Functional independent component analysis: an extension of fourth-order blind identification

Bing Li

Penn State University, University Park, United States.

Germain Van Bever†

The Open University, Milton Keynes, United Kingdom.

Hannu Oja

University of Turku, Turku, Finland.

Radka Sabolová

The Open University, Milton Keynes, United Kingdom.

Frank Critchley

The Open University, Milton Keynes, United Kingdom.

Summary.

We extend Independent Component Analysis, and in particular Fourth-Order Blind Identification, to functional data. Two major problems arise in this extension: (i) the notion of "marginals" is not naturally defined for functional data and (ii) the covariance operator is, in general, non invertible. These limitations are tackled by reformulating the problem in a coordinate-free manner and by imposing natural restrictions on the mixing model. The proposed procedure, which involves simultaneous diagonalisation of second- and fourth-order scatter operators, is shown to be Fisher consistent and affine invariant. A sample estimator is provided and illustrated on simulated and real datasets. In particular, it is shown to uncover particular structures that are missed by classical PCA in an Australian precipitation dataset.

Keywords: Affine invariance and equivariance, Coordinate-free approach, FOBI operator, Functional PCA, Hilbert space, Simultaneous diagonalisation

1. Introduction

Independent Component Analysis (ICA) was originally introduced as a technique to isolate several independent signals based on the data produced by several receivers, which contain the signals only in mixed forms, a problem sometimes referred to as the *cocktail-party problem* (Hyvärinen and Oja,

†Correspondence should be addressed to: germain.van-bever@open.ac.uk. Postal address: Germain Van Bever, The Open University, MCT Faculty, Walton Hall, MK7 6AA, Milton Keynes, United Kingdom.

2000). From a statistical perspective, ICA can also be viewed as a refinement of Principal Component Analysis (PCA), in the sense that it can recover the patterns hidden in the higher moments of random vectors that cannot be identified by classical PCA. Since its introduction, ICA has found wide application far beyond its original context, such as separating brain activities from artifacts in Magnetoencephalography data (Vigário et al., 1998), extracting features in financial time series (Kiviluoto and Oja, 1998), denoising images (Hyvärinen, 1999), and separating users' signals from interfering signals in telecommunication (Ristaniemi and Joutsensalo, 1999; Cristescu et al., 2000).

In this paper, we extend ICA to functional data, where the observed units are random functions rather than random vectors. Functional data are increasingly prevalent in modern research and daily life. Stock prices, brain signals, daily temperatures and precipitations, and longitudinal studies are but a few examples of the many forms of functional data arising in modern research. Meanwhile, devices such as smartphones and smart wristbands make real-time recordings of personal positions, movements, and vital signs commonplace. The past two decades have seen vigorous developments of theories and methodologies for processing and analysing functional data. Many statistical methods, such as linear regression, principal component analysis, canonical correlation, and sufficient dimension reduction have already been extended to the functional setting. See, for example, Ramsay and Silverman (2005), Yao, Müller and Wang (2005a,b), Ferraty and Vieu (2006), Horváth and Kokoszka (2012), Ferré and Yao (2003, 2005), and Hsing and Ren (2009). To distinguish between classical ICA for random vectors and its extension to functional data considered in this paper, we refer to the former as classical ICA, and to the latter as functional ICA.

Let X be a p-dimensional random vector, whose entries are statistically dependent, representing mixtures of independent signals. The classical independent component model assumes the existence of a nonsingular $p \times p$ matrix Ψ such that

$$X = \mu + \Psi Z, \tag{1}$$

where Z is a random vector whose entries are independent, representing the source signals. The goal of ICA is then to recover, under this model, a matrix A such that the entries of AX are independent. This process is called *demixing*, see Tyler et al. (2009) and Miettinen et al. (2015). Many demixing methods have been developed for classical ICA and can be roughly divided into three types: (i) those that are based on higher moments, such as fourth-order blind identification (FOBI, Cardoso, 1989; Oja, Sirkiä and Eriksson, 2006) and joint approximate diagonalisation of eigenmatrices (JADE, Cardoso and Souloumiac, 1993; Hyvärinen, 1999); (ii) those that are derived from semiparametric principles (Chen and Bickel, 2006; Samworth and Yuan, 2012); (iii) those that are derived from invariance arguments and semiparametrically efficient inference and based on multivariate ranks and signed ranks (Ilmonen and Paindaveine, 2011; Hallin and Mehta, 2015). In this paper we focus on the extension of the FOBI procedure, which is the original and most commonly used ICA method. FOBI has a simple structure and its extension is the most transparent. Moreover, the basic idea developed

for the extension of FOBI can be applied to extend other ICA procedures, such as JADE.

The first conceptual hurdle in our generalisation is that the classical formulation of (1) relies on the assumption that the entries of the vector Z are independent. This notion of "entries of a vector" does not arise naturally for random functions and we thus first reformulate the classical ICA in a coordinate-free manner, so that it no longer relies on the marginals of Z. This reformulation not only serves as a stepping stone for the extension, but also provides fresh insights into classical ICA itself.

Our overall approach to the generalisation is to replace (i) the Euclidean space in classical ICA by a Hilbert space of functions on an interval; (ii) moment matrices such as the covariance matrix by linear operators defined on the Hilbert space and (iii) eigenvectors by eigenfunctions. Essentially, the process of demixing a random function boils down to that of simultaneous diagonalisation of two self-adjoint linear operators defined on the functional space in which the random function X takes values. At the sample level, these operators can be written as finite-dimensional matrices using coordinate representation, which are then diagonalised simultaneously using linear algebra.

Gutch and Theis (2012) proposes an extension of classical ICA to the case where X is a sequence of random variables. This is related but different to our generalisation: in our case, the support of the random function X is an interval rather than the set of natural numbers. In particular, in their setting, there still exist natural components of Z, and therefore a more immediate analogy to the finite-dimensional case. Peña et al. (2014) explores an extension of kurtosis to functional data as a way to identify outliers and cluster structures. Their focus is, however, put on classification, implementation and theoretical properties under mixtures of Gaussian processes rather than a functional version of ICA.

The rest of the paper is organised as follows. In Section 2, we give a coordinate-free reformulation of classical ICA. We then (Section 3) introduce the functional independent component model by extending the coordinate-free version of (1). While doing so, we also develop machinery and notation of function spaces, operators, and tensor products that are useful for later discussions. In Section 4, we introduce the FOBI operator and the related FOBI demixing procedure; we also develop the population-level properties of the FOBI operator. Fisher consistency of the FOBI demixing procedure (that is, under the functional independent component model, it produces a random function that has independent components) is established in Section 5, while its affine invariance is discussed in Section 6. The sample-level implementation of the procedure is developed in Section 7 via coordinate representations. Its performance on simulated and real datasets is illustrated in Section 8. Finally, we make some concluding remarks in Section 9, while all proofs are collected in the Appendix.

2. Coordinate-free formulation of classical ICA

In this section, we reformulate the classical independent component analysis (ICA) for random vectors in a coordinate-free manner, so that it can be generalised to random functions. We also outline a

corresponding reformulation of the classical FOBI demixing procedure.

Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $Z : \Omega \to \mathbb{R}^p$ be a p-dimensional random vector on $(\Omega, \mathcal{F}, P)$. For simplicity and without loss of generality, assume $E(Z) = 0$. In classical ICA, Z is said to have independent components if $Z^1 \perp\!\!\!\perp \cdots \perp\!\!\!\perp Z^p$, where $\perp\!\!\!\perp$ indicates independence and $Z^1, \ldots, Z^p$ are the components of Z. A random vector $X : \Omega \to \mathbb{R}^p$ is said to follow an independent component model with respect to Z if and only if $X = \Psi Z$ for some nonsingular matrix $\Psi \in \mathbb{R}^{p \times p}$.

The difficulty with generalising ICA to the functional setting using this formulation is that we do not have a natural analogue of "components" of a random function. Specifically, if Z is a random function on an interval $J \subset \mathbb{R}$, then it is by no means clear what aspects of Z should be regarded as the "most natural" components of Z. It might be tempting to think of the set of variables $\{Z(t) : t \in J\}$ as corresponding to $Z^1, \ldots, Z^p$ above. However, because J is an uncountable set, using this analogy is difficult, involving too many technicalities that may cloud the main idea of the extension. Moreover, there is no order of importance between $Z(t_1)$ and $Z(t_2)$ for two points $t_1, t_2 \in J$, which is important for our development.

For these reasons, we first reformulate classical ICA in a way that does not rely on Euclidean coordinates, but rather in terms of *any* orthonormal basis (ONB) $\{v_1, \ldots, v_p\}$ in $\mathbb{R}^p$. Recall that our goal is to find A such that $AX = Z$, where Z has independent components. Writing $B = VA$ for $V = (v_1 \cdots v_p)$, it follows immediately that $(AX)^1 \perp\!\!\!\perp \cdots \perp\!\!\!\perp (AX)^p$ is equivalent to $(BX)^{\mathsf{T}} v_1 \perp\!\!\!\perp \cdots \perp\!\!\!\perp (BX)^{\mathsf{T}} v_p$, where Y^i stands for the ith entry of the vector Y.

This leads to the following definition of independent component model that is not specific to Euclidean coordinates. In order to generalise to abstract spaces later on, we will adopt the set of eigenvectors of $\Sigma_X = \mathrm{var}(X)$ as natural choice of ONB. To avoid ambiguity caused by multiplicity in eigenvalues, we say that an ONB $\{v_1, \ldots, v_p\}$ is a Σ_X-ONB if $\Sigma_X v_i \propto v_i$ for each i.

Definition 2.1 *Suppose $E\|X\|^2 < \infty$ and $\{v_1, \ldots, v_p\}$ is a Σ_X-ONB. We say that X follows an independent component model if there is a matrix $\Gamma \in \mathbb{R}^{p \times p}$ such that*

$$(\Gamma X)^{\mathsf{T}} v_1 \perp\!\!\!\perp \cdots \perp\!\!\!\perp (\Gamma X)^{\mathsf{T}} v_p.$$

If this condition is satisfied then we write $X \sim \mathrm{ICM}(\Gamma)$.

Note that Γ plays the role of Ψ^{-1} in the original formulation (1). We now turn to the corresponding reformulation of the FOBI estimator at the population level. For a generic random vector $S \in \mathbb{R}^p$, define the following matrix-valued function of its fourth moments

$$\mathrm{fobi}(S) = E[(SS^{\mathsf{T}})^2]. \tag{2}$$

Under reasonably mild conditions it can be shown that, if $O\Lambda O^{\mathsf{T}}$ is the spectral decomposition of $\mathrm{fobi}(\Sigma_X^{-1/2} X)$, where O is an orthogonal matrix and Λ is a diagonal matrix, then $O^{\mathsf{T}} \Sigma_X^{-1/2} X$ has independent components; that is, $(O^{\mathsf{T}} \Sigma_X^{-1/2} X)^1 \perp\!\!\!\perp \cdots \perp\!\!\!\perp (O^{\mathsf{T}} \Sigma_X^{-1/2} X)^p$.

This result can be restated in a coordinate free fashion as follows. For a given Σ_X-ONB $V = (v_1, \ldots, v_p)$, reexpress any vector $u \in \mathbb{R}^p$ as coordinates relative to the Σ_X-ONB; that is, $u = V V^\mathsf{T} u$, where $V^\mathsf{T} u$ are the coordinates of u in the basis $\{v_1, \ldots, v_p\}$. The coordinates of X with respect to this basis are $\tilde{X} = V^\mathsf{T} X$. Then finding B such that $(BX)^\mathsf{T} v_1 \perp\!\!\!\perp \cdots \perp\!\!\!\perp (BX)^\mathsf{T} v_p$ is equivalent to classical ICA problem of finding $\tilde{B}$ such that $(\tilde{B}\tilde{X})^1 \perp\!\!\!\perp \cdots \perp\!\!\!\perp (\tilde{B}\tilde{X})^p$, which can be achieved by the spectral decomposition of $\mathrm{fobi}(\Sigma_{\tilde{X}}^{-1/2} \tilde{X})$.

3. Functional independent component model

In order to provide the functional version of Definition 2.1, we now introduce the following notations and definitions.

Let J denote an interval in $\mathbb{R}$ and let $\mathscr{H}$ denote a separable Hilbert space of functions from J to $\mathbb{R}$. Let $\mathcal{B}$ be the Borel σ-field generated by the open sets in $\mathscr{H}$, as defined according to its norm $\| \cdot \|_{\mathscr{H}} = \langle \cdot, \cdot \rangle_{\mathscr{H}}^{1/2}$. A random element $X \in \mathscr{H}$ is any mapping from Ω to $\mathscr{H}$ that is $\mathcal{F}/\mathcal{B}$ measurable. We first define the mean and variance of X. The former is a member of $\mathscr{H}$, the latter is a linear operator from $\mathscr{H}$ to $\mathscr{H}$.

