.12	−0.11	0.10
2	2	−0.16	0.30	−0.04	−0.09	0.13	−0.16	−0.07	0.01	−0.04	−0.02	0.06
2	3	−0.20	−0.24	−0.86	0.83	0.86	0.27	0.52	0.05	−0.17	0.12	−0.11
3	1	0.16	0.39	0.31	−0.43	−0.26	−0.42	0.22	−0.00	−0.14	−0.04	−0.14
3	2	0.78	0.51	−0.10	−0.30	−0.35	−0.12	0.09	−0.03	0.10	0.02	−0.07
3	3	0.15	0.52	0.72	−0.52	−0.99	−0.49	−0.53	−0.08	0.24	−0.14	0.13

the gestures are drawn in the air roughly vertically, occupying mostly the *x-z* plane. Similar explanations could also be produced for the JADE and principal component analysis solutions (not shown here).

7. Discussion

We conclude the paper by discussing some directions for future research. First, while the provided rule of thumb of choosing $d = p$ proved useful in the examples, the logical next step is to provide a more analytical approach, e.g., in the form of sequential hypothesis testing.

Second, Theorem 3 shows how the independent component scores are obtained but tells nothing about the division of the scores into the independent subvectors. In our real data example this was not an issue as visual inspection already revealed us the scores of interest, but in the case of less visual data some kind of testing procedure is called for. A similar problem was encountered in [37] where an approach based on scatter matrices with the independence property was used to identify the independent subvectors, and a likewise procedure could possibly also be used here.

Third, asymptotic properties of the proposed estimator could likely be obtained as consistency results of the form, $\|\Gamma_k^\bullet(F_n) - \Gamma_k^\bullet(F)\|_{\mathrm{OP}} = \mathcal{O}_p(h_n)$, where $\Gamma_k^\bullet(F) = (\psi_k^\bullet \otimes \psi_k^\bullet)\Sigma(X^{(d)})^{-1/2}$ is one of the operators of Theorem 3, F_n is the empirical distribution function of a sample of size n from the distribution F, h_n is some sufficient rate of convergence and $k \in \{1, \ldots, d\}$. This asymptotic regime could further be combined with the assumption that the number of interesting principal components $d = d_n$ depends on the sample size, as mentioned in Section 4, to allow for a more flexible reduction into the finite-dimensional subspace.

Fourth, in Section 5 it was shown that the extensions of both FOBI and JADE to multivariate functional data can be applied in practice by projecting the observed functions into the space spanned by the first d eigenvectors of the covariance matrix operator and then subjecting the obtained standardized principal component coefficients to regular FOBI or JADE. This naturally begs for the question whether also some other standard multivariate methods can be meaningfully extended to multivariate functional data simply by applying them to the principal component coefficients. Some preliminary testing shows that this is certainly the case for FastICA, a projection pursuit-based family of independent component methods [18].

CRediT authorship contribution statement

Joni Virta: Methodology, Software, Writing — original draft. **Bing Li:** Conceptualization, Methodology, Writing — review & editing, Supervision. **Klaus Nordhausen:** Writing — review & editing, Supervision. **Hannu Oja:** Writing — review & editing, Supervision.

Acknowledgments

Appendix. Proofs

Proof of Lemma 1. The self-adjointness of Σ_{XX} follows simply from the earlier discussion of the adjoints of the components $\Sigma_{X_iX_j}$. Furthermore, by expanding element-wise we have for any $f \in \mathcal{H}$:

$$\langle \Sigma_{XX} f, f \rangle_{\mathcal{H}} = \mathrm{E} \{ \langle (X \otimes X) f, f \rangle_{\mathcal{H}} \} = \mathrm{E} \left(\langle X, f \rangle_{\mathcal{H}}^2 \right) \geq 0,$$

showing that Σ_{XX} is non-negative.

Let $\{e_k\}_{k=1}^{\infty}$ be an orthonormal basis of $\mathcal{H}$. Using the same reasoning as above, the trace of the self-adjoint, non-negative operator Σ_{XX} is then

$$\mathrm{trace}(\Sigma_{XX}) = \sum_{k=1}^{\infty} \langle \Sigma_{XX} e_k, e_k \rangle_{\mathcal{H}} = \mathrm{E} \left(\sum_{k=1}^{\infty} \langle X, e_k \rangle_{\mathcal{H}}^2 \right) = \mathrm{E} \|X\|_{\mathcal{H}}^2,$$

where the last equality uses Parseval's identity. Now, by our assumptions $\mathrm{E}\|X\|_{\mathcal{H}}^2$ is finite, making Σ_{XX} a trace-class operator.

To show that the affine equivariance holds, let $A \in \mathcal{L}(\mathcal{H})$ and write

$$\Sigma(AX) = \mathrm{E}(AX \otimes AX) = \mathrm{E}\left\{ A(X \otimes X) A^* \right\}.$$

Thus $\Sigma(AX)$ is the unique operator C satisfying $\langle f, Cg \rangle_{\mathcal{H}} = \mathrm{E}\{\langle f, A(X \otimes X) A^* g \rangle_{\mathcal{H}}\}$, for all $f, g \in \mathcal{H}$. Using again the definition of the expected value of a random operator, the right-hand side equals $\langle f, A\Sigma(X)A^* g \rangle_{\mathcal{H}}$, showing that $\Sigma(AX) = A\Sigma(X)A^*$.

