Biometrika (2012), **99**, 4, *pp.* 813–832

Printed in Great Britain

doi: 10.1093/biomet/ass037
Advance Access publication 26 August 2012

Dispersion operators and resistant second-order functional data analysis

By DAVID KRAUS and VICTOR M. PANARETOS

Institute of Mathematics, Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland

david.kraus@epfl.ch victor.panaretos@epfl.ch

Summary

Inferences related to the second-order properties of functional data, as expressed by covariance structure, can become unreliable when the data are non-Gaussian or contain unusual observations. In the functional setting, it is often difficult to identify atypical observations, as their distinguishing characteristics can be manifold but subtle. In this paper, we introduce the notion of a dispersion operator, investigate its use in probing the second-order structure of functional data, and develop a test for comparing the second-order characteristics of two functional samples that is resistant to atypical observations and departures from normality. The proposed test is a regularized M-test based on a spectrally truncated version of the Hilbert–Schmidt norm of a score operator defined via the dispersion operator. We derive the asymptotic distribution of the test statistic, investigate the behaviour of the test in a simulation study and illustrate the method on a structural biology dataset.

Some key words: Covariance operator; Karhunen–Loève expansion; M-estimation; Resistant test; Spectral truncation; Two-sample testing.

1. Introduction

The second-order structure of a random function is key to understanding the nature of the functional observations that it induces, as it is inextricably linked with the smoothness properties of the stochastic fluctuations of the function. Given a suitable random function in a separable Hilbert space, e.g., $L^2[0, 1]$, these second-order properties are encapsulated in the covariance operator. The link with the smoothness properties of the random function is then given by the Karhunen–Loève expansion (e.g., Adler, 1990), which provides an optimal Fourier representation of the random function, using a basis comprised by the eigenfunctions of this operator. Consequently, a significant part of functional data analysis has concentrated on estimating the covariance operator, and employing its spectral decomposition in order to probe the smoothness properties of the functional data; see Bosq (2000), Dauxois et al. (1982), Hall & Hosseini-Nasab (2006), Ramsay & Silverman (2005), Gervini (2006), Hall et al. (2006) and Yao & Lee (2006), to name but a few. A natural inference problem is that of comparing the covariance structures of two samples of functional data, in order to decide whether they share the same fluctuation properties. Aspects of this problem were considered in Benko et al. (2009), who employed a bootstrap procedure to compare subsets of eigenfunctions or eigenvalues of the two samples in a financial context. The more global problem of testing whether two samples share the same covariance operator was investigated in the Gaussian case by Panaretos et al. (2010), motivated by the study of mechanical properties of DNA, and subsequently by Boente et al. (2011) through

a simulation-based approach. In a slightly different setting, Gabrys & Kokoszka (2007) and Horváth et al. (2010) investigated second-order tests to detect the presence or change of serial correlation in functional data. The goal of this paper is to study the problem of second-order inference in a more general setting. We focus on situations where the data are not Gaussian, and indeed may be characterized by the presence of influential observations. That we do not use the word outlier is deliberate: in the functional case, observations can significantly impact the empirical covariance operator, though they may not be outlying. The infinite-dimensional nature of the data means that an observation can be atypical in many ways, the deviation from the mean being only one; observations close to the mean may contain unusual frequency components. Detection of such observations via exploratory techniques may be nontrivial (Sun & Genton, 2011).

Such influential observations might significantly influence the estimation of the covariance, and, even more profoundly, the quality of the estimators of its spectrum. For these reasons, robustified estimates of the spectrum have been proposed, based on the spectra of robust estimators of the covariance operator. Locantore et al. (1999) proposed the use of the spectrum of the so-called spherical covariance operator in a discretized setting (Boente & Fraiman, 1999). Gervini (2008) introduced the functional median and further studied the properties of the spherical covariance spectrum for functional data concentrated on an unknown finite-dimensional hyperplane. Bali et al. (2012) adapted the projection-pursuit method of Li & Chen (1985) in the functional case. The sensitivity of the empirical covariance operator and its spectrum to the presence of influential observations can have an impact on testing procedures for the covariance operator. This is already observed in the finite-dimensional case (Layard, 1974; Olson, 1974), where deviations from a Gaussian assumption, or the presence of influential observations, can completely ruin a testing procedure even in one dimension (Box, 1953; Hampel et al., 1986). Finite-dimensional robust or resistant tests for covariance matrices cannot be directly extended to the functional case, as they often depend on the assumption of an invertible empirical covariance, which will by default be violated in the functional case for all sample sizes (Tiku & Balakrishnan, 1985; O'Brien, 1992; Zhang et al., 1991; Anderson, 2006). Even if a pseudo-inverse operator is employed, one immediately runs into the problem of ill-posedness.