Under the assumption $E\|X\|_{\mathscr{H}} < \infty$, the linear functional $f \mapsto E\langle f, X \rangle_{\mathscr{H}}$ is bounded, because $|E\langle f, X \rangle_{\mathscr{H}}| \le \|f\|_{\mathscr{H}} E\|X\|_{\mathscr{H}}$. By Riesz's representation theorem, there is a member μ of $\mathscr{H}$ such that $\langle f, \mu \rangle_{\mathscr{H}} = E\langle f, X \rangle_{\mathscr{H}}$. We define μ as the mean of X and write it as EX. Throughout the paper, we assume $EX = 0$, the function in $\mathscr{H}$ that is everywhere 0. We can do so without loss of generality because all the operators concerned are invariant under translation $X \mapsto X + f$ for any fixed $f \in \mathscr{H}$.

To define the variance operator of X, we need the notions of tensor product, random operator, and expectation of a random operator. Let $\mathscr{B}(\mathscr{H})$ be the class of bounded linear operators from $\mathscr{H}$ to $\mathscr{H}$. For $f, g \in \mathscr{H}$, their tensor product $f \otimes g$ is the operator $\mathscr{H} \to \mathscr{H} : h \mapsto \langle g, h \rangle_{\mathscr{H}} f$. A random operator is a mapping from Ω to $\mathscr{B}(\mathscr{H})$ measurable with respect to the Borel σ-field in $\mathscr{B}(\mathscr{H})$. If W is a random operator such that $E\|W\| < \infty$, then the mapping $(f, g) \mapsto E\langle f, Wg \rangle_{\mathscr{H}}$ is a bounded bilinear form, which induces an operator $A \in \mathscr{B}(\mathscr{H})$ such that $\langle f, Ag \rangle_{\mathscr{H}} = E\langle f, Wg \rangle_{\mathscr{H}}$ (Conway, 1990, Theorem 2.2). We define A as the expectation of W and write it as EW.

Under $E\|X\|_{\mathscr{H}}^2 < \infty$, the linear operator

$$E[(X - EX) \otimes (X - EX)] \tag{3}$$

is bounded since for any $f \in \mathscr{H}$,

$$\|E[(X - EX) \otimes (X - EX)]f\|_{\mathscr{H}} = \|E[(X - EX)\langle X - EX, f \rangle_{\mathscr{H}}]\|_{\mathscr{H}}$$
$$\le E\left(|\langle X - EX, f \rangle_{\mathscr{H}}| \, \|X - EX\|_{\mathscr{H}}\right)$$
$$\le E(\|X - EX\|_{\mathscr{H}}^2)\|f\|_{\mathscr{H}}.$$

The operator (3) is called the variance operator and will be denoted by $\mathrm{var}(X)$. This type of construction originated in the functional analysis context by Baker (1973). See also Ferré and Yao (2003)

and Hsing and Eubank (2015). Moreover, it can be shown (Bonaccorsi and Priola, 2006) that $\mathrm{var}(X)$ is a trace-class operator, that is $tr(\mathrm{var}(X)) := \sum \langle \mathrm{var}(X)e_k, e_k \rangle$ is finite for any ONB $\{e_k : k \in \mathbb{N}_0\}$, if $E\|X\|_{\mathscr{H}}^2 < \infty$.

Similar to the finite-dimensional case, we call an ONB $\{f_i : i \in \mathbb{N}_0\}$ in $\mathscr{H}$ a Σ_x-ONB if $\Sigma_x f_i \propto f_i$ for each $i \in \mathbb{N}_0$. By Conway (1990), a Σ_x-ONB exists if Σ_x is a compact operator, which is true because Σ_x is trace-class. We now define a functional independent component model.

Definition 3.1 *Suppose X is a random element in $\mathscr{H}$ with $E\|X\|_{\mathscr{H}}^2 < \infty$ and let $\{f_i : i \in \mathbb{N}_0\}$ be a Σ_x-ONB. We say that X follows a functional independent component model with respect to an operator $\Gamma \in \mathscr{B}(\mathscr{H})$ if the sequence of random variables*

$$\{\langle \Gamma X, f_i \rangle_{\mathscr{H}} : \ i \in \mathbb{N}_0\}$$

is independent. In this case we write $X \sim \mathrm{FICM}(\Gamma)$.

The Σ_x-ONB may not be unique when some of its eigenvalues have multiplicity greater than 1. In that case we simply choose and fix a Σ_x-ONB. If $X \sim \mathrm{FICM}(I)$ for the identity mapping I, then the set of variables $\{\langle X, f_i \rangle_{\mathscr{H}} : \ i \in \mathbb{N}_0\}$ is independent. Following the tradition of classical ICA, we say that the random function X has independent components. In this sense, the process of demixing can be described as finding an operator $A \in \mathscr{B}(\mathscr{H})$ such that $AX \sim \mathrm{FICM}(I)$.

There is an additional difficulty in generalising estimators for classical ICA to the functional setting. Taking FOBI for example, in the classical setting, it relies on the standardised random vector $\Sigma_x^{-1/2}X$, which has covariance matrix equal to I_p, the $p \times p$ identity matrix. In the functional setting, however, such a standardised random function is not well defined, because Σ_x, being a trace-class operator, cannot be the identity mapping. While it is possible to generalise the notion of a random function to satisfy this condition, we choose to make an additional assumption to keep the structure simple.

Let $\mathscr{T}$ be a subspace of $\mathscr{H}$. We define $\mathscr{B}(\mathscr{T})$ as the class of all bounded linear operators from $\mathscr{T}$ to $\mathscr{T}$. Let $P_{\mathscr{T}}$ be the projection onto $\mathscr{T}$, and $Q_{\mathscr{T}} = I - P_{\mathscr{T}}$ the projection onto $\mathscr{T}^\perp$. We define

$$\mathscr{B}(\mathscr{H}|\mathscr{T}) = \{P_{\mathscr{T}}AP_{\mathscr{T}} + Q_{\mathscr{T}} : \ A \in \mathscr{B}(\mathscr{T})\}. \tag{4}$$

$\mathscr{B}(\mathscr{H}|\mathscr{T})$ is monotone in $\mathscr{T}$, in the sense that, if $\mathscr{T}' \subseteq \mathscr{T}''$ are subsets of $\mathscr{H}$, then $\mathscr{B}(\mathscr{H}|\mathscr{T}') \subseteq \mathscr{B}(\mathscr{H}|\mathscr{T}'')$. Moreover, it has upper bound $\mathscr{B}(\mathscr{H})$ (for $\mathscr{T} = \mathscr{H}$), and lower bound $\{I\}$ (for $\mathscr{T} = \{0\}$). In the next definition we impose a strong condition on functional independent component model to avoid standardisation in $\mathscr{H}$, as mentioned previously.

Definition 3.2 *Suppose X is a random element in $\mathscr{H}$ with $E\|X\|_{\mathscr{H}}^2 < \infty$ and let $\{f_i : i \in \mathbb{N}_0\}$ be a Σ_x-ONB. We say that X follows a kth order functional independent component model if there is a mapping $\Gamma \in \mathscr{B}(\mathscr{H}|\mathscr{T}_k)$ such that $X \sim \mathrm{FICM}(\Gamma)$, where $\mathscr{T}_k$ is the subspace spanned by $f_1, \ldots, f_k$. In this case we write $X \sim \mathrm{FICM}_k(\Gamma)$.*

Note that $\mathrm{FICM}(\Gamma)$ can be rewritten as $\mathrm{FICM}_\infty(\Gamma)$ and $X \sim \mathrm{FICM}_0(\Gamma)$ if and only if X has independent components. Also note that stating $X \sim \mathrm{FICM}(\Gamma)$ for $\Gamma \in \mathscr{B}(\mathscr{H}|\mathscr{T}_k)$ or stating $X \sim \mathrm{FICM}_k(\Gamma)$ is equivalent.

One can interpret the meaning of the additional assumption in Definition 3.2 as follows. For a gaussian random element $X \in \mathscr{H}$, the collection of random variables $\{\langle X, f_i \rangle_{\mathscr{H}} : i \in \mathbb{N}_0\}$ is always independent, because, whenever $i \neq j$,

$$\mathrm{cov}(\langle X, f_i \rangle_{\mathscr{H}}, \langle X, f_j \rangle_{\mathscr{H}}) = \langle f_i, \Sigma_X f_j \rangle_{\mathscr{H}} = 0.$$

Thus the components that are independent along the eigenfunctions of Σ_X are "gaussian like". In the signal processing context, gaussian components are usually regarded as uninteresting because they are isotropic. Thus Definition 3.2 can be interpreted as "there are only finitely many interesting signals in the random function X". Besides avoiding the standardisation problem, Definition 3.2 is a natural mechanism for regularisation. Indeed, ignoring small eigenvalues amounts to removing the high frequency components of X, which is a form of smoothing.

Our approach in Definition 3.2 places the interesting components of Z in the first k eigenfunctions of Σ_X. This is justified by the fact that Z is an unobserved random function, meaning that Γ can always be redefined so that the interesting components occur in these eigenfunctions. This also explains why it is reasonable to choose and fix a set $f_1, \ldots, f_k$ when the eigenvalues of Σ_X have multiplicities greater than 1.

Under the assumption $X \sim \mathrm{FICM}_k(\Gamma)$, our goal is to find an operator $A \in \mathscr{B}(\mathscr{H}|\mathscr{T}_k)$ such that AX has independent components. As in classical ICA, we call this process *demixing the random function X* and the operator A a *demixing operator*.

4. Functional FOBI operator and unitary equivariance

Let X be a random element in $\mathscr{H}$ such that $E\|X\|^4_{\mathscr{H}} < \infty$. Then the bilinear form

$$\mathscr{H} \times \mathscr{H} \to \mathbb{R}: \quad (f_1, f_2) \mapsto E\langle f_1, (X \otimes X)^2 f_2 \rangle_{\mathscr{H}}$$

is bounded because, by standard properties of the tensor product (see for example Lemma A.1 in the appendix) and the Cauchy-Schwarz inequality,

$$|E\langle f_1, (X \otimes X)^2 f_2 \rangle_{\mathscr{H}}| \leq |E\{\|X\|^2_{\mathscr{H}} \langle f_1, (X \otimes X) f_2 \rangle_{\mathscr{H}}\}|$$

$$\leq |E\{\|X\|^2_{\mathscr{H}} \langle f_1, X \rangle_{\mathscr{H}} \langle f_2, X \rangle_{\mathscr{H}}\}| \leq E(\|X\|^4_{\mathscr{H}}) \|f_1\|_{\mathscr{H}} \|f_2\|_{\mathscr{H}}.$$

This operator is well defined (following the same argument as for the variance operator) and is the analogue of the FOBI matrix in the functional setting. We give its formal definition.

Definition 4.1 *Let X be a random element of $\mathscr{H}$ with $E\|X\|^4_{\mathscr{H}} < \infty$. We call the operator $E[(X \otimes X)^2]: \mathscr{H} \to \mathscr{H}$ the FOBI operator, and write it as* $\mathrm{fobi}(X)$.

Like the variance operator, the FOBI operator is a trace-class operator, as shown in the next theorem.

Theorem 4.2 *If $E\|X\|^4 < \infty$, then* $\mathrm{fobi}(X)$ *is a trace-class operator.*

Operators such as $\mathrm{var}(X)$ and $\mathrm{fobi}(X)$ can be viewed as mappings from the class of all distributions of X to $\mathscr{B}(\mathscr{H})$. Such mappings are called ($\mathscr{B}(\mathscr{H})$-valued) statistical functionals. As statistical functionals, $\mathrm{var}(X)$ and $\mathrm{fobi}(X)$ both enjoy a type of equivariance that is important for our development. This property and its role in classical ICA were extensively studied in Tyler et al. (2009). As usual, let $P \circ X^{-1}$ denote the probability measure induced by X. Let $\mathscr{A}(\mathscr{H})$ be a class of operators in $\mathscr{B}(\mathscr{H})$. For each $A \in \mathscr{A}(\mathscr{H})$ and $b \in \mathscr{H}$, let $T_{A,b}$ be the transformation

$$\mathscr{H} \to \mathscr{H}: \quad h \mapsto Ah + b.$$

We now give a formal definition of equivariance of a statistical functional. Recall that the adjoint operator A^* of $A \in \mathscr{B}(\mathscr{H})$ is defined uniquely through $\langle Af, g \rangle_{\mathscr{H}} = \langle f, A^*g \rangle_{\mathscr{H}}$.