Finally, the full independence property follows by assuming that X_i and X_j are independent and checking that

$$\mathrm{E}\left\{ \langle f_i, (X_i \otimes X_j) g_j \rangle_i \right\} = \mathrm{E}\left(\langle X_i, f_i \rangle_i \right) \mathrm{E}\left(\langle X_j, g_j \rangle_j \right) = 0,$$

for all $f_i \in \mathcal{H}_i$ and $g_j \in \mathcal{H}_j$, and thus by definition $\mathrm{E}(X_i \otimes X_j) = 0$. $\square$

Proof of Lemma 2. Since Σ is affine equivariant, and since $\{\phi_k\}_{k=1}^{\infty}$ are eigenvectors of $\Sigma(X)$, we have

$$\Sigma(X^{(d)}) = P_{\mathcal{M}_d} \Sigma(X) P_{\mathcal{M}_d} = \sum_{k=1}^{d} \lambda_k (\phi_k \otimes \phi_k),$$

which further implies that $\Sigma(X^{(d)})^{-1/2} = \sum_{k=1}^{d} \lambda_k^{-1/2} (\phi_k \otimes \phi_k)$. Next, for $Z^{(d)} = \Gamma_0 X^{(d)}$ we have

$$\Sigma(Z^{(d)}) = \sum_{k=1}^{d} \lambda_k (\Gamma_0 \phi_k \otimes \Gamma_0 \phi_k) = \Gamma_0 \Sigma(X^{(d)}) \Gamma_0^*.$$

As Γ_0 is boundedly invertible, the inverse square root of $\Sigma(Z^{(d)})$ exists as a bounded operator, and we can write

$$\Sigma(Z^{(d)})^{-1/2} Z^{(d)} = \left\{ \Sigma(Z^{(d)})^{-1/2} \Gamma_0 \Sigma(X^{(d)})^{1/2} \right\} \Sigma(X^{(d)})^{-1/2} X^{(d)}. \tag{10}$$

What remains is to prove that $A_0 = \Sigma(Z^{(d)})^{-1/2} \Gamma_0 \Sigma(X^{(d)})^{1/2}$ is unitary which follows by directly verifying,

$$A_0 A_0^* = \Sigma(Z^{(d)})^{-1/2} \Gamma_0 \Sigma(X^{(d)}) \Gamma_0^* \Sigma(Z^{(d)})^{-1/2},$$

where $\Gamma_0 \Sigma(X^{(d)}) \Gamma_0^*$ is equal to $\Sigma(Z^{(d)})$, showing that $A_0 A_0^* = P_{\mathcal{M}_d}$. The operator A_0 is thus unitary and consequently also $A_0^* A_0 = P_{\mathcal{M}_d}$. Applying now A_0^* from left to both sides of (10) shows that $U_0 = A_0^*$, concluding the proof. $\square$

Proof of Lemma 3. We provide the proof for the second term in (6), the proofs for the third and fourth terms following similarly. Using the definition of the expected value of a random operator, the second term is the unique operator $A \in \mathcal{L}(\mathcal{M}_d)$ with

$$\langle f, Ag \rangle_{\mathcal{H}} = \mathrm{E}\left(\langle X', \phi_i \rangle_{\mathcal{H}} \langle X', \phi_j \rangle_{\mathcal{H}} \langle X, f \rangle_{\mathcal{H}} \langle X, g \rangle_{\mathcal{H}} \right),$$

for all $f, g \in \mathcal{M}_d$. The independence of X and X' further implies that the right-hand side can be written in the form

$$\mathrm{E}\left\{ \langle (X' \otimes X') \phi_i, \phi_j \rangle_{\mathcal{H}} \right\} \mathrm{E}\left\{ \langle (X \otimes X) f, g \rangle_{\mathcal{H}} \right\} = \langle \Sigma(X') \phi_i, \phi_j \rangle_{\mathcal{H}} \langle \Sigma(X) f, g \rangle_{\mathcal{H}},$$

which equals $\langle f, \delta_{ij} P_{\mathcal{M}_d} g \rangle$ under our assumptions, concluding the proof. $\square$

Proof of Theorem 1. Consider only the first term in the expansion of C^{ij} in Lemma 3. Plugging in UZ we get $U\mathrm{E}\{\langle UZ, \phi_i \rangle_{\mathcal{H}} \langle UZ, \phi_j \rangle_{\mathcal{H}} (Z \otimes Z)\} U^* = UMU^*$. The (k, ℓ) component operator of the expected value M is defined as the operator $A_{k\ell} \in \mathcal{L}(\mathcal{H}_\ell, \mathcal{H}_k)$ satisfying

$$\langle f_k, A_{k\ell} g_\ell \rangle_k = \sum_{s,t=1}^{p} \mathrm{E}\left(\langle Z_s, \xi_{is} \rangle_s \langle Z_t, \xi_{jt} \rangle_t \langle Z_k, f_k \rangle_k \langle Z_\ell, g_\ell \rangle_\ell \right), \tag{11}$$

for all $f_k \in \text{span}(\{\phi_{mk}\}_{m=1}^d)$ and $g_\ell \in \text{span}(\{\phi_{m\ell}\}_{m=1}^d)$ where $\xi_i = (\xi_{i1}, \ldots, \xi_{ip}) = U^*\phi_i$. Concentrate first on the off-diagonal case $k \neq \ell$. Then either $s = k, t = \ell$ or $s = \ell, t = k$ as otherwise the independence and zero means of the component functions reduce the sum to zero. Consider the first of these cases:

$$\text{E}\{\langle f_k, (Z_k \otimes Z_k)\xi_{ik}\rangle_k\} \text{E}\{\langle g_\ell, (Z_\ell \otimes Z_\ell)\xi_{j\ell}\rangle_\ell\} = \langle f_k, (\xi_{ik} \otimes \xi_{j\ell})g_\ell\rangle_k.$$