To cope with these issues, this paper introduces a class of operators that we term dispersion operators that are implicitly defined through a variational problem, motivated by M-estimators of location for the tensor product of the centred functional observations. It is then proposed that these operators be used as proxies for the covariance operator, when inferences on the second-order structure are to be drawn for non-Gaussian and potentially contaminated functional samples. The implicit definition of a dispersion operator gives rise to a score equation, as the dispersion operator is a zero of the Fréchet derivative of the variational problem with respect to the operator argument. This functional score equation is then used as a basis to construct a test for the second-order comparison of two functional samples. The test is based on the distance of the functional score equation under the null hypothesis from zero, measured by an appropriately renormalized Hilbert–Schmidt distance.

2. SECOND-ORDER INFERENCE BASED ON THE DISPERSION OPERATOR

2·1. *Covariance operators*

To describe the second-order properties of a random element X in a separable Hilbert space of functions $\mathcal{H}$, often taken to be $L^2[0, 1]$, with norm $\|\cdot\|$ and inner product $\langle\cdot, \cdot\rangle$, one typically considers the covariance operator of X, $\mathscr{C} : \mathcal{H} \to \mathcal{H}$, defined as

$$\mathscr{C}(f) = E\{\langle f, X - \mu\rangle(X - \mu)\};$$

here $\mu = E(X)$ represents the mean of the function X. For example, in the case $\mathcal{H} \equiv L^2[0, 1]$, with inner product $\langle f, g\rangle = \int_0^1 f(t)g(t)\,\mathrm{d}t$, the covariance operator is represented as an integral operator

$$\mathscr{C}(f) = \int_0^1 r(\cdot, s) f(s)\,\mathrm{d}s,$$

where $r(s, t) = E[\{X(s) - \mu(s)\}\{X(t) - \mu(t)\}]$ stands for the covariance kernel of the process X. For the purposes of this paper, it will be more fruitful to think of the covariance operator as an operator related to tensor products on $\mathcal{H}$, rather than through the sample path perspective based on the covariance kernel. In particular, we will think of the covariance operator as

$$\mathscr{C} = E\{(X - \mu) \otimes (X - \mu)\},$$

where $\otimes$ stands for the tensor product on $\mathcal{H}$: for $f, g \in \mathcal{H}$, $f \otimes g$ defines an operator on $\mathcal{H}$ through $(f \otimes g)(h) = \langle g, h\rangle f$, where $h \in \mathcal{H}$. In this setting, and provided that $E(\|X\|^2) < \infty$, the covariance operator $\mathscr{C}$ can itself be thought of as an element of a Hilbert space, the space $\mathrm{HS}(\mathcal{H}, \mathcal{H})$ of Hilbert–Schmidt operators acting on $\mathcal{H}$. This is the space of linear operators $\mathscr{R}$ on $\mathcal{H}$ such that

$$\|\mathscr{R}\|_{\mathrm{HS}} = \left(\sum_{k=1}^{\infty} \|\mathscr{R}e_k\|^2\right)^{1/2} < \infty,$$

where $\{e_k\}$ is any orthonormal basis of $\mathcal{H}$. Here, $\|\cdot\|_{\mathrm{HS}}$ defines a norm on $\mathrm{HS}(\mathcal{H}, \mathcal{H})$, corresponding to the inner product $\langle \mathscr{R}_1, \mathscr{R}_2\rangle_{\mathrm{HS}} = \sum_{k=1}^{\infty} \langle \mathscr{R}_1 e_k, \mathscr{R}_2 e_k\rangle$. In what follows, we will usually omit the subscript HS, as the nature of the norm or inner product employed, whether it is an operator or an element norm, will be clearly implied from the space where its argument belongs.