Definition 4.3 *A $\mathscr{B}(\mathscr{H})$-valued statistical functional ν is said to be affine equivariant with respect to $\mathscr{A}(\mathscr{H})$ if, for every $A \in \mathscr{A}(\mathscr{H})$, $b \in \mathscr{H}$,*

$$\nu(P \circ (AX + b)^{-1}) = A\nu(P \circ X^{-1})A^*.$$

Let $\mathscr{U}(\mathscr{H})$ denote the class of all unitary operators in $\mathscr{B}(\mathscr{H})$. We now show that $\mathrm{var}(X)$ and $\mathrm{fobi}(X)$ are equivariant with respect to $\mathscr{B}(\mathscr{H})$ and $\mathscr{U}(\mathscr{H})$, respectively.

Theorem 4.4 *Whenever they are defined, $\mathrm{var}(X)$ is affine equivariant with respect to $\mathscr{B}(\mathscr{H})$, and $\mathrm{fobi}(X)$ is unitary equivariant with respect to $\mathscr{U}(\mathscr{H})$. Specifically,*

(a) for each $A \in \mathscr{B}(\mathscr{H})$, $b \in \mathscr{H}$, $\mathrm{var}(AX + b) = A\,\mathrm{var}(X)A^*$;

(b) for each $U \in \mathscr{U}(\mathscr{H})$, $b \in \mathscr{H}$, $\mathrm{fobi}(UX + b) = U\,\mathrm{fobi}(X)U^*$.

5. Fisher consistency of functional FOBI

We now establish at the population-level that one can use the spectral decomposition of the FOBI operator to demix a random function $X \sim \mathrm{FICM}_k(\Gamma)$. This is a form of Fisher consistency since it means that, when evaluated at the true distribution, the FOBI procedure provides the targeted parameter, which in our case is the demixing operator.

We need to introduce a few more classes of operators similar to $\mathscr{B}(\mathscr{H}|\mathscr{T})$. As a general rule, for a subset $\mathscr{A}(\mathscr{H})$ of $\mathscr{B}(\mathscr{H})$, let

$$\mathscr{A}(\mathscr{H}|\mathscr{T}) = \{P_{\mathscr{T}}AP_{\mathscr{T}} + Q_{\mathscr{T}}, \; A \in \mathscr{A}(\mathscr{T})\}. \tag{5}$$

Similar to $\mathscr{B}(\mathscr{H}|\mathscr{T})$, $\mathscr{A}(\mathscr{H}|\mathscr{T})$ is monotone increasing in $\mathscr{T}$ with lower bound $\{I\}$ and upper bound $\mathscr{A}(\mathscr{H})$. The following special cases will be useful to our discussion. Recall that $\mathscr{T}_k \subseteq \mathscr{H}$ is the subspace spanned by $f_1, \ldots, f_k$, the first k members of the Σ_X-ONB of $\mathscr{H}$. Let

$$\mathscr{D}(\mathscr{T}_k) = \{\textstyle\sum_{i=1}^{k} d_i(f_i \otimes f_i) : \ d_1, \ldots, d_k \in \mathbb{R}\},$$

$$\mathscr{R}(\mathscr{T}_k) = \{\textstyle\sum_{i=1}^{k} d_i(f_i \otimes f_i) : \ d_1, \ldots, d_k \in \mathbb{R}, \ |d_1| = \cdots = |d_k| = 1\},$$

$$\mathscr{P}(\mathscr{T}_k) = \{\textstyle\sum_{i=1}^{k} (f_{\pi(i)} \otimes f_i) : \ \pi \text{ is an injection from } \{1, \ldots, k\} \text{ to } \{1, \ldots, k\}\},$$

$$\mathscr{U}(\mathscr{T}_k) = \text{the class of unitary operators on } \mathscr{T}_k.$$

A member of $\mathscr{D}(\mathscr{T}_k)$ resembles a diagonal matrix which diagonal elements $d_1, \ldots, d_k$; a member of $\mathscr{R}(\mathscr{T}_k)$ a diagonal matrix whose diagonal elements are either 1 or -1; a member of $\mathscr{P}(\mathscr{T}_k)$ a permutation matrix; and a member in $\mathscr{U}(\mathscr{T}_k)$ an orthogonal matrix. According to the rule in (5), we define

$$\mathscr{D}(\mathscr{H}|\mathscr{T}_k), \quad \mathscr{R}(\mathscr{H}|\mathscr{T}_k), \quad \mathscr{P}(\mathscr{H}|\mathscr{T}_k), \quad \mathscr{U}(\mathscr{H}|\mathscr{T}_k). \tag{6}$$

The next proposition gives some elementary properties of these classes. Their simple proofs are omitted.

Proposition 5.1 *The following statements hold true.*

(a) If $D_1, D_2 \in \mathscr{D}(\mathscr{H}|\mathscr{T}_k)$, then $D_1 D_2 \in \mathscr{D}(\mathscr{H}|\mathscr{T}_k)$;

(b) $\mathscr{P}(\mathscr{H}|\mathscr{T}_k) \subseteq \mathscr{U}(\mathscr{H}|\mathscr{T}_k)$, $\mathscr{R}(\mathscr{H}|\mathscr{T}_k) \subseteq \mathscr{U}(\mathscr{H}|\mathscr{T}_k)$;

(c) $U \in \mathscr{U}(\mathscr{T}_k)$ iff there is an ONB $\{g_1, \ldots, g_k\}$ of $\mathscr{T}_k$ such that $U = \sum_{i=1}^{k} g_i \otimes f_i$;

(d) $U \in \mathscr{U}(\mathscr{H}|\mathscr{T}_k)$ iff there is an ONB $\{g_1, \ldots, g_k\}$ such that

$$U = P_{\mathscr{T}_k}(\textstyle\sum_{i=1}^{k} g_i \otimes f_i)P_{\mathscr{T}_k} + Q_{\mathscr{T}_k} = \sum_{i=1}^{k} g_i \otimes f_i + Q_{\mathscr{T}_k}.$$

The next proposition shows that the independent component property in Definition 3.2 is invariant under transformations in $\mathscr{R}(\mathscr{H}|\mathscr{T}_k)$, $\mathscr{P}(\mathscr{H}|\mathscr{T}_k)$, and $\mathscr{P}(\mathscr{H}|\mathscr{T}_k)$ and their arbitrary compositions.

Proposition 5.2 *If $X \sim \mathrm{FICM}_k(\Gamma)$ for some $\Gamma \in \mathscr{B}(\mathscr{H}|\mathscr{T}_k)$, then*

$$X \sim \mathrm{FICM}_k(R\Gamma), \quad X \sim \mathrm{FICM}_k(D\Gamma), \quad X \sim \mathrm{FICM}_k(\Pi\Gamma).$$

for any $R \in \mathscr{R}(\mathscr{H}|\mathscr{T}_k)$, $D \in \mathscr{D}(\mathscr{H}|\mathscr{T}_k)$, and $\Pi \in \mathscr{P}(\mathscr{H}|\mathscr{T}_k)$. Consequently, $X \sim \mathrm{FICM}_k(A\Gamma)$, where A is the product of any permutation of these operators.

For a generic compact and self-adjoint operator $A \in \mathscr{B}(\mathscr{H})$ with first k eigenvalues $\lambda_1 \geq \cdots \geq \lambda_k$, define its partial power relative to k as

$$A^{\alpha(k)} = \textstyle\sum_{i=1}^{k} \lambda_i^{\alpha}(h_i \otimes h_i) + Q_{\mathscr{S}_k}, \tag{7}$$

where $h_1, \ldots, h_k$ are eigenfunctions corresponding to $\lambda_1, \ldots, \lambda_k$, $\mathscr{S}_k$ is the subspace spanned by $h_1, \ldots, h_k$, and $Q_{\mathscr{S}_k}$ is the projection onto $\mathscr{S}_k^{\perp}$. That is, we only raise to the power α the first k eigenvalues of A. In this notation, the usual power A^{α} may be written as $A^{\alpha(\infty)}$. The possibility of $\lambda_1, \ldots, \lambda_k$ with multiplicities higher than 1 in Definition (7) is not an issue here since $f_1, \ldots, f_k$ are chosen and fixed in our construction – as justified in the penultimate paragraph of Section 3.

Fisher consistency of the FOBI procedure for functional data is established in the following theorem.

Theorem 5.3 *Assume* $X \sim \mathrm{FICM}_k(\Gamma)$, $E\|X\|_{\mathscr{H}}^4 < \infty$, *and* $\mathrm{kurt}(\langle \Gamma X, f_i \rangle)$, $i = 1, \ldots, k$ *are distinct numbers. Let* $\sum_{i=1}^k \tau_i(h_i \otimes h_i)$ *be the spectral decomposition of*

$$P_{\mathscr{T}_k}\mathrm{fobi}(\Sigma_X^{-1/2(k)}X)P_{\mathscr{T}_k},$$

and let $V = \sum_{i=1}^k (h_i \otimes f_i) + Q_{\mathscr{T}_k}$. *Then* $\{\langle V^*\Sigma_X^{-1/2(k)}X, f_i \rangle_{\mathscr{H}} : i \in \mathbb{N}_0\}$ *is independent.*

Proof of Theorem 5.3 relies on the three following lemmas. We state them here as they provide beneficial understanding on the main steps of the proof. Lemma 5.4 shows that, after partial standardisation, X and Z only differ by an unitary transformation. Lemma 5.5 shows that if Z has independent components, then $\mathrm{fobi}(Z)$ must be diagonal. This is the key property that leads to Fisher consistency of the FOBI method. Finally, Lemma 5.6 shows that spectral decomposition of a particular operator is essentially unique (up to permutation and sign change).

Lemma 5.4 *Suppose* $E\|X\|_{\mathscr{H}}^2 < \infty$ *and* $X \sim \mathrm{FICM}_k(\Gamma)$. *Let* $Z = \Gamma X$. *Then there exists* $U \in \mathscr{U}(\mathscr{H}|\mathscr{T}_k)$ *such that* $\Sigma_X^{-1/2(k)}X = U\Sigma_Z^{-1/2(k)}Z$.

Lemma 5.5 *If* $E\|Z\|_{\mathscr{H}}^4 < \infty$ *and the sequence of random variables* $\{\langle Z, f_i \rangle_{\mathscr{H}} : i \in \mathbb{N}_0\}$ *is independent, then*

$$\mathrm{fobi}(Z) = \sum_{i \in \mathbb{N}_0} \nu_i(Z)(f_i \otimes f_i), \tag{8}$$

where $\nu_i(Z) = [\mathrm{var}(\langle Z, f_i \rangle_{\mathscr{H}})]^2[\mathrm{kurt}(\langle Z, f_i \rangle_{\mathscr{H}}) + E\|Z\|_{\mathscr{H}}^2/\mathrm{var}(\langle Z, f_i \rangle_{\mathscr{H}}) - 1]$.

Lemma 5.6 *Suppose* $\{f_1, \ldots, f_k\}$, $\{g_1, \ldots, g_k\}$ *and* $\{h_1, \ldots, h_k\}$ *are ONBs in* $\mathscr{T}_k$, *and* $\{c_1, \ldots, c_k\}$ *and* $\{d_1, \ldots, d_k\}$ *are sets of real numbers, where* $c_1, \ldots, c_k$ *are distinct. If* $\sum_{i=1}^k c_i(g_i \otimes g_i) = \sum_{i=1}^k d_i(h_i \otimes h_i)$, *then there exist* $R_0 \in \mathscr{R}(\mathscr{T}_k)$ *and* $\Pi_0 \in \mathscr{P}(\mathscr{T}_k)$ *such that*

$$\sum_{i=1}^k (g_i \otimes f_i) = \sum_{i=1}^k (h_i \otimes f_i)R_0\Pi_0.$$

Moreover, $\{d_1, \ldots, d_k\}$ *is a permutation of* $\{c_1, \ldots, c_k\}$.

We note that, when $i > k$, $\langle V^*\Sigma_X^{-1/2(k)}X, f_i \rangle_{\mathscr{H}}$ reduces to $\langle X, f_i \rangle$. When $i \in \{1, \ldots, k\}$,

$$\langle V^*\Sigma_X^{-1/2(k)}X, f_i \rangle_{\mathscr{H}} = \langle \Sigma_X^{-1/2(k)}X, Vf_i \rangle_{\mathscr{H}} = \langle \Sigma_X^{-1/2(k)}X, h_i \rangle_{\mathscr{H}} := \langle \tilde{X}, h_i \rangle_{\mathscr{H}}.$$

So Theorem 5.3 simply states that the sequence of random variables

$$\langle \tilde{X}, h_1 \rangle_{\mathscr{H}}, \ldots, \langle \tilde{X}, h_k \rangle_{\mathscr{H}}, \ \langle X, f_{k+1} \rangle_{\mathscr{H}}, \ldots$$

is independent. Because all the interesting signals are assumed to be in the first k components, we regard random variables $\langle \tilde{X}, h_1 \rangle_{\mathscr{H}}, \ldots, \langle \tilde{X}, h_k \rangle_{\mathscr{H}}$ as the final product of the FOBI demixing procedure.