The expected value of an arbitrary (k, ℓ)th off-diagonal component operator of M is thus $(\xi_{ik} \otimes \xi_{j\ell}) + (\xi_{jk} \otimes \xi_{i\ell})$, which can be recognized to be also the (k, ℓ)th component operator of $U^*\{(\phi_i \otimes \phi_j) + (\phi_j \otimes \phi_i)\}U$.

The general form for an arbitrary (k, k)th diagonal component operator of M can be found in a similar manner. Notice first that if $k = \ell$ in (11) then it must be that $s = t$ or otherwise the sum is again zero by independence and zero means. The summation over s can then be divided into two cases, $s = k$ and $s \neq k$. Similar manipulation as done above yields then the expected value $\text{E}\{(Z_k \otimes Z_k)(\xi_{ik} \otimes \xi_{jk})(Z_k \otimes Z_k)\} + (\delta_{ij} - \langle \xi_{ik}, \xi_{jk}\rangle_k)P_k$ for the (k, k)th diagonal operator where the first summand comes from the former case and the second from the latter.

Putting now everything together into a matrix of operators shows that the three last terms in the alternative form for C^{ij} in Lemma 3 cancel out, leaving us with the claimed result. □

Proof of Theorem 2. For an arbitrary $X \in \mathcal{X}^4(\mathcal{M}_d)$ with $\Sigma(X) = P_{\mathcal{M}_d}$ we have by Lemma 3

$$C(X) = \sum_{i=1}^d C^{ii}(X) = \text{E}\left\{\sum_{i=1}^d \langle X, \phi_i\rangle_{\mathcal{H}}^2 (X \otimes X)\right\} - (d+2)P_{\mathcal{M}_d}, \tag{12}$$

the argument of the expectation being further simplified by Parseval's identity to $\sum_{i=1}^d \langle X, \phi_i\rangle_{\mathcal{H}}^2 (X \otimes X) = \|X\|_{\mathcal{H}}^2 (X \otimes X) = (X \otimes X)^2$. The first claimed equality now follows from the form $C(X) = \text{E}\{(X \otimes X)^2\} - (d+2)P_{\mathcal{M}_d}$.

By Theorem 1, an arbitrary off-diagonal element of the operator $C(Z) = \sum_{i=1}^d D^{ii}$ is zero. The exact form for its diagonal elements D_{kk} could also be derived from (7) but the seeming dependency of $D_{kk} = \sum_{i=1}^d D_{kk}^{ii}$ on the operator U needlessly complicates things and it is simpler to proceed straight from the form $C(Z) = \text{E}\{(Z \otimes Z)^2\} - (d+2)P_{\mathcal{M}_d}$. The (k, k)th diagonal operator of the first term is then defined as the unique operator $A_{kk} \in \mathcal{L}(\mathcal{H}_k, \mathcal{H}_k)$ satisfying

$$\langle f_k, A_{kk}g_k\rangle_k = \text{E}\left(\sum_{j=1}^d \langle Z_j, Z_j\rangle_j \langle Z_k, f_k\rangle_k \langle Z_k, g_k\rangle_k\right),$$

for all $f_k \in \text{span}(\{\phi_{mk}\}_{m=1}^d)$ and $g_k \in \text{span}(\{\phi_{mk}\}_{m=1}^d)$. Divide the summation over j into two cases, $j = k$ and $j \neq k$. The former yields the term $\text{E}\{\langle f_k, \|Z_k\|_k^2 (Z_k \otimes Z_k), g_k\rangle_k\}$ contributing $\text{E}\{\|Z_k\|_k^2 (Z_k \otimes Z_k)\} = \text{E}\{(Z_k \otimes Z_k)^2\}$ to the final expected value. The latter yields the term

$$\left\{\sum_{j\neq k} \text{E}\left(\|Z_j\|_j^2\right)\right\} \text{E}\{\langle f_k, (Z_k \otimes Z_k), g_k\rangle_k\} = \left\{\sum_{j\neq k} \text{E}\left(\|Z_j\|_j^2\right)\right\} \langle f_k, g_k\rangle_k,$$

where the first multiplicand can be written as $\text{E}\left(\|Z\|_{\mathcal{H}}^2\right) - \text{E}\left(\|Z_k\|_k^2\right)$, whose first term equals by Parseval's identity

$$\text{E}\left(\|Z\|_{\mathcal{H}}^2\right) = \sum_{i=1}^d \text{E}\left(\langle Z, \phi_i\rangle_{\mathcal{H}}^2\right) = \sum_{i=1}^d \langle \phi_i, \text{E}\left(Z \otimes Z\right) \phi_i\rangle = d.$$