In this Hilbert–Schmidt setting, the covariance operator can be seen as the operator $\mathscr{C} \in \mathrm{HS}(\mathcal{H}, \mathcal{H})$ that solves the variational problem

$$\min_{\mathscr{R} \in \mathrm{HS}(\mathcal{H}, \mathcal{H})} E\{\|(X - \mu) \otimes (X - \mu) - \mathscr{R}\|^2\}.$$

The sample counterpart of the covariance operator, the empirical covariance operator,

$$\hat{\mathscr{C}}_n = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X}) \otimes (X_i - \bar{X}),$$

can be represented as the solution to the problem

$$\min_{\mathscr{R} \in \mathrm{HS}(\mathcal{H}, \mathcal{H})} \frac{1}{n} \sum_{i=1}^{n} \|(X_i - \bar{X}) \otimes (X_i - \bar{X}) - \mathscr{R}\|^2,$$

where $X_1, \ldots, X_n$ is a collection of independent and identically distributed copies of X, and $\bar{X} = n^{-1} \sum_{i=1}^{n} X_i$ stands for their empirical mean. This being essentially a least squares problem, both the empirical covariance operator and methods based on it will be sensitive to the presence of atypical observations in the dataset $X_1, \ldots, X_n$. In fact, it can also be seen that the empirical covariance operator admits a Gaussian maximum likelihood estimator interpretation, in a Cramér–Wold sense: if X is assumed Gaussian, then $\hat{\mathscr{C}}_n$ is the unique element of $\mathrm{HS}(\mathcal{H}, \mathcal{H})$

such that, for every $f \in \mathcal{H}$, $\langle f, \hat{\mathscr{C}}_n f \rangle$ is the unique maximum likelihood estimator of the variance of $\langle f, X \rangle$. The law of X is completely determined by the laws of the collection $\{\langle f, X \rangle : f \in \mathcal{H}\}$, and of course $\langle f, X \rangle$ is Gaussian with mean $\langle f, \mu \rangle$ and variance $\langle f, \mathscr{C} f \rangle$.

The basic strategy of this paper will be to obtain procedures pertaining to the second-order structure of X that are more resistant to departures from normality and to the presence of influential observations by replacing the squared norm in the variational problem defining the covariance by a less sensitive loss function. This gives rise to a new class of second-order characteristics, which we call dispersion operators.

2·2. *Dispersion operators*

Let P be a distribution on the separable Hilbert space $\mathcal{H}$ and let X be a random element with this distribution. The usual covariance is the integral of the operator

$$\mathscr{P}(x; \mu) = (x - \mu) \otimes (x - \mu), \quad x \in \mathcal{H},$$

with respect to P. This suggests that a dispersion operator could be defined as an M-estimator of the location of $\mathscr{P}(X; \mu)$. Let ρ be a nonnegative, differentiable, strictly increasing and convex function on $\mathbb{R}_0^+$ with $\rho(0) = 0$. We define the ρ-dispersion operator of the distribution P as

$$\mathscr{R}(\mathsf{P}) = \underset{\mathscr{R} \in \mathrm{HS}(\mathcal{H}, \mathcal{H})}{\arg\min} \; M(\mathsf{P}; \mathscr{R}, \mu), \tag{1}$$

where

$$\begin{aligned} M(\mathsf{P}; \mathscr{R}, \mu) &= E_{\mathsf{P}}[\rho\{\|\mathscr{P}(X; \mu) - \mathscr{R}\|\} - \rho\{\|\mathscr{P}(X; \mu)\|\}] \\ &= \int [\rho\{\|\mathscr{P}(x; \mu) - \mathscr{R}\|\} - \rho\{\|\mathscr{P}(x; \mu)\|\}] \, \mathrm{d}\mathsf{P}(x). \end{aligned} \tag{2}$$

In the definition of the dispersion operator, μ is chosen to be some suitable element of $\mathcal{H}$ with the interpretation of a location parameter. It is natural to use μ equal to the ρ-centre

$$\mu(\mathsf{P}) = \underset{\mu \in \mathcal{H}}{\arg\min} \, L(\mathsf{P}; \mu),$$

where

$$L(\mathsf{P}; \mu) = E_{\mathsf{P}}\{\rho(\|X - \mu\|) - \rho(\|X\|)\} = \int \{\rho(\|x - \mu\|) - \rho(\|x\|)\} \, \mathrm{d}\mathsf{P}(x).$$