These variables can be further simplified as follows. Note that

$$\tilde{X} = [\textstyle\sum_{i=1}^k \lambda_i^{-1/2}(f_i \otimes f_i) + \sum_{k+1}^{\infty}(f_i \otimes f_i)](\sum_{i \in \mathbb{N}_0} \langle X, f_i \rangle_{\mathscr{H}} f_i),$$

where λ_i are eigenvalues of Σ_X. Since $h_1, \ldots, h_k \in \mathscr{T}_k$, they are orthogonal to f_i for any $i > k$. Consequently,

$$\langle \tilde{X}, h_i \rangle_{\mathscr{H}} = \textstyle\sum_{\ell=1}^k \lambda_\ell^{-1/2} \langle X, f_\ell \rangle_{\mathscr{H}} \langle f_\ell, h_i \rangle_{\mathscr{H}}, \quad i = 1, \ldots, k.$$

These formulae will prove convenient for constructing the FOBI-demixed random variables at the sample level, in Section 7 below.

We finish this section by commenting on the distinct kurtoses assumption in Theorem 5.3. From the proof of Theorem 5.3, it is easy to see that all components with common kurtosis (in particular, all gaussian components) will be indistinguishable by the FOBI operator. In such a case, the spectral decomposition is no longer unique up to permutation or sign changes but also invariant under orthogonal transformation within the subspace spanned by these components. As a result, the components obtained from FOBI might not be independent, but will span the same subspace as components with common kurtosis do. This is a well-known property of FOBI (see, for example, Miettinen et al. (2015)). For simplicity, Theorem 5.3 and subsequent results are stated with the distinct kurtoses assumption, but can be easily generalised to common kurtoses in the obvious way.

6. Affine invariance of the FOBI demixing procedure

In Theorem 4.4 we showed that the FOBI operator is equivariant under any unitary transformation. Since the random function resulting from the FOBI demixing procedure is $V^* \Sigma_X^{-1/2(k)} X$, it is of interest to ask whether this random function also enjoys a degree of invariance. For ease of reference, we refer to the random function $V^* \Sigma_X^{-1/2(k)} X$ as the FOBI transformation of X, and denote it by $\mathrm{fobit}(X)$. It turns out that $\mathrm{fobit}(X)$ enjoys a stronger form of invariance than $\mathrm{fobi}(X)$ itself: $\mathrm{fobi}(X)$ is equivariant under transformations in the unitary class $\mathscr{U}(\mathscr{H})$, whereas $\mathrm{fobit}(X)$ is invariant under transformations in $\mathscr{B}(\mathscr{H}|\mathscr{T}_k)$.

Theorem 6.1 *Suppose*

a. X is a random function in $\mathscr{H}$ with $E\|X\|_{\mathscr{H}}^4 < \infty$ and $\{f_i : i \in \mathbb{N}_0\}$ is a Σ_X-ONB;

b. $X \sim \mathrm{FICM}_k(\Gamma)$ for some $\Gamma \in \mathscr{B}(\mathscr{H}|\mathscr{T}_k)$;

c. $\mathrm{kurt}(\langle \Gamma X, f_i\rangle_{\mathscr{H}})$, $i = 1, \dots, k$ *are all distinct.*

Then, for any injective operator $A \in \mathscr{B}(\mathscr{H}|\mathscr{T}_k)$, we have $\mathrm{fobit}(AX) = \mathrm{fobit}(X)$.

Note that this theorem hinges crucially on the IC assumption on X (see Lemma 5.4) as, for a general random function X, $\mathrm{fobit}(X)$ and $\mathrm{fobit}(AX)$ might differ. This is in clear parallel with the multivariate case, where the scatter matrix functional $\mathrm{fobi}(S) = E[(SS^{\mathsf{T}})^2]$ is orthogonally equivariant and the resulting fobi procedure is fully affine invariant under the IC model. Replacing $\mathrm{fobi}(X) = E[(SS^{\mathsf{T}})^2]$ by its fully affine equivariant version $E[(S\Sigma_S^{-1}S^{\mathsf{T}})SS^{\mathsf{T}}]$ provides an affine equivariant procedure for any random vector S. In our functional case, however, such an affine equivariant operator does not exist, since standardisation of X cannot be achieved.

7. Estimation

We turn now to the sample estimation procedure for functional ICA based on FOBI. The general principle here, as often, is to replace, wherever possible, the true distribution of $X(\cdot)$ by its empirical distribution generated by n independent copies $X_1(\cdot), \dots, X_n(\cdot)$. A further particularity of functional data is the fact that observations are only measured finitely many times on J, leaving the user with a sample

$$\left\{ X_{i,t_j} = X_i(t_j) : i = 1, \dots, n; \; j = 1, \dots, T_i \right\},$$

rather than the functions themselves. The first step in any statistical procedure dealing with functional data is therefore to reconstruct $X_i(t)$ based on its measurements. This is typically achieved by selecting a basis $\{\phi_i(t) : i \in \mathbb{N}_0\}$ of $\mathscr{H}$ and, denoting $\mathscr{H}_m = \{\sum_{i=1}^m c_i \phi_i(t) : (c_1, \dots, c_m) \in \mathbb{R}^m\}$, by assuming that $X_i(t)$ belongs to $\mathscr{H}_m$, the subset of $\mathscr{H}$ spanned by the first m basis elements. The reconstructed curves are $\hat{X}_i(t) = \sum_{i=1}^m \hat{c}_i \phi_i(t)$, for

$$(\hat{c}_1, \dots, \hat{c}_m) = \mathrm{argmin}_{(c_1, \dots, c_m)} \sum_{j=1}^{T_i} \left(X_{i,t_j} - \sum_{i=1}^m c_i \phi_i(t_j) \right)^2.$$

The nature of the data and personal preferences allow for a guided choice of basis functions and m. Ramsay and Silverman (2005) and Ferraty and Vieu (2006), for example, make abundant use of Fourier or spline bases.

For any $f = \sum_{i=1}^m c_i \phi_i(t) \in \mathscr{H}_m$, we call $(c_1, \dots, c_m)^{\mathsf{T}}$ its coordinates and write them as $[f]$. For a linear operator $A : \mathscr{H}_m \to \mathscr{H}_m$, $[A]$ denotes its matrix of coordinates, which is the $m \times m$ matrix $([A\phi_1], \dots, [A\phi_m])$. In particular, we have $[Af] = [A][f]$ for any $A \in \mathscr{B}(\mathscr{H}_m)$, $f \in \mathscr{H}_m$. Let $E_n \hat{X} = n^{-1} \sum_{i=1}^n \hat{X}_i$, and let

$$\Sigma_{\hat{X}} = \mathrm{var}_n(\hat{X}) = n^{-1} \sum_{i=1}^n (\hat{X}_i - E_n \hat{X}) \otimes (\hat{X}_i - E_n \hat{X}).$$

Let $\tilde{X} = \Sigma_{\hat{X}}^{-1/2(k)}(\hat{X} - E_n \hat{X})$. Lemma 7.1 below provides the eigendecomposition of $\mathrm{fobi}_n(\tilde{X}) = E_n[(\tilde{X} \otimes \tilde{X})^2]$ based on its coordinates in $\mathscr{H}_m$.

Lemma 7.1 *The following holds:*

(a) $[\Sigma_{\hat{X}}] = \text{var}_n([\hat{X}])$ *and* $[\Sigma_{\hat{X}}^{\alpha(k)}] = \sum_{i=1}^{k} \lambda_i^{\alpha}[f_i][f_i]^{\mathsf{T}} + \sum_{i=k+1}^{r}[f_i][f_i]^{\mathsf{T}}$, *where* r *is the rank of* $\Sigma_{\hat{X}}$, $\{\lambda_i\}$ *are the eigenvalues of* $\Sigma_{\hat{X}}$, $\{f_i\}$ *are the eigenfunctions of* $\Sigma_{\hat{X}}$.

(b) *A function* f *is an eigenfunction of* $\Sigma_{\hat{X}}$ *if and only if* $[f]$ *is an eigenvector of* $\text{var}_n([\hat{X}])$. *A function* $g \in \mathscr{H}_m$ *is an eigenfunction of* $\text{fobi}_n(\tilde{X})$ *if and only if the vector* $[g]$ *is an eigenvector of* $\text{fobi}_n([\tilde{X}]) = E_n(([\tilde{X}][\tilde{X}]^{\mathsf{T}})^2)$.

(c) *Denoting by* $u_1, \ldots, u_k$ *and* $v_1, \ldots, v_k$ *the first* k *eigenvalues of* $\text{var}_n([\hat{X}])$ *and* $\text{fobi}_n([\tilde{X}])$, *respectively, the* i*th demixed variable is given by*

$$([\hat{X}] - E_n[\hat{X}])^{\mathsf{T}}(\sum_{l=1}^{k} \lambda_l^{-1/2} u_l u_l^{\mathsf{T}}) v_i.$$

Algorithmically, it can therefore be seen that the demixed components result from conducting FOBI on the first k standardised principal scores of $\hat{X}$. The choice of k is guided by the natural limitations of conducting PCA: take k large enough to explain most of the variance (say, 99%), but not too large to cause estimation inaccuracies in $\lambda_l^{-1/2}$.

8. Simulation and application

In this section, the FOBI procedure is applied to several datasets and its usefulness is shown. We start with simulated data before turning to a real dataset.

8.1. Simulated dataset

We first illustrate the use of functional FOBI (hereafter FFOBI) and compare the components obtained to the principal scores given by functional PCA (FPCA). To this end, an i.i.d. sample $X_1, \ldots, X_n$ is generated from a random function X on $[0, 1]$, which is defined as $X(t) = \sum_{\ell=1}^{m_0} C_\ell \phi_{0,\ell}(t)$, where $\{\phi_{0,\ell} : \ell = 1, \ldots, m_0\}$ is the Fourier basis

$$\{1, \ \cos(2\pi t), \ \sin(2\pi t), \ \ldots \ \cos(2r_0 \pi t), \ \sin(2r_0 \pi t)\}, \ \text{where } 2r_0 + 1 = m_0,$$

with $m_0 = 15$, $r_0 = 7$ and different values of n – namely, $n = 500$, $n = 2000$ and $n = 5000$ respectively.

The coefficient vector $C = (C_1, \ldots, C_{m_0})^{\mathsf{T}}$ follows an IC model and is constructed as $A_\rho Z_\tau$, where $A_\rho = \left(\text{diag}(A_4, |\rho| I_{(m_0-4)})\right)^{1/2}$, for A_4 a 4×4 matrix with diagonal 1 and off-diagonal elements ρ and the components Z^i of Z are independent power exponential random variables with density

$$f_{z^i}(x) = \frac{\beta_i}{2\alpha_i \Gamma(1/\beta_i)} \exp\{-(|x|/\alpha)^{\beta_i}\},$$

where $\Gamma(x)$ is Euler's Gamma function. The shape parameters α_i and β_i are uniquely chosen such that Z^i has variance 1 and $\beta_i := \beta_i(\tau) = 1 + 2\tau i$. Note that, in this case, $kurt(Z^i) = \kappa(\beta_i)$, where

$\kappa(x) = (\Gamma(5/x)\Gamma(1/x))/(\Gamma(3/x)^2)$. It is easy to see that $\kappa(x)$ is monotonely decreasing in x and converging to a constant.

The mixing matrix is such that the covariance matrix of C has eigenvalues $1 + 3\rho$ (associated to the direction $(1_4^\intercal, 0_4^\intercal)^\intercal$), $(1 - \rho)$ and ρ of respective multiplicities 1, 3 and $(m_0 - 4)$, the latter being associated to the last $(m_0 - 4)$ "noise" components. The correlation ρ controls the relative weight given to the first eigendirection in the signal. It is taken within $[-1/3, 1/2]$ to ensure positive definiteness of the scatter operator and a variance of the noise smaller than that of the signal.

The rationale behind the choice of parameters is the following: for $\tau = 0$, all the components have the same kurtosis. As τ increases, the difference $\Delta_i(\tau) = \kappa(\beta_i(\tau)) - \kappa(\beta_{i-1}(\tau))$ between the kurtoses changes. And, although $\Delta_2(\tau)$ is monotonely increasing, this is not so for $\Delta_i(\tau)$, $i \geq 3$, where separation initially increases before decreasing again, therefore varying the difficulty of the separation problem.

This simulation aims at measuring how both methods are reconstructing the original independent components. To do so, we generate 500 datasets and measure in each case the minimum distance correlation index between the obtained (FPCA or FFOBI) components and the original $(Z^1, \cdots, Z^k)$ as

$$C(Y, Z) = \frac{1}{\sqrt{k}} \min_{L,P} \|LPR - I\|,$$

where R is the correlation matrix between the random vectors Y and Z, P is a permutation matrix and L is a sign-changing matrix (i.e. diagonal with entries $+1$ or -1). See, for example, Nordhausen et al. (2011). Note that, in the functional case covered in this paper, this criterion not only measures the performance of the procedure under consideration but also the good reconstruction of the signal subspace, as different choice of bases might lead to different performances.