Similarly, by choosing an orthonormal basis for the kth component space one can show that $\text{E}\left(\|Z_k\|_k^2\right) = d_k := \dim\{\text{span}(\{\phi_{mk}\}_{m=1}^d)\}$. The total contribution of the case $j \neq k$ to the expected value of the (k, k)th diagonal operator is thus $(d - d_k)P_k$. Finally, putting everything together with (12) yields the desired result. □

Proof of Lemma 4. Inspect without loss of generality the first eigenvector ψ_1 and assume that it is not canonical, $\psi_1 = (\psi_{11}, \ldots, \psi_{1p})$, where, without loss of generality, ψ_{11} and ψ_{12} are assumed to be non-zero. Then the linearly independent vectors $(\psi_{11}, 0, 0, \ldots, 0)$ and $(0, \psi_{12}, 0, \ldots, 0)$ are both eigenvectors of D associated with the same eigenvalue τ_1, making the eigenspace associated with the eigenvalue τ_1 have dimension of at least 2, a contradiction as the assumption on the distinctness of eigenvalues implies unit rank. Thus only one of $\psi_{11}, \ldots, \psi_{1p}$ can be non-zero. □

Proof of Lemma 5. By the Cauchy–Schwarz inequality and the unit length of ψ_k we have

$$w(\psi_1, \ldots, \psi_d) \leq \sum_{i=1}^l \sum_{k=1}^d \langle \psi_k, \psi_k\rangle_{\mathcal{H}} \langle S_i\psi_k, S_i\psi_k\rangle_{\mathcal{H}} = \sum_{i=1}^l \sum_{k=1}^d \langle S_i\psi_k, S_i\psi_k\rangle_{\mathcal{H}}.$$

Now, $\sum_{k=1}^d \langle S_i\psi_k, S_i\psi_k\rangle_{\mathcal{H}} = \|S_i\|_{HS}^2$ for any orthonormal basis $\{\psi_k\}_{k=1}^d$ and we have shown the first part of the claim. To see when the equality holds recall that the Cauchy–Schwarz inequality preserves equality if and only if the two vectors

in question are proportional. We must thus have $\psi_k = a_{ik} S_i \psi_k$ for some $a_{ik} \in \mathbb{R}$ for all $i \in \{1, \ldots, I\}$, $k \in \{1, \ldots, d\}$, which is equivalent to saying that each ψ_k is an eigenvector of each S_i. $\square$

Proof of Theorem 3. Recall first that by Lemma 2 we have $\tilde{X} = U_0 \tilde{Z}$ where $\tilde{Z} = \Sigma(Z^{(d)})^{-1/2} Z^{(d)}$. By Lemma 1 the operator $\Sigma(Z^{(d)})$ is diagonal and thus one possible choice for the inverse square root of the operator $\Sigma(Z^{(d)})$ is also a diagonal operator, namely the diagonal operator G with some inverse square roots of the diagonal elements of $\Sigma(Z^{(d)})$ as its diagonal elements. With this choice, $\Sigma(Z^{(d)})^{-1/2} = G$, also $\tilde{Z}$ has then independent component functions. A reasoning similar to the one used in Remark 2.1 in [20] shows that all inverse square roots of $\Sigma(Z^{(d)})$ are of the form VG where V is unitary and can by the unitary equivariance be taken out of $C(\tilde{Z})$, "merging" it with U_0. We may thus assume that $\Sigma(Z^{(d)})^{-1/2}$ is a diagonal operator. Invoking then Theorem 2 shows that $C(\tilde{Z})$ is also a diagonal operator.

Let $\{h_k^F\}_{k=1}^d$ be the eigenvectors of $C(\tilde{Z})$. Then by Theorem 2 the FOBI-basis of X is given by $\{\psi_k^F\}_{k=1}^d = \{U_0 h_k^F\}_{k=1}^d$. Then, by the unitarity of U_0 we have

$$\hat{Z}_k^F = U_0(h_k^F \otimes h_k^F)\tilde{Z}.$$

As $C(\tilde{X})$ and $C(\tilde{Z})$ share the same eigenvalues all the assumptions of Lemma 4 are satisfied and only the $\ell(k)$th element of h_k^F is non-zero, $k \in \{1, \ldots, d\}$. Consequently

$$\hat{Z}_k = \langle h_{k\ell(k)}^F, \tilde{Z}_{\ell(k)} \rangle_{\ell(k)} U_0 h_k^F,$$

showing that $\hat{Z}_k$ depends only on the $\ell(k)$th component of Z.

The result for the JADE-basis follows similarly. We first notice that by Theorem 1 the operators C^{ij} are semi-unitary equivariant in the sense that we may again assume that $\Sigma(Z^{(d)})^{-1/2}$ is a diagonal operator and that the random function $\tilde{Z}$ has independent component functions. Let then $\{\psi_k^J\}_{k=1}^d$ be the joint diagonalizer of $\mathcal{C}$. Now, by Theorem 1 we have $C^{ij}(\tilde{X}) = \mathbf{U}_0 D^{ij} \mathbf{U}_0^*$ where $D^{ij} = D^{ij}(U_0, \tilde{Z})$ are diagonal operators, $i, j \in \{1, \ldots, d\}$. By Lemma 5 the joint diagonalizer of the set $\{D^{ij}\}_{i,j=1}^d$ is $\{h_k^J\}_{k=1}^d$ where each h_k^J is canonical. Consequently, the joint diagonalizer of $\mathcal{C}$ is $\{\psi_k^J\}_{k=1}^d = \{U_0 h_k^J\}_{k=1}^d$ and the desired result follows as above with FOBI. $\square$

Proof of Theorem 4. First, our space being finite-dimensional, for every fixed pair of bases $\mathcal{B}, \mathcal{G}$ every linear operator A in $\mathcal{M}$ has with it associated the unique matrix $_\mathcal{G}[A]_\mathcal{B} \in \mathbb{R}^{pK \times pK}$ that satisfies $[Af]_\mathcal{G} = (_\mathcal{G}[A]_\mathcal{B})[f]_\mathcal{B}$, for all $f \in \mathcal{M}$. Furthermore, a function f is an eigenfunction of the operator A associated with the eigenvalue λ if and only if $[f]_\mathcal{B}$ is an eigenvector of the matrix $_\mathcal{B}[A]_\mathcal{B}$ associated with the same eigenvalue λ.

The inner product of two elements $f_1, f_2 \in \mathcal{M}$ expressed in the same basis $\mathcal{G} = \{g_k\}_{k=1}^K$ is given simply by

$$\langle f_1, f_2 \rangle = \sum_{k=1}^K \sum_{k'=1}^K ([f_1]_\mathcal{G})_k ([f_2]_\mathcal{G})_{k'} \langle g_k, g_{k'} \rangle = [f_1]_\mathcal{G}^\top \mathbf{G}_\mathcal{G} [f_2]_\mathcal{G},$$

where $\mathbf{G}_\mathcal{G} = (\langle g_k, g_{k'} \rangle)_{k,k'=1}^K$ is the Gram matrix of the basis $\mathcal{G}$. The tensor product between two elements $f_1, f_2 \in \mathcal{M}$ has the following coordinate

$$_\mathcal{G}[f_1 \otimes f_2]_\mathcal{G} = [f_1]_\mathcal{G} [f_2]_\mathcal{G}^\top \mathbf{G}_\mathcal{G}. \tag{13}$$

These and more properties about the coordinate system were used and further developed in [26].

We begin with the coordinate representation of the standardization step. An estimate for $\Sigma_{X_r X_s}$ is

$$\hat{\Sigma}_{X_r X_s} = \frac{1}{n} \sum_{i=1}^n (X_{ir} \otimes X_{is}).$$

The coordinate of $\hat{\Sigma}_{XX}$ is the matrix $\{_{\mathcal{G}_0}[\hat{\Sigma}_{X_r X_s}]_{\mathcal{G}_0}\}_{r,s=1}^p$, which, by (13), has

$$_{\mathcal{G}_0}[\hat{\Sigma}_{X_r X_s}]_{\mathcal{G}_0} = \frac{1}{n} \sum_{i=1}^n [X_{ir}]_{\mathcal{G}_0} [X_{is}]_{\mathcal{G}_0}^\top \mathbf{G}_{\mathcal{G}_0}.$$

Assemble these matrices together to obtain

$$_\mathcal{G}[\hat{\Sigma}_{XX}]_\mathcal{G} = \frac{1}{n} [X]_\mathcal{G}^\top [X]_\mathcal{G} (\mathbf{I}_p \otimes \mathbf{G}_{\mathcal{G}_0}),$$

where $\otimes$ is the Kronecker product between matrices.

We next fix the dimension $d \leq pK$ and estimate the coordinate $[\hat{\phi}_\ell]_\mathcal{G}$ of the first d eigenfunctions $\hat{\phi}_\ell$ of $\hat{\Sigma}_{XX}$. As shown in [26], $\hat{\phi}_\ell$ is the ℓth eigenfunction of the operator $\hat{\Sigma}_{XX}$ if and only if $(\mathbf{I}_p \otimes \mathbf{G}_{\mathcal{G}_0}^{1/2})[\hat{\phi}_\ell]_\mathcal{G}$ is the ℓth eigenvector of the matrix $(\mathbf{I}_p \otimes \mathbf{G}_{\mathcal{G}_0}^{1/2})(_\mathcal{G}[\hat{\Sigma}_{XX}]_\mathcal{G})(\mathbf{I}_p \otimes \mathbf{G}_{\mathcal{G}_0}^{-1/2})$. The orthogonal projection of $f \in \mathcal{M}$ onto $span(\mathcal{V})$, where $\mathcal{V} = \{\hat{\phi}_\ell\}_{\ell=1}^d$, is then

$$\sum_{\ell=1}^d (\hat{\phi}_\ell \otimes \hat{\phi}_\ell)f = \sum_{k=1}^d [\hat{\phi}_\ell]_\mathcal{G}^\top (\mathbf{I}_p \otimes \mathbf{G}_{\mathcal{G}_0})[f]_\mathcal{G} \hat{\phi}_\ell,$$

and the coordinates of the observations in the eigenbasis $\mathcal{V}$ are thus

$$[X^{i(d)}]_{\mathcal{V}} = [\hat{\Phi}]_{\mathcal{G}}^{\top}(\mathbf{I}_p \otimes \mathbf{G}_{\mathcal{G}_0})[X^i]_{\mathcal{G}} \in \mathbb{R}^d, \quad i \in \{1, \ldots, n\},$$

where $[\hat{\Phi}]_{\mathcal{G}} = ([\hat{\phi}_1]_{\mathcal{G}}, \ldots, [\hat{\phi}_d]_{\mathcal{G}})$. Let $[X^{(d)}]_{\mathcal{V}} = ([X^{1(d)}]_{\mathcal{V}}, \ldots, [X^{n(d)}]_{\mathcal{V}})^{\top}$. Then the above equations can be written as

$$[X^{(d)}]_{\mathcal{V}} = [X]_{\mathcal{G}}(\mathbf{I}_p \otimes \mathbf{G}_{\mathcal{G}_0})[\hat{\Phi}]_{\mathcal{G}}.$$