To ensure an independent choice of test basis, we used $\mathscr{H}_m$ generated by the first $m = 10$ spline basis functions, together with $k = 4$. Results were actually very similar for all classical bases (Fourier, spline, polynomial, exponential, etc.; the impact of not knowing the basis being negligible) and values of $k \geq 4$ (understandably, k too small may not permit correct recovery of the signal components).

Figure 1 shows the mean value of $C(Y, Z)$ for Z as above and Y being either FPCA or FFOBI scores for (i) a fixed value of $\rho_0 = 0.4$ and equally spaced values $\tau = 0, 0.005, 0.01, \cdots$ ranging in $[0, 1]$ and (ii) a fixed value of $\tau_0 = 0.5$ and $\rho = -0.3, -0.29, \cdots$ varying in its domain.

Note that the principal directions given by the covariance operator of X are not independent in this example. This is indeed confirmed by the overall poor performance of FPCA, for any value of τ and ρ. Note also that, when ρ gets closer to 0, the noise component disappears, so that the signal subspace is correctly reconstructed. FPCA still fails, however, to reconstruct the independent directions.

From both panels, we see that FFOBI reconstructs the original components quite well, using the fourth moments information to further rotate in the signal subspace. Interestingly, this good
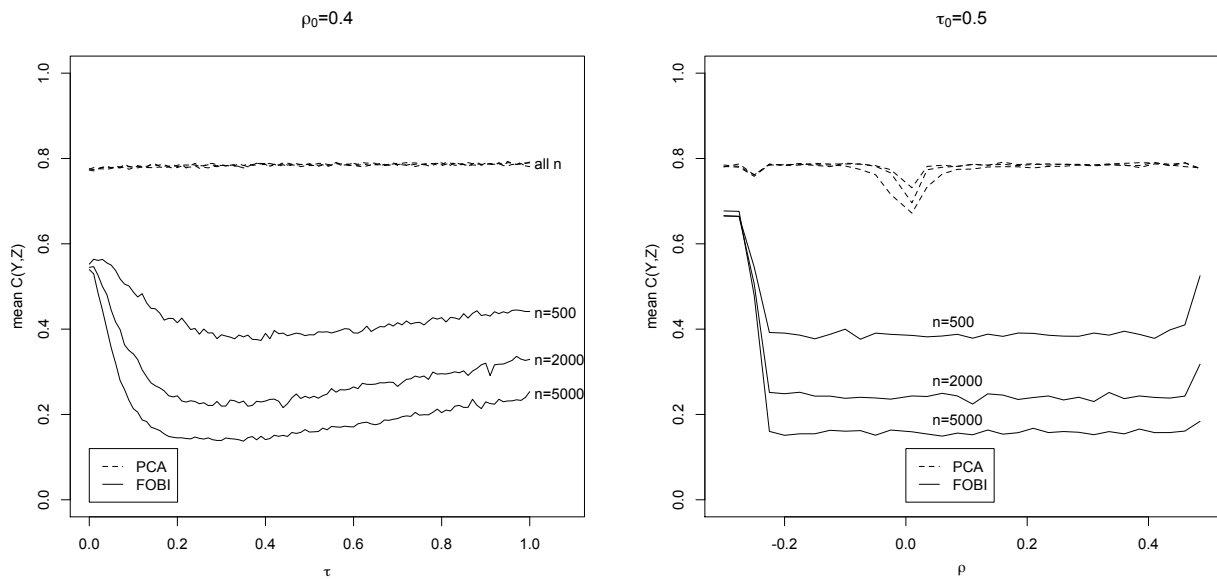
Fig. 1. Mean value of $C(Y, Z)$ for (left) fixed $\rho_0 = 0.4$ and $\tau \in [0, 1]$ and (right) fixed $\tau_0 = 0.5$ and $\rho \in [-0.3, 0.5]$. In both plots, plain lines are for Y being the FPCA scores and dashed lines for Y obtained from FFOBI. In both cases, $k = 4$.

performance holds for most values ρ, not just for those near 0, where the noise is absent. To illustrate this fact, the value ρ_0 chosen in the left panel is close to one of the boundaries and of comparable performance to those obtained with ρ closer to 0. This overall good performance is again illustrated in the right panel, where, for a fixed value of the shape parameter, FFOBI and FPCA behaves similarly for all values of ρ within the domain. Also, we see from the left panel that the minimum distance index is lowest when the values of the kurtoses are well separated. As n increases, consistent estimation of $\kappa(\beta_i)$ results in a decrease in the mean index.

Summarising, FFOBI provides a good methodology to reconstruct independent components, particularly so when the covariance operator has multiple eigenvalues or eigenvalues with close values.

8.2. Real dataset

Let us now turn to a real data example. This dataset consists of daily rainfall measurements in 191 weather stations in Australia, measured (non regularly) between 1840 and 1990 (a more detailed description can be found in Delaigle and Hall (2010)). The data can be accessed at http://rda.ucar.edu.

After removing an obviously outlying station from the data, averaging over the years and spline smoothing, our dataset consists of 190 curves $\hat{X}_i(t)$ representing the rainfall at time t (time passed in a given year) at the ith weather station. This dataset, presented in Figure 2 below, is known to have two types of station: (i) "tropical" ones (usually located north) for which the precipitations are heavier in summer (early in the year) and (ii) those for which the high precipitations occur in the cooler months.
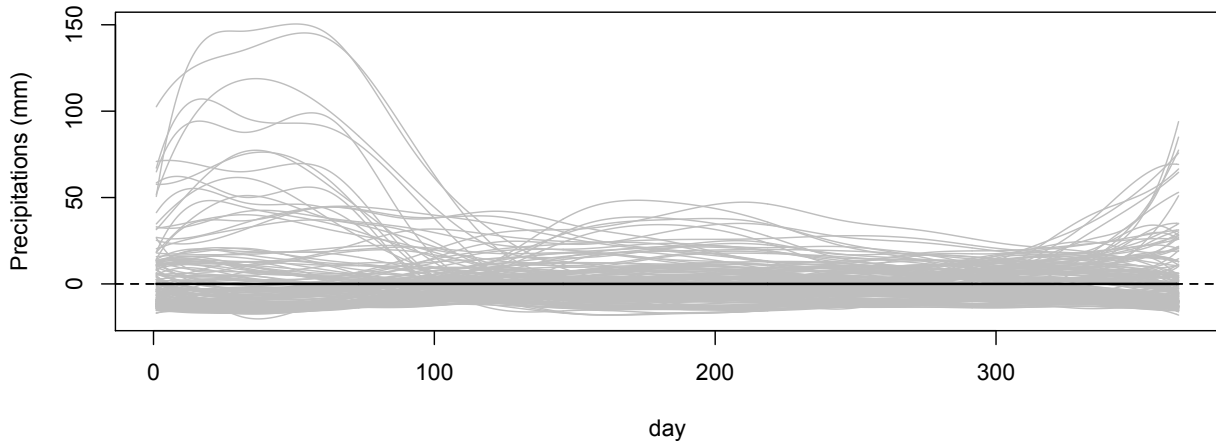
Fig. 2. Precipitations as a function of time in 190 different locations in Australia.

Figure 3 presents a histogram of the first principal score (which accounts for 73.2% of the total variation), together with the functional principal components. Interestingly, discrimination between the two behaviors exposed above is well known to be accounted for by the first principal component, for which large values are associated to the tropical stations. This can be seen in the left panel of Figure 5, where the locations of the weather stations are shown, coloured according to the result of a 2-means clustering along FPCA's first component.
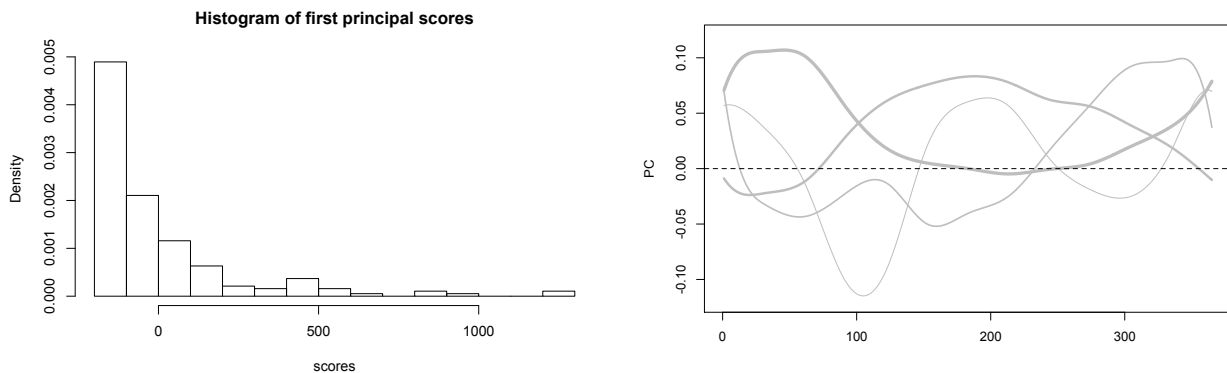
Fig. 3. Left: Histogram of the first principal scores. Right: $k = 4$ first principal functions. From thickest to thinnest, they account respectively for 73.2%, 22%, 3.1% and 1.1%.

Remarkably, the components obtained from conducting the FFOBI procedure allow to uncover another source of discrimination. Figure 4 presents the histogram of the fourth FFOBI scores (we chose in this simulation $k = 4$ as four principal components account for 99.4%). A clear bimodal structure is exhibited. Figure 4 also shows the eigenfunctions of the FFOBI operator. Interestingly, we see that the fourth functions emphasises strongly a period towards the end of the year and also, to a lesser extent, a few months between summer and winter. The right panel of Figure 5 colours the
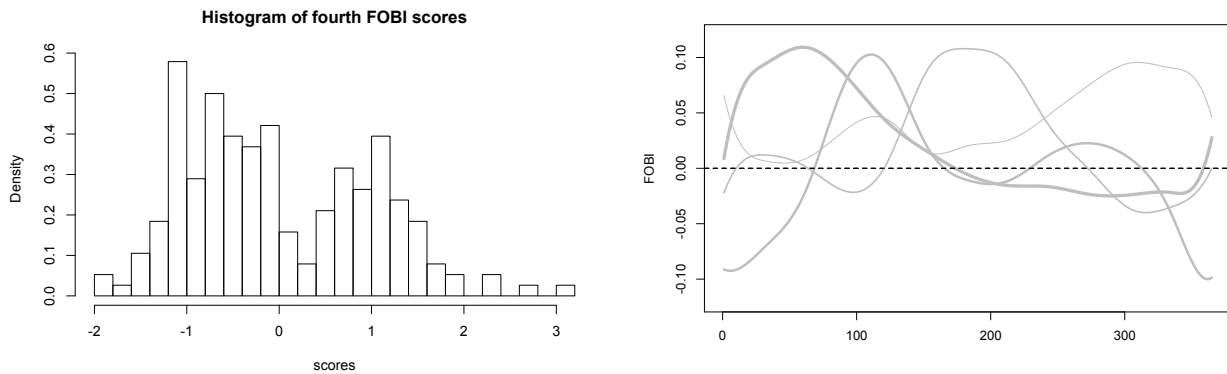
Fig. 4. Left: Histogram of the fourth FFOBI scores. Right: $k = 4$ FOBI functions. From thickest to thinnest, in decreasing order of kurtosis.

location of the weather stations according to the result of a 2-means clustering along FFOBI's fourth component. The East-West discrimination is striking (a few weather stations are wrongly classified, though, but are given a score that lies between both modes in the distribution) with the eastern stations famously known to suffer heavy rains during fall and spring, being hit by "Australian east coast lows", extratropical cyclones generated by the particular geography of the region (the Great Divide) together with the El Niño Southern Oscillation phenomenon (see, for example, Hopkins and Holland (1997)).