Since the principal component scores satisfy $_{\mathcal{V}}[\hat{\Sigma}(X^{(d)})]_{\mathcal{V}} = \hat{\mathbf{\Lambda}}_d$, where $\hat{\mathbf{\Lambda}}_d = \operatorname{diag}(\hat{\lambda}_1, \ldots, \hat{\lambda}_d)$ contains the eigenvalues of $\hat{\Sigma}_{XX}$, the coordinates of the standardized observations $\tilde{X}^i$ in the eigenbasis are

$$[\tilde{X}^i]_{\mathcal{V}} = (_{\mathcal{V}}[\hat{\Sigma}(X^{(d)})]_{\mathcal{V}})^{-1/2}[X^{i(d)}]_{\mathcal{V}} = \hat{\mathbf{\Lambda}}_d^{-1/2}[\hat{\Phi}]_{\mathcal{G}}^{\top}(\mathbf{I}_p \otimes \mathbf{G}_{\mathcal{G}_0})[X^i]_{\mathcal{G}}.$$

Turning our attention to the fourth cross-cumulant operators we have for fixed $k, \ell \in \{1, \ldots, d\}$ the estimate

$$\hat{C}^{k\ell}(\tilde{X}) = \frac{1}{n}\sum_{i=1}^{n}\langle\tilde{X}^i, \hat{\phi}_k\rangle_{\mathcal{V}}\langle\tilde{X}^i, \hat{\phi}_\ell\rangle_{\mathcal{V}}(\tilde{X}^i \otimes \tilde{X}^i) - \delta_{k\ell}P_{\mathcal{M}_d} - \hat{\phi}_k \otimes \hat{\phi}_\ell - \hat{\phi}_\ell \otimes \hat{\phi}_k,$$

where the inner product $\langle\tilde{X}^i, \hat{\phi}_k\rangle_{\mathcal{V}}$ just extracts the kth element of the coordinate vector $[\tilde{X}^i]_{\mathcal{V}}$. Reasoning then as above with the covariance matrix operator it is straightforward to obtain the following coordinate representation:

$$_{\mathcal{V}}[\hat{C}^{k\ell}(\tilde{X})]_{\mathcal{V}} = \frac{1}{n}\sum_{i=1}^{n}([\tilde{X}^i]_{\mathcal{V}}^{\top}\mathbf{e}_k)([\tilde{X}^i]_{\mathcal{V}}^{\top}\mathbf{e}_\ell) \cdot [\tilde{X}^i]_{\mathcal{V}}[\tilde{X}^i]_{\mathcal{V}}^{\top} - \delta_{k\ell}\mathbf{I}_d - \mathbf{e}_k\mathbf{e}_\ell^{\top} - \mathbf{e}_\ell\mathbf{e}_k^{\top},$$

where $\mathbf{e}_k$ is the kth canonical basis vector of $\mathbb{R}^d$ and $\mathbf{I}_d$ is the $d \times d$ identity matrix. The similarity of this form to (2) already suggests that the functional independent component analysis solutions are found by performing regular FOBI or JADE on the coordinates $[\tilde{X}^i]_{\mathcal{V}}$ of the standardized observations.

The coordinate representation of the estimate of the FOBI-operator (8) is now simply

$$_{\mathcal{V}}[\hat{C}(\tilde{X})]_{\mathcal{V}} = \frac{1}{n}\sum_{i=1}^{n}[\tilde{X}^i]_{\mathcal{V}}^{\top}[\tilde{X}^i]_{\mathcal{V}} \cdot [\tilde{X}^i]_{\mathcal{V}}[\tilde{X}^i]_{\mathcal{V}}^{\top} - (d+2)\mathbf{I}_d,$$

and an estimate $\mathcal{U}^F = \{\hat{\psi}_m^F\}_{m=1}^d$ for the FOBI-basis is found from its eigendecomposition. Letting next $[\hat{\Psi}^F]_{\mathcal{V}} = ([\hat{\psi}_1^F]_{\mathcal{V}}, \ldots, [\hat{\psi}_d^F]_{\mathcal{V}}) \in \mathbb{R}^{d\times d}$ be the coordinate representation of the eigenvectors of $\hat{C}(\tilde{X})$ in $\mathcal{V}$, the vector of the FOBI independent component scores, $\langle\hat{\psi}_m^F, \tilde{X}^i\rangle = [\hat{\psi}_m^F]_{\mathcal{V}}^{\top}[\tilde{X}^i]_{\mathcal{V}}$, $m \in \{1, \ldots, d\}$, is then finally obtained as

$$[\hat{\Psi}^F]_{\mathcal{V}}^{\top}[\tilde{X}^i]_{\mathcal{V}} = [\hat{\Psi}^F]_{\mathcal{V}}^{\top}\hat{\mathbf{\Lambda}}_d^{-1/2}[\hat{\Phi}]_{\mathcal{G}}^{\top}(\mathbf{I}_p \otimes \mathbf{G}_{\mathcal{G}_0})[X^i]_{\mathcal{G}}.$$