Note that this latter discrimination is not picked up by FPCA components, while the North-South behavior is still taken into account by the first FFOBI score, thereby showing the ability of the FFOBI transformation to uncover extra structure.
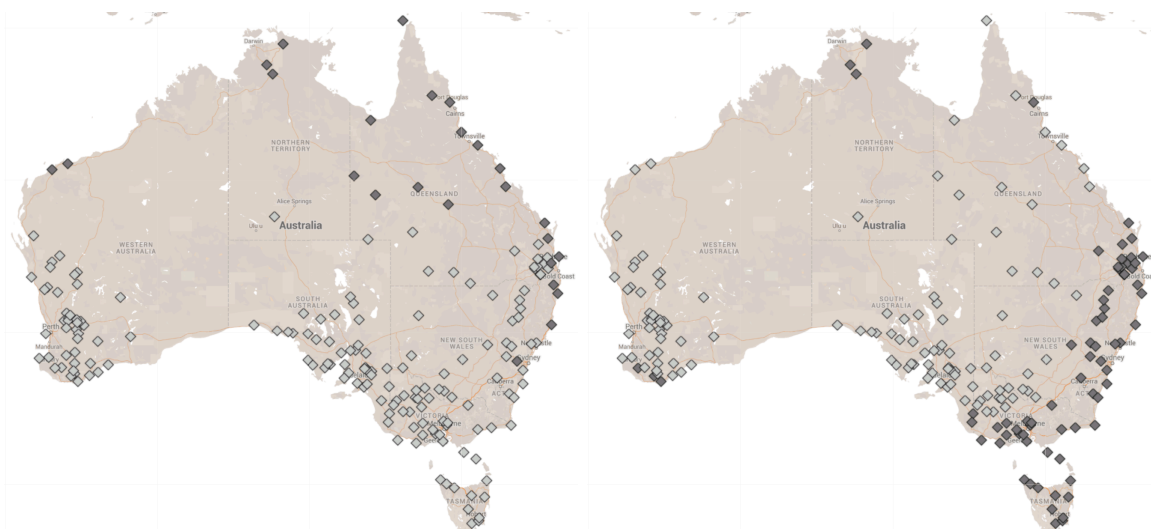

Fig. 5. Locations† of the 190 Australian weather stations together with their classification based on 2-means clustering on first principal scores (left) and fourth FOBI scores (right).

†Map copyright: Google©.

9. Concluding remarks

In this paper, we introduced a functional version of the fourth-order blind identification procedure which, under natural assumptions, reconstructs the independent components of a functional IC model. This novel construction is, however, one of few attempts to provide a dimension reduction methodology in the functional setup that goes *beyond* classical PCA.

While its Fisher consistency has been stated in this paper, many properties of functional FOBI remain to be explored but go certainly beyond the scope of this paper. We already stressed the importance of the selection of k and the sensitivity to the estimation of the covariance operator's eigenvalues. While asymptotic results are known for FPCA (see, for example, Bosq (2000) or Hall and Hosseini-Nasab (2006)), such results have never been extended to other scatter operators on Hilbert spaces. We hope to provide such a study in future work.

FOBI has also been proved to work well with mixtures of distributions. In more generality, Invariant Coordinate Selection (Tyler et al., 2009) has been proved to recover independent components in IC models as well as the Fisher subspace in certain mixture distributions. The possibility to provide a functional version of ICS is, however, far more complicated, as, to the best of our knowledge, there does not exist another fully affine equivariant scatter operator in $\mathscr{H}$. If one restricts purely to the FIC models, extensions of other methodologies based on moments (such as JADE, Cardoso and Souloumiac (1993)) using the same operator diagonalisation ideas could be considered.

A. Appendix

This appendix collects proofs of technical results. The following properties of tensor products are standard (see, for example, Fukumizu and Bach, 2007) but we record them here as a lemma for easy reference.

Lemma A.1 *If $f, g, h \in \mathscr{H}$, and $A, B \in \mathscr{B}(\mathscr{H})$, then*

(a) $(g \otimes f)^* = f \otimes g$;

(b) $(Ag) \otimes (Bf) = A(g \otimes f)B^*$;

(c) $(h \otimes g)(g \otimes f) = \|g\|_{\mathscr{H}}^2 (h \otimes f)$.

Proof of Theorem 4.2: Let $\{f_i : i \in \mathbb{N}_0\}$ be an orthonormal basis of $\mathscr{H}$. Then, by Lemma A.1,

$$\operatorname{tr}(\operatorname{fobi}(X)) = \sum_{i \in \mathbb{N}_0} \langle f_i, \operatorname{fobi}(X) f_i \rangle_{\mathscr{H}} = \sum_{i \in \mathbb{N}_0} E(\|X\|_{\mathscr{H}}^2 \langle f_i, X \rangle_{\mathscr{H}}^2)$$

The right-hand side is the limit

$$\lim_{n \to \infty} \sum_{i=1}^n E(\|X\|_{\mathscr{H}}^2 \langle f_i, X \rangle_{\mathscr{H}}^2) = \lim_{n \to \infty} E(\|X\|_{\mathscr{H}}^2 \sum_{i=1}^n \langle f_i, X \rangle_{\mathscr{H}}^2).$$

Because

$$\|X\|^2_{\mathscr{H}} \sum_{i=1}^n \langle f_i, X\rangle^2_{\mathscr{H}} \le \|X\|^2_{\mathscr{H}} \sum_{i=1}^\infty \langle f_i, X\rangle^2_{\mathscr{H}} = \|X\|^4_{\mathscr{H}}, \quad E\|X\|^4_{\mathscr{H}} < \infty,$$

by the dominated convergence theorem

$$\lim_{n\to\infty} \sum_{i=1}^n E\left(\|X\|^2_{\mathscr{H}} \sum_{i=1}^n \langle f_i, X\rangle^2_{\mathscr{H}}\right) = E\left(\|X\|^2_{\mathscr{H}} \lim_{n\to\infty} \sum_{i=1}^n \langle f_i, X\rangle^2_{\mathscr{H}}\right) = E\|X\|^4_{\mathscr{H}} < \infty.$$

Hence $\mathrm{tr}(\mathrm{fobi}(X)) < \infty$. $\qquad\square$

Proof of Theorem 4.4: *(a)* We note that

$$\mathrm{var}((AX+b)) = E\left\{[(AX+b) - E(AX+b)] \otimes [(AX+b) - E(AX+b)]\right\}$$
$$= E[(AX) \otimes (AX)].$$

By Lemma A.1, the right-hand side is $AE(X \otimes X)A^* = A\,\mathrm{var}(X)A^*$.

(b) Similarly,

$$\mathrm{fobi}(UX+b) = E[\|UX\|^2_{\mathscr{H}}(UX) \otimes (UX)] \tag{9}$$

Note that $\|UX\|^2_{\mathscr{H}} = \langle X, U^*UX\rangle_{\mathscr{H}} = \|X\|^2_{\mathscr{H}}$ and, by Lemma A.1, $(UX) \otimes (UX) = U(X \otimes X)U^*$. Hence the right-hand side of (9) is $U\mathrm{fobi}(X)U^*$. $\qquad\square$

Proof of Proposition 5.2: The assertions about D and R are obvious. Consider now the assertion for $\Pi \in \mathscr{P}(\mathscr{H}|\mathscr{T}_k)$. Let $Z = \Gamma X$. Because $E\|Z\|^2_{\mathscr{H}} < \infty$, Σ_z is trace-class and Z can be expressed as $\sum_{i\in\mathbb{N}_0}\langle Z, f_i\rangle_{\mathscr{H}}f_i$. Consequently,

$$\Pi Z = \sum_{i=1}^k \sum_{j=1}^k (f_{\pi(i)} \otimes f_i)\langle Z, f_j\rangle_{\mathscr{H}}f_j + \sum_{i=k+1}^\infty \langle Z, f_i\rangle_{\mathscr{H}}f_i$$
$$= \sum_{i=1}^k \langle Z, f_i\rangle_{\mathscr{H}}f_{\pi(i)} + \sum_{i=k+1}^\infty \langle Z, f_i\rangle_{\mathscr{H}}f_i.$$

Hence,

$$\langle \Pi Z, f_i\rangle_{\mathscr{H}} = \begin{cases} \langle Z, f_{\pi^{-1}(i)}\rangle_{\mathscr{H}} & \text{for } i = 1, \ldots, k, \text{ and} \\ \langle Z, f_i\rangle_{\mathscr{H}} & \text{for } i = k+1, k+2, \ldots \end{cases}.$$

Thus the set of random variables $\{\langle \Pi Z, f_i\rangle_{\mathscr{H}} : i \in \mathbb{N}_0\}$ is independent. The last statement follows immediately. $\qquad\square$

Proof of Lemma 5.4: Since $X \sim \mathrm{FICM}_k(\Gamma)$, there exists a $\Gamma_0 \in \mathscr{B}(\mathscr{T}_k)$ such that

$$X = (P_{\mathscr{T}_k}\Gamma_0 P_{\mathscr{T}_k} + Q_{\mathscr{T}_k})Z.$$

By Theorem 4.4, part (a),

$$\Sigma_x = (P_{\mathscr{T}_k}\Gamma_0 P_{\mathscr{T}_k} + Q_{\mathscr{T}_k})\Sigma_z(P_{\mathscr{T}_k}\Gamma_0^* P_{\mathscr{T}_k} + Q_{\mathscr{T}_k}) = P_{\mathscr{T}_k}\Gamma_0 P_{\mathscr{T}_k}\Sigma_z P_{\mathscr{T}_k}\Gamma_0^* P_{\mathscr{T}_k} + Q_{\mathscr{T}_k}\Sigma_z Q_{\mathscr{T}_k}.$$

Hence $\Sigma_X^{-1/2(k)} = P_{\mathscr{T}_k}(\Gamma_0 P_{\mathscr{T}_k}\Sigma_Z P_{\mathscr{T}_k}\Gamma_0^*)^{-1/2}P_{\mathscr{T}_k} + Q_{\mathscr{T}_k}$, which implies

$$\Sigma_X^{-1/2(k)}X = \Sigma_X^{-1/2(k)}(P_{\mathscr{T}_k}\Gamma_0 P_{\mathscr{T}_k} + Q_{\mathscr{T}_k})\Sigma_Z^{1/2(k)}\Sigma_Z^{-1/2(k)}Z \equiv A\Sigma_Z^{-1/2(k)}Z.$$

It remains to show that $A \in \mathscr{U}(\mathscr{H}|\mathscr{T}_k)$. Put $\Sigma_Z^0 = P_{\mathscr{T}_k}\Sigma_Z P_{\mathscr{T}_k}$. Then,

$$\begin{aligned}
A &= (P_{\mathscr{T}_k}(\Gamma_0\Sigma_Z^0\Gamma_0^*)^{-1/2}P_{\mathscr{T}_k} + Q_{\mathscr{T}_k})(P_{\mathscr{T}_k}\Gamma_0 P_{\mathscr{T}_k} + Q_{\mathscr{T}_k})(P_{\mathscr{T}_k}(\Sigma_Z^0)^{1/2}P_{\mathscr{T}_k} + Q_{\mathscr{T}_k}) \\
&= P_{\mathscr{T}_k}(\Gamma_0\Sigma_Z^0\Gamma_0^*)^{-1/2}\Gamma_0(\Sigma_Z^0)^{1/2}P_{\mathscr{T}_k} + Q_{\mathscr{T}_k}.
\end{aligned}$$

So, showing $A \in \mathscr{U}(\mathscr{H}|\mathscr{T}_k)$ amounts to show $U = (\Gamma_0\Sigma_Z^0\Gamma_0^*)^{-1/2}\Gamma_0(\Sigma_Z^0)^{1/2} \in \mathscr{U}(\mathscr{T}_k)$. We have

$$\begin{aligned}
UU^* &= [(\Gamma_0\Sigma_Z^0\Gamma_0^*)^{-1/2}\Gamma_0(\Sigma_Z^0)^{1/2}][(\Gamma_0\Sigma_Z^0\Gamma_0^*)^{-1/2}\Gamma_0(\Sigma_Z^0)^{1/2}]^* \\
&= [(\Gamma_0\Sigma_Z^0\Gamma_0^*)^{-1/2}\Gamma_0(\Sigma_Z^0)^{1/2}][(\Sigma_Z^0)^{1/2}\Gamma_0^*(\Gamma_0\Sigma_Z^0\Gamma_0^*)^{-1/2}] = I_{\mathscr{T}_k},
\end{aligned}$$

where $I_{\mathscr{T}_k}$ denotes the identity operator in $\mathscr{T}_k$. Similarly,

$$\begin{aligned}
U^*U &= [(\Gamma_0\Sigma_Z^0\Gamma_0^*)^{-1/2}\Gamma_0(\Sigma_Z^0)^{1/2}]^*[(\Gamma_0\Sigma_Z^0\Gamma_0^*)^{-1/2}\Gamma_0(\Sigma_Z^0)^{1/2}] \\
&= [(\Sigma_Z^0)^{1/2}\Gamma_0^*(\Gamma_0\Sigma_Z^0\Gamma_0^*)^{-1/2}][(\Gamma_0\Sigma_Z^0\Gamma_0^*)^{-1/2}\Gamma_0(\Sigma_Z^0)^{1/2}] = I_{\mathscr{T}_k}.
\end{aligned}$$