For the JADE-solution, an estimate $\mathcal{U}^J = \{\hat{\psi}_m^J\}_{m=1}^d$ for the JADE-basis, i.e., the joint diagonalizer of the set $\{\hat{C}^{k\ell}(\tilde{X})\}_{k,\ell=1}^d$, is found by maximizing the quantity (9), the maximization problem now having the form

$$[\hat{\Psi}^J]_{\mathcal{V}} = \operatorname*{argmax}_{[\hat{\Psi}^J]_{\mathcal{V}}^{\top}[\hat{\Psi}^J]_{\mathcal{V}}=\mathbf{I}_d} \sum_{k=1}^{d}\sum_{\ell=1}^{d}\sum_{m=1}^{d}\left\{[\hat{\psi}_m^J]_{\mathcal{V}}^{\top}(_{\mathcal{V}}[\hat{C}^{k\ell}(\tilde{X})]_{\mathcal{V}})[\hat{\psi}_m^J]_{\mathcal{V}}\right\}^2,$$

where $[\hat{\Psi}^J]_{\mathcal{V}} = ([\hat{\psi}_1^J]_{\mathcal{V}}, \ldots, [\hat{\psi}_d^J]_{\mathcal{V}}) \in \mathbb{R}^{d\times d}$. As with FOBI above, the vectors of the JADE independent component scores are then

$$[\hat{\Psi}^J]_{\mathcal{V}}^{\top}[\tilde{X}^i]_{\mathcal{V}} = [\hat{\Psi}^J]_{\mathcal{V}}^{\top}\hat{\mathbf{\Lambda}}_d^{-1/2}[\hat{\Phi}]_{\mathcal{G}}^{\top}(\mathbf{I}_p \otimes \mathbf{G}_{\mathcal{G}_0})[X^i]_{\mathcal{G}}. \quad \square$$

References

[1] F.R. Bach, M.I. Jordan, Kernel independent component analysis, J. Mach. Learn. Res. 3 (Jul) (2002) 1–48.
[2] C.R. Baker, Joint measures and cross-covariance operators, Trans. Amer. Math. Soc. 186 (1973) 273–289.
[3] J.R. Berrendero, A. Justel, M. Svarc, Principal components for multivariate functional data, Comput. Statist. Data Anal. 55 (9) (2011) 2619–2634.
[4] S. Bonhomme, J.-M. Robin, Consistent noisy independent component analysis, J. Econometrics 149 (1) (2009) 12–25.
[5] D. Bosq, Linear Processes in Function Spaces: Theory and Applications, Vol. 149, Springer Science & Business Media, 2012.
[6] V.D. Calhoun, T. Adali, L.K. Hansen, J. Larsen, J.J. Pekar, ICA of functional MRI data: an overview, in: International Workshop on Independent Component Analysis and Blind Signal Separation, 2003.
[7] J.-F. Cardoso, Multidimensional independent component analysis, in: Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, volume 4, 1998, pp. 1941–1944.
[8] J.-F. Cardoso, Source separation using higher order moments, in: International Conference on Acoustics, Speech, and Signal Processing, 1989, 1989, pp. 2109–2112.
[9] J.-F. Cardoso, A. Souloumiac, Blind beamforming for non-Gaussian signals, in: IEE Proceedings F-Radar and Signal Processing, volume 140, 1993, pp. 362–370.
[10] J.-F. Cardoso, A. Souloumiac, Jacobi angles for simultaneous diagonalization, SIAM J. Matrix Anal. Appl. 17 (1) (1996) 161–164.
[11] J.-M. Chiou, Y.-T. Chen, Y.-F. Yang, Multivariate functional principal component analysis: A normalization approach, Statist. Sinica 24 (2014) 1571–1596.

[12] P. Comon, C. Jutten, Handbook of Blind Source Separation: Independent Component Analysis and Applications, Academic Press, 2010.
[13] J.B. Conway, A Course in Functional Analysis, Vol. 96, Springer Science & Business Media, 2013.
[14] W. Dai, M.G. Genton, Multivariate functional data visualization and outlier detection, J. Comput. Graph. Statist. 27 (4) (2018) 923–934.
[15] H.W. Gutch, F.J. Theis, To infinity and beyond: On ICA over Hilbert spaces, in: LVA/ICA, Springer, 2012, pp. 180–187.
[16] C. Happ, S. Greven, Multivariate functional principal component analysis for data observed on different (dimensional) domains, J. Amer. Statist. Assoc. 113 (522) (2018) 649–659.
[17] M. Hubert, P.J. Rousseeuw, P. Segaert, Multivariate functional outlier detection, Stat. Methods Appl. 24 (2) (2015) 177–202.
[18] A. Hyvärinen, E. Oja, A fast fixed-point algorithm for independent component analysis, Neural Comput. 9 (7) (1997) 1483–1492.
[19] P. Ilmonen, K. Nordhausen, H. Oja, E. Ollila, A new performance index for ICA: properties, computation and asymptotic analysis, in: International Conference on Latent Variable Analysis and Signal Separation, Springer, 2010, pp. 229–236.
[20] P. Ilmonen, H. Oja, R. Serfling, On invariant coordinate system (ICS) functionals, Internat. Statist. Rev. 80 (1) (2012) 93–110.
[21] J. Jacques, C. Preda, Model-based clustering for multivariate functional data, Comput. Statist. Data Anal. 71 (2014) 92–106.
[22] M. Kayano, K. Dozono, S. Konishi, Functional cluster analysis via orthonormalized Gaussian basis expansions and its application, J. Classification 27 (2) (2010) 211–230.
[23] J. Kela, P. Korpipää, J. Mäntyjärvi, S. Kallio, G. Savino, L. Jozzo, S. Di Marca, Accelerometer-based gesture control for a design environment, Pers. Ubiquitous Comput. 10 (5) (2006) 285–299.
[24] Z. Koldovsky, P. Tichavsky, E. Oja, Efficient variant of algorithm FastICA for independent component analysis attaining the Cramer-Rao lower bound, IEEE Trans. Neural Netw. 17 (5) (2006) 1265–1277.
[25] B. Li, H. Chun, H. Zhao, On an additive semigraphoid model for statistical networks with application to pathway analysis, J. Amer. Statist. Assoc. 109 (2014) 1188–1204.
[26] B. Li, E. Solea, A nonparametric graphical model for functional data with application to brain networks based on fMRI, J. Amer. Statist. Assoc. 113 (2018) 1637–1655.
[27] B. Li, J. Song, Dimension reduction for functional data based on weak conditional moments, 2017. Submitted.
[28] B. Li, J. Song, Nonlinear sufficient dimension reduction for functional data, Ann. Statist. 45 (2017) 1059–1095.
[29] B. Li, G. Van Bever, H. Oja, R. Sabolová, F. Critchley, Functional independent component analysis: an extension of the fourth-order blind identification, 2019, Unpublished manuscript.
[30] J. Liu, L. Zhong, J. Wickramasuriya, V. Vasudevan, UWave: Accelerometer-based personalized gesture recognition and its applications, Pervasive Mob. Comput. 5 (6) (2009) 657–675.
[31] M. Matilainen, K. Nordhausen, H. Oja, New independent component analysis tools for time series, Statist. Probab. Lett. 105 (2015) 80–87.
[32] J. Miettinen, K. Nordhausen, H. Oja, S. Taskinen, Deflation-based FastICA with adaptive choices of nonlinearities, IEEE Trans. Signal Process. 62 (21) (2014) 5716–5724.
[33] J. Miettinen, K. Nordhausen, H. Oja, S. Taskinen, J. Virta, The squared symmetric FastICA estimator, Signal Process. 131 (2017) 402–411.
[34] J. Miettinen, K. Nordhausen, S. Taskinen, Blind source separation based on joint diagonalization in R: The packages JADE and BSSasymp, J. Stat. Softw. 76 (2) (2017) 1–31.
[35] J. Miettinen, S. Taskinen, K. Nordhausen, H. Oja, Fourth moments and independent component analysis, Statist. Sci. 30 (3) (2015) 372–390.
[36] E. Moreau, A generalization of joint-diagonalization criteria for source separation, IEEE Trans. Signal Process. 49 (3) (2001) 530–541.
[37] K. Nordhausen, H. Oja, Independent subspace analysis using three scatter matrices, Aust. J. Stat. 40 (1&2) (2011) 93–101.
[38] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2016.
[39] J. Ramsay, B. Silverman, Functional Data Analysis, Springer, 2005.
[40] J.O. Ramsay, H. Wickham, S. Graves, G. Hooker, Fda: Functional data analysis, 2014, R package version 2.4.4.
[41] B.B. Risk, D.S. Matteson, D. Ruppert, Linear non-Gaussian component analysis via maximum likelihood, J. Amer. Statist. Assoc. 114 (525) (2019) 332–343.
[42] P.J. Rousseeuw, J. Raymaekers, M. Hubert, A measure of directional outlyingness with applications to image data and video, J. Comput. Graph. Statist. 27 (2) (2018) 345–359.
[43] Y. Sato, Theoretical considerations for multivariate functional data analysis, in: Proceedings 59th ISI World Statistics Congress, 2013, pp. 25–30.
[44] J. Song, B. Li, On additive functional principal component analysis, 2017. Submitted.
[45] S. Tokushige, H. Yadohisa, K. Inada, Crisp and fuzzy k-means clustering algorithms for multivariate functional data, Comput. Stat. 22 (1) (2007) 1–16.
[46] W.N. Venables, B.D. Ripley, Modern Applied Statistics with S, fourth ed., Springer, New York, ISBN: 0-387-95457-0, 2002.
[47] J. Virta, B. Li, K. Nordhausen, H. Oja, Independent component analysis for tensor-valued data, J. Multivariate Anal. 162 (2017) 172–192.
[48] J. Virta, B. Li, K. Nordhausen, H. Oja, JADE for tensor-valued observations, J. Comput. Graph. Statist. 27 (3) (2018) 628–637.
[49] H. Wickham, ggplot2: Elegant Graphics for Data Analysis, Springer-Verlag New York, 2009.
[50] M. Yamamoto, Clustering of functional data in a low-dimensional subspace, Adv. Data Anal. Classif. 6 (3) (2012) 219–247.
[51] M. Yamamoto, H. Hwang, Dimension-reduced clustering of functional data via subspace separation, J. Classification 34 (2) (2017) 294–326.
[52] M. Yamamoto, Y. Terada, Functional factorial K-means analysis, Comput. Statist. Data Anal. 79 (2014) 133–148.