Hence $U \in \mathscr{U}(\mathscr{T}_k)$. $\qquad\square$

Proof of Lemma 5.5: Since Z can be expanded as $\sum_{i\in\mathbb{N}_0}\langle Z, f_i\rangle_{\mathscr{H}}f_i$, we have

$$\begin{aligned}
\mathrm{fobi}(Z) &= E[\langle Z, Z\rangle_{\mathscr{H}}(Z \otimes Z)] \\
&= \sum_{i,j,k,\ell\in\mathbb{N}_0}E[\langle\langle Z, f_i\rangle_{\mathscr{H}}f_i, \langle Z, f_j\rangle_{\mathscr{H}}f_j\rangle_{\mathscr{H}}(\langle Z, f_k\rangle_{\mathscr{H}}f_k \otimes \langle Z, f_\ell\rangle_{\mathscr{H}}f_\ell)] \\
&= \sum_{i,j,k,\ell\in\mathbb{N}_0}\langle f_i, f_j\rangle_{\mathscr{H}}E[\langle Z, f_i\rangle_{\mathscr{H}}\langle Z, f_j\rangle_{\mathscr{H}}(\langle Z, f_k\rangle_{\mathscr{H}}f_k \otimes \langle Z, f_\ell\rangle_{\mathscr{H}}f_\ell)] \\
&= \sum_{i,k,\ell\in\mathbb{N}_0}E[\langle Z, f_i\rangle_{\mathscr{H}}^2\langle Z, f_k\rangle_{\mathscr{H}}\langle Z, f_\ell\rangle_{\mathscr{H}}](f_k \otimes f_\ell).
\end{aligned} \qquad (10)$$

Because the sequence of random variables $\{\langle Z, f_i\rangle_{\mathscr{H}} : i \in \mathbb{N}_0\}$ is independent, we have, for any $k \neq \ell$ and any $i \in \mathbb{N}_0$, $E[\langle Z, f_i\rangle_{\mathscr{H}}^2\langle Z, f_k\rangle_{\mathscr{H}}\langle Z, f_\ell\rangle_{\mathscr{H}}] = 0$. Hence, (10) reduces to

$$\begin{aligned}
\mathrm{fobi}(Z) &= \sum_{i,j\in\mathbb{N}_0}E[\langle Z, f_i\rangle_{\mathscr{H}}^2\langle Z, f_j\rangle_{\mathscr{H}}^2](f_i \otimes f_i) \\
&= \sum_{i\in\mathbb{N}_0}E\langle Z, f_i\rangle_{\mathscr{H}}^4(f_i \otimes f_i) + E[\langle Z, f_i\rangle_{\mathscr{H}}^2](f_i \otimes f_i)\sum_{j\neq i}E[\langle Z, f_j\rangle_{\mathscr{H}}^2].
\end{aligned}$$

However, since $E(Z) = 0$, we have $E\langle Z, f_i\rangle_{\mathscr{H}}^2 = \mathrm{var}(\langle Z, f_i\rangle_{\mathscr{H}})$. Moreover, $\mathrm{kurt}(\langle Z, f_i\rangle_{\mathscr{H}})$ is defined through the relation $E\langle Z, f_i\rangle_{\mathscr{H}}^4 = [\mathrm{var}(\langle Z, f_i\rangle_{\mathscr{H}})]^2\mathrm{kurt}(\langle Z, f_i\rangle_{\mathscr{H}})$. Hence the right-hand side above can be rewritten as

$$\sum_{i\in\mathbb{N}_0}\{[\mathrm{var}(\langle Z, f_i\rangle_{\mathscr{H}})]^2\mathrm{kurt}(\langle Z, f_i\rangle_{\mathscr{H}}) + [\mathrm{var}(\langle Z, f_i\rangle_{\mathscr{H}})]\sum_{j\neq i}[\mathrm{var}(\langle Z, f_j\rangle_{\mathscr{H}})]\}(f_i \otimes f_i).$$

The desired equality now follows from $E\|Z\|_{\mathscr{H}}^2 = \sum_{j\in\mathbb{N}_0}\mathrm{var}(\langle Z, f_i\rangle_{\mathscr{H}})$. $\qquad\square$

Proof of Lemma 5.6: Let $A = \sum_{i=1}^k c_i(f_i \otimes f_i)$. Then A is a self-adjoint operator in $\mathscr{B}(\mathscr{T}_k)$ with spectral decomposition $\sum_{i=1}^k c_i(f_i \otimes f_i)$. Because $c_1, \dots, c_k$ are distinct, the assertion follows from

the uniqueness of the spectral decomposition of a self adjoint operator with finite-dimensional range (Conway, 1990). $\hfill\square$

Proof of Theorem 5.3: Let $Z = \Gamma X$, $\tilde{X} = \Sigma_X^{-1/2(k)} X$, and $\tilde{Z} = \Sigma_Z^{-1/2(k)} Z$. Then, by Lemma 5.4, $\tilde{X} = U\tilde{Z}$ for some $U \in \mathscr{U}(\mathscr{H}|\mathscr{T}_k)$. Since $\mathscr{U}(\mathscr{H}|\mathscr{T}_k) \subseteq \mathscr{U}(\mathscr{H})$, by the unitary equivariance of the FOBI operator in Theorem 4.4, $\mathrm{fobi}(\tilde{X}) = U\mathrm{fobi}(\tilde{Z})U^*$. Since Z has independent components, and $\Sigma_Z^{-1/2(k)} \in \mathscr{D}(\mathscr{H}|\mathscr{T}_k)$, by Proposition 5.2, $\tilde{Z}$ also has independent components. By Lemma 5.5, then

$$\mathrm{fobi}(\tilde{Z}) = \sum_{i=1}^{\infty}\nu_i(\tilde{Z})(f_i \otimes f_i). \tag{11}$$

Because $U \in \mathscr{U}(\mathscr{H}|\mathscr{T}_k)$, by Proposition 5.1, it can be written as $\sum_{i=1}^{k} g_i \otimes f_i + Q_{\mathscr{T}_k}$ for some ONB $\{g_1, \ldots, g_k\}$ of $\mathscr{T}_k$. Thus

$$\mathrm{fobi}(\tilde{X}) = (\sum_{i=1}^{k} g_i \otimes f_i + Q_{\mathscr{T}_k})[\sum_{i=1}^{k}\nu_i(\tilde{Z})(f_i \otimes f_i) + \sum_{i=k+1}^{\infty}\nu_i(\tilde{Z})(f_i \otimes f_i)]$$
$$(\sum_{i=1}^{k} g_i \otimes f_i + Q_{\mathscr{T}_k})^*.$$

The right-hand side is of the form $(A_1+B_1)(A_2+B_2)(A_3+B_3)$, where A_1, A_2, A_3 are members of $\mathscr{B}(\mathscr{T}_k)$ and B_1, B_2, B_3 are members of $\mathscr{B}(\mathscr{T}_k^{\perp})$. Hence the product can be written as $A_1A_2A_3 + B_1B_2B_3$. These two terms are computed as follows

$$A_1A_2A_3 = (\sum_{i=1}^{k} g_i \otimes f_i)[\sum_{i=1}^{k}\nu_i(\tilde{Z})(f_i \otimes f_i)](\sum_{i=1}^{k} f_i \otimes g_i) = \sum_{i=1}^{k}\nu_i(\tilde{Z})(g_i \otimes g_i).$$
$$B_1B_2B_3 = \sum_{i=k+1}^{\infty}\nu_i(\tilde{Z})(f_i \otimes f_i).$$

Since $A_1A_2A_3 \in \mathscr{B}(\mathscr{T}_k)$ and $B_1B_2B_3 \in \mathscr{B}(\mathscr{T}_k^{\perp})$, we have

$$P_{\mathscr{T}_k}\mathrm{fobi}(\tilde{X})P_{\mathscr{T}_k} = \sum_{i=1}^{k}\nu_i(\tilde{Z})(g_i \otimes g_i).$$

By construction, $\mathrm{var}(\langle \tilde{Z}, f_i \rangle_{\mathscr{H}}) = 1$, and hence, for any $i = 1, \ldots, k$,

$$\nu_i(\tilde{Z}) = \mathrm{kurt}(\langle Z, f_i \rangle_{\mathscr{H}}) + E\|\tilde{Z}\|_{\mathscr{H}}^2 - 1,$$

which are distinct as $\mathrm{kurt}(\langle Z, f_1 \rangle_{\mathscr{H}}), \ldots, \mathrm{kurt}(\langle Z, f_k \rangle_{\mathscr{H}})$ are distinct. Because

$$\sum_{i=1}^{k}\nu_i(\tilde{Z})(g_i \otimes g_i) = \sum_{i=1}^{k}\tau_i(h_i \otimes h_i) = P_{\mathscr{T}_k}\mathrm{fobi}(\tilde{X})P_{\mathscr{T}_k},$$

where $\nu_1(\tilde{Z}), \ldots, \nu_k(\tilde{Z})$ are distinct, by Lemma 5.6,

$$\sum_{i=1}^{k}(g_i \otimes f_i) = \sum_{i=1}^{k}(h_i \otimes f_i)R_0\Pi_0$$

for some $R_0 \in \mathscr{R}(\mathscr{T}_k)$ and $\Pi_0 \in \mathscr{P}(\mathscr{T}_k)$. Let $R = P_{\mathscr{T}_k}R_0P_{\mathscr{T}_k} + Q_{\mathscr{T}_k}$ and $\Pi = P_{\mathscr{T}_k}\Pi_0P_{\mathscr{T}_k} + Q_{\mathscr{T}_k}$. Then $R \in \mathscr{R}(\mathscr{H}|\mathscr{T}_k)$, $\Pi \in \mathscr{P}(\mathscr{H}|\mathscr{T}_k)$, and $U = VR\Pi$. Hence

$$V^*\tilde{X} = \Pi^*R^*U^*U\tilde{Z} = \Pi^*R^*\tilde{Z} = \Pi^{-1}R\tilde{Z}.$$

Because $\tilde{Z}$ has independent components, $R \in \mathscr{R}(\mathscr{H}|\mathscr{T}_k)$, $\Pi^{-1} \in \mathscr{P}(\mathscr{H}|\mathscr{T}_k)$, by Proposition 5.2, $V^*\tilde{X}$ has independent components. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

Proof of Theorem 6.1: Since A is a injective operator in $\mathscr{B}(\mathscr{H}|\mathscr{T}_k)$, there is an invertible operator $A_0 \in \mathscr{B}(\mathscr{T}_k)$ such that $A = P_{\mathscr{T}_k} A_0 P_{\mathscr{T}_k} + Q_{\mathscr{T}_k}$, $A_0^{-1} \in \mathscr{B}(\mathscr{T}_k)$. Let $Y = AX$. Then

$$\Sigma_Y^{-1/2(k)} Y = \Sigma_Y^{-1/2(k)} AX = \Sigma_Y^{-1/2(k)} A \Sigma_X^{1/2(k)} \Sigma_X^{-1/2(k)} X.$$

Let λ_i denote the eigenvalue of Σ_X associated with the eigenfunction f_i, and let

$$\Sigma_{X,0} = \sum_{i=1}^{k} \lambda_i (f_i \otimes f_i), \quad \Sigma_{X,1} = \sum_{i=k+1}^{\infty} \lambda_i (f_i \otimes f_i).$$

Then $\Sigma_X = \Sigma_{X,0} + \Sigma_{X,1}$, and

$$\Sigma_Y = A \Sigma_X A^* = (P_{\mathscr{T}_k} A_0 P_{\mathscr{T}_k} + Q_{\mathscr{T}_k})(\Sigma_{X,0} + \Sigma_{X,1})(P_{\mathscr{T}_k} A_0^* P_{\mathscr{T}_k} + Q_{\mathscr{T}_k})$$
$$= P_{\mathscr{T}_k} A_0 \Sigma_{X,0} A_0^* P_{\mathscr{T}_k} + \Sigma_{X,1}.$$

Hence

$$\Sigma_Y^{-1/2(k)} = P_{\mathscr{T}_k} (A_0 \Sigma_{X,0} A_0^*)^{-1/2} P_{\mathscr{T}_k} + Q_{\mathscr{T}_k}.$$

Substituting this and $Y = (P_{\mathscr{T}_k} A_0 P_{\mathscr{T}_k} + Q_{\mathscr{T}_k})X$ into $\Sigma_Y^{-1/2(k)} Y$, we obtain

$$\Sigma_Y^{-1/2(k)} Y = [P_{\mathscr{T}_k} (A_0 \Sigma_{X,0} A_0^*)^{-1/2} P_{\mathscr{T}_k} + Q_{\mathscr{T}_k}](P_{\mathscr{T}_k} A_0 P_{\mathscr{T}_k} + Q_{\mathscr{T}_k}) \Sigma_X^{1/2(k)} \Sigma_X^{-1/2(k)} X$$
$$= [P_{\mathscr{T}_k} (A_0 \Sigma_{X,0} A_0^*)^{-1/2} A_0 \Sigma_{X,0}^{1/2} P_{\mathscr{T}_k} + Q_{\mathscr{T}_k}] \Sigma_X^{-1/2(k)} X.$$

Following the proof of Lemma 5.4, it is easy to see that

$$P_{\mathscr{T}_k} (A_0 \Sigma_{X,0} A_0^*)^{-1/2} A_0 \Sigma_{X,0}^{1/2} P_{\mathscr{T}_k} + Q_{\mathscr{T}_k} \in \mathscr{U}(\mathscr{H}|\mathscr{T}_k).$$

Denote this unitary operator U_{YX}. By Lemma 5.4, there exist unitary operators U_{XZ}, U_{YZ} such that

$$\Sigma_X^{-1/2(k)} X = U_{XZ} \Sigma_Z^{-1/2(k)} Z, \quad \Sigma_Y^{-1/2(k)} Y = U_{YZ} \Sigma_Z^{-1/2(k)} Z.$$

Hence, by Lemma 5.5, with $U_{ZX} = U_{XY}^*$ and $U_{XY} = U_{YX}^*$,

$$\mathrm{fobi}(\Sigma_Y^{-1/2(k)} Y) = \mathrm{fobi}(U_{YX} U_{XZ} \Sigma_Z^{-1/2(k)} Z) = U_{YX} U_{XZ} \mathrm{fobi}(\Sigma_Z^{-1/2(k)} Z) U_{ZX} U_{XY},$$
$$\mathrm{fobi}(\Sigma_X^{-1/2(k)} X) = U_{XZ} \mathrm{fobi}(\Sigma_Z^{-1/2(k)} Z) U_{ZX}.$$

Now, consider the spectral decompositions of $\mathrm{fobi}(\Sigma_X^{-1/2(k)} X)$ and $\mathrm{fobi}(\Sigma_Y^{-1/2(k)} Y)$ projected on $\mathscr{T}_k$:

$$P_{\mathscr{T}_k} \mathrm{fobi}(\Sigma_X^{-1/2(k)} X) P_{\mathscr{T}_k} = V_{X,0} D_0 V_{X,0}^*, \quad P_{\mathscr{T}_k} \mathrm{fobi}(\Sigma_Y^{-1/2(k)} Y) P_{\mathscr{T}_k} = V_{Y,0} D_0 V_{Y,0}^*,$$

where $V_{X,0}, V_{Y,0} \in \mathscr{U}(\mathscr{T}_k)$ and $D_0 \in \mathscr{D}(\mathscr{T}_k)$. Because $\mathrm{kurt}(\langle \Gamma X, f_i \rangle_{\mathscr{H}})$, $i = 1, \ldots, k$, are distinct, we have, by Lemma 5.6, $V_{Y,0} = U_{YX,0} U_{XZ,0} R_0 \Pi_0$, and $V_{X,0} = U_{XZ,0} R_0 \Pi_0$, where $U_{YX,0}, U_{XZ,0}, R_0$ and Π_0 are

the $\mathscr{T}_k$-components of U_{YX}, U_{XZ}, R and Π defined according to the rule $C = P_{\mathscr{T}_k} C_0 P_{\mathscr{T}_k} + Q_{\mathscr{T}_k}$ for any $C \in \mathscr{B}(\mathscr{H}|\mathscr{T}_k)$. This last expression can be re-written as relations between operators in $\mathscr{B}(\mathscr{H}|\mathscr{T}_k)$ as follows:

$$V_Y = U_{YX} U_{XZ} R\Pi, \quad V_X = U_{XZ} R\Pi,$$

where V_Y and V_X are defined according to the same rule. Then

$$V_Y = U_{YX} V_X (R\Pi)^{-1}(R\Pi) = U_{YX} V_X.$$

Consequently,

$$V_Y^* \Sigma_Y^{-1/2(k)} Y = V_X^* U_{XY} \Sigma_Y^{-1/2(k)} A \Sigma_X^{1/2(k)} \Sigma_X^{-1/2(k)} X. \tag{12}$$

However, note that

$$U_{XY} \Sigma_Y^{-1/2(k)} A \Sigma_X^{1/2(k)} = [P_{\mathscr{T}_k} \Sigma_{X,0}^{1/2} A_0^* (A_0 \Sigma_{X,0} A_0^*)^{-1/2} P_{\mathscr{T}_k} + Q_{\mathscr{T}_k}]$$

$$[P_{\mathscr{T}_k}(A_0 \Sigma_{X,0} A_0^*)^{-1/2} P_{\mathscr{T}_k} + Q_{\mathscr{T}_k}](P_{\mathscr{T}_k} A_0 \Sigma_{X,0}^{1/2} P_{\mathscr{T}_k} + Q_{\mathscr{T}_k}) \tag{13}$$

$$= P_{\mathscr{T}_k} \Sigma_{X,0}^{1/2} A_0^* (A_0 \Sigma_{X,0} A_0^*)^{-1} A_0 \Sigma_{X,0}^{1/2} P_{\mathscr{T}_k} + Q_{\mathscr{T}_k} = I.$$

Substituting (13) into (12), we obtain the desired identity.   □

Proof of Lemma 7.1: *(a)*. It is easy to show that the inner product in $\mathscr{H}_m$ is $\langle f, g \rangle_{\mathscr{H}} = [f]^{\mathsf{T}}[g]$, and that the coordinate of tensor product in $\mathscr{H}_m$ is $[f \otimes g] = [f][g]^{\mathsf{T}}$. By the definition of $\Sigma_{\hat{X}}$ and linearity of the coordinate mapping,

$$[\Sigma_{\hat{X}}] = [E_n(\hat{X} \otimes \hat{X})] - [(E_n \hat{X}) \otimes (E_n \hat{X})]. \tag{14}$$

Hence

$$[E_n(\hat{X} \otimes \hat{X})] = E_n[\hat{X} \otimes \hat{X}] = E_n([\hat{X}][\hat{X}]^{\mathsf{T}})$$

$$[(E_n \hat{X}) \otimes (E_n \hat{X})] = [E_n \hat{X}][E_n \hat{X}]^{\mathsf{T}} = (E_n[\hat{X}])(E_n[\hat{X}])^{\mathsf{T}}.$$

Substitute these into the right hand side of (14) to prove the first equality in *1*. By the definition of partial power (7), as applied to $\Sigma_{\hat{X}}$, we have

$$\Sigma_{\hat{X}} = \sum_{i=1}^{k} \lambda_i^\alpha (f_i \otimes f_i) + \sum_{i=k+1}^{r} (f_i \otimes f_i).$$

Taking coordinates on both sides proves the second equality in *1*.

(b). If $\Sigma_{\hat{X}} f = \lambda f$, then $[\Sigma_{\hat{X}}][f] = \lambda[f]$, which, by *(a)*. above implies $\mathrm{var}_n([\hat{X}])[f] = \lambda[f]$. The same argument proves the second assertion.

(c). The ith demixed variable is

$$\langle \tilde{X}, g_i \rangle_{\mathscr{H}} = \sum_{\ell=1}^{k} \lambda_\ell^{-1/2} \langle \hat{X} - E_n \hat{X}, f_\ell \rangle_{\mathscr{H}} \langle f_\ell, g_i \rangle_{\mathscr{H}}$$

$$= \sum_{\ell=1}^{k} \lambda_\ell^{-1/2} ([\hat{X}] - E_n[\hat{X}])^{\mathsf{T}} [f_\ell][f_\ell]^{\mathsf{T}}[g_i],$$

which is the desired equality because, by *1.*, $[f_\ell]$ and $[g_i]$ are the eigenvectors of $\mathrm{var}_n([\hat{X}])$ and $\mathrm{fobi}_n([\tilde{X}])$, respectively. $\hspace{2cm} \square$

Acknowledgments

References

Baker, C. R. (1973) Joint measures and cross-covariance operators. *Transactions of the American Mathematical Society*, **186**, 273–289.

Bonaccorsi, S. and Priola, E. (2006) From brownian motion to stochastic differential equations. 10th Internet Seminar.

Bosq, D. (2000) *Linear Processes in Function Spaces.* Springer, New York.

Cardoso, J. F. (1989) Source separation using higher order moments. *Proceedings of IEEE International Conference on Accoustics, Speech and Signal Processing, Glasgow, UK*, 2109–2112.

Cardoso, J. F. and Souloumiac, A. (1993) Blind beamforming for non gausssian signals. *IEE Proceedings-F*, **140**, 362–370.

Chen, A. and Bickel, P. (2006) Efficient independent component analysis. *The Annals of Statistics*, **34**, 2825–2855.

Conway, J. (1990) *A Course in Functional Analysis.* Springer.

Cristescu, R., Ristaniemi, T., Joutsensalo, J. and Karhunen, J. (2000) Cdma delay estimation using fast ica algorithm. In *The 11th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, London*, 1117–1120.

Delaigle, A. and Hall, P. (2010) Defining probability density for a distribution of random functions. *The Annals of Statistics*, **38**, 1171–1193.

Ferraty, F. and Vieu, P. (2006) *Nonparametric Functional Data Analysis: Theory and Practice.* Springer.

Ferré, L. and Yao, A. F. (2003) Functional sliced inverse regression analysis. *Statistics: A Journal of Theoretical and Applied Statistics*, **37**, 475–488.

— (2005) Smoothed functional inverse regression. *Statistica Sinica*, **15**, 665–683.

Fukumizu, K. and Bach, F. R. (2007) Statistical Consistency of Kernel Canonical Correlation Analysis. *Journal of Machine Learning Research*, **8**, 361–383.

Gutch, H. W. and Theis, F. J. (2012) To infinity and beyond: on ica over hilbert spaces. In *Latent Variable Analysis and Signal Separation* (eds. F. Theis, A. Cichocki, A. Yeredor and M. Zibulevsky). Springer.

Hall, P. and Hosseini-Nasab, M. (2006) On properties of functional principal components analysis. *Journal of the Royal Statistical Society, Series B*, 109–126.

Hallin, M. and Mehta, C. (2015) R-estimation for asymmetric independent component analysis. *Journal of the American Statistical Association*, **110**, 218–232.

Hopkins, L. C. and Holland, G. J. (1997) Australian heavy-rain days and associated east coast cyclones: 1958-1992. *J. Climate*, **10**, 621–635.

Horváth, L. and Kokoszka, P. (2012) *Inference for Functional Data with Applications*. Springer.

Hsing, T. and Eubank, R. (2015) *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. John Wiley & Sons, Ltd.

Hsing, T. and Ren, H. (2009) An rkhs formulation of the inverse regression dimension-reduction problem. *The Annals of Statistics*, **37**, 726–755.

Hyvärinen, A. (1999) Sparse code shrinkage: denoising of nongaussian data by maximum likelihood estimation. *Neural Computation*, **11(7)**, 1483–1492.

Hyvärinen, A. and Oja, E. (2000) Independent component analysis: algorithms and applications. *Neural Networks*, **13**, 411–430.

Ilmonen, P. and Paindaveine, D. (2011) Semiparametrically efficient inference based on signed ranks in symmetric independent component models. *The Annals of Statistics*, **39**, 2448–2476.

Kiviluoto, K. and Oja, E. (1998) Independent component analysis for parallel financial times series. *Proceedings of the International Conference on Neural Information Processing*, **2**, 895–898.

Miettinen, J., Taskinen, S., Nordhausen, K. and Oja, H. (2015) Fourth moments and independent components. *Statistical Science*, **30**, 372–390.

Nordhausen, K., Oja, H. and Ollila, E. (2011) On the performance indices of ICA and Blind Source Separation. In *IEEE 12th International Workshop on Signal Processing Advances in Wireless Communications*, 486–490.

Oja, H., Sirkiä, S. and Eriksson, J. (2006) Scatter matrices and independent component analysis. *Australian Journal of Statistics*, **35**, 175–189.

Peña, D., Prieto, J. and Rendón, C. (2014) Independent components techniques based on kurtosis for functional data analysis. *Working Paper 14-10 Statistics and Econometrics Series (06)*, Universidad

Carlos III de Madrid.

Ramsay, J. and Silverman, B. (2005) *Functional Data Analysis, 2nd Ed.* Springer-Verlag.

Ristaniemi, T. and Joutsensalo, J. (1999) On the performance of blind source separation in cdma downlink. In *Proceedings of the International Workshop on Independent Component Analysis and Signal Separation*, 1483–1492.

Samworth, R. J. and Yuan, M. (2012) Independent component analysis via nonparametric maximum likelihood estimation. *The Annals of Statistics*, **40**, 2973–3002.

Tyler, D. E., Critchley, F., Dümbgen, L. and Oja, H. (2009) Invariant co-ordinate selection. *Journal of the Royal Statistical Society, Series B*, **71**, 549–592.

Vigário, R., Jousmäki, V., Hämäläinen, M., Hari, R. and Oja, E. (1998) Independent component analysis for identification of artifacts in magnetoencephalographic recordings. *Advances in Neural Information Processing Systems*, **10**, 229–235.

Yao, F., Müller, H. G. and Wang, J. L. (2005a) Functional data analysis for sparse longitudinal data. *Journal of American Statistical Association*, **100**, 577–590.

— (2005b) Functional linear regression analysis for longitudinal data. *The Annals of Statistics*, **33**, 2873–2903.