Equivalently, one may define $\mu(\mathsf{P})$ and $\mathscr{R}(\mathsf{P})$ as solutions to score equations. The objective functionals $L(\mathsf{P}; \mu)$ and $M(\mathsf{P}; \mathscr{R}, \mu)$ are real-valued functionals defined on the Hilbert spaces $\mathcal{H}$ and $\mathrm{HS}(\mathcal{H}, \mathcal{H})$, respectively. The corresponding scores are their Fréchet derivatives, that is, linear functionals on the corresponding Hilbert space that can be uniquely identified with an element of that Hilbert space. Specifically, the centre $\mu(\mathsf{P})$ is the solution to the functional equation

$$G(\mathsf{P}; \mu) = 0,$$

where the element

$$G(\mathsf{P}; \mu) = \frac{\partial}{\partial \mu} L(\mathsf{P}; \mu) = E_{\mathsf{P}} \left\{ \frac{\rho'(\|X - \mu\|)}{\|X - \mu\|} (\mu - X) \right\} = \int \frac{\rho'(\|x - \mu\|)}{\|x - \mu\|} (\mu - x) \; \mathrm{d}\mathsf{P}(x)$$

of $\mathcal{H}$ determines the Fréchet derivative of L with respect to μ. The dispersion operator is defined as the solution to the operator equation

$$\mathcal{G}(\mathsf{P}; \mathcal{R}, \mu) = \mathcal{O}, \tag{3}$$

where $\mathcal{O}$ is the zero operator on $\mathcal{H}$ and the operator

$$\begin{aligned}\mathcal{G}(\mathsf{P}; \mathcal{R}, \mu) &= \frac{\partial}{\partial \mathcal{R}} M(\mathsf{P}; \mathcal{R}, \mu) = E_{\mathsf{P}} \left[\frac{\rho'\{\|\mathcal{P}(X; \mu) - \mathcal{R}\|\}}{\|\mathcal{P}(X; \mu) - \mathcal{R}\|} \{\mathcal{R} - \mathcal{P}(X; \mu)\} \right] \\ &= \int \frac{\rho'\{\|\mathcal{P}(x; \mu) - \mathcal{R}\|\}}{\|\mathcal{P}(x; \mu) - \mathcal{R}\|} \{\mathcal{R} - \mathcal{P}(x; \mu)\} \, \mathrm{d}\mathsf{P}(x)\end{aligned}$$

determines the Fréchet derivative of M with respect to $\mathcal{R}$.

The empirical dispersion operator based on the sample $X_1, \ldots, X_n$ is the dispersion operator of the empirical distribution $\hat{\mathsf{P}}$ of the sample, that is, $\mathcal{R}(\hat{\mathsf{P}})$. The empirical dispersion operator can be in general computed around any element $\mu \in \mathcal{H}$; in practice, one naturally uses the empirical centre $\mu(\hat{\mathsf{P}})$, i.e., the centre of the empirical distribution.

PROPOSITION 1. *Let* P *be a distribution on the separable Hilbert space* $\mathcal{H}$ *that is not concentrated on a line in* $\mathcal{H}$ *or on four points of* $\mathcal{H}$. *Assume that* ρ *is nonnegative, strictly increasing on* $[0, \infty)$ *and convex. Then, the objective function* $M(\mathsf{P}; \mathcal{R}, \mu)$ *as a functional of* $\mathcal{R}$ *is strictly convex for any* $\mu \in \mathcal{H}$ *and thus the* ρ*-dispersion operator around* μ *exists and is unique.*

Proposition 1 holds without any moment assumptions because the subtraction of $\rho\{\|\mathcal{P}(X; \mu)\|\}$ and $\rho(\|X\|)$ in the definition of $M(\mathsf{P}; \mathcal{R}, \mu)$ and $L(\mathsf{P}; \mu)$, respectively, guarantees the existence and finiteness of the objective functions. Under fairly weak further assumptions, we may also deduce that the empirical dispersion operator is well defined and consistent.

COROLLARY 1. *Let* $X_1, \ldots, X_n$ *be independent random elements with law* P *that has no discrete component and is such that the probability that* $X_1, \ldots, X_n$ *be collinear is zero* $(n \geqslant 3)$. *Then, for* $n \geqslant 5$, *the empirical* ρ*-dispersion operator corresponding to* $X_1, \ldots, X_n$ *exists and is almost surely unique. Moreover, if* $\hat{\mu}$ *is consistent for a location parameter* μ, *then the empirical dispersion operator around* $\hat{\mu}$ *is itself consistent for the dispersion operator around* μ.

We remark, for example, that the empirical functional median, i.e., the empirical centre corresponding to $\rho(u) = u$, was proven to be consistent for its theoretical counterpart in Gervini (2008). In fact, in the setting of Corollary 1, this result can be extended to location parameters corresponding to strictly increasing convex ρ-functions.

It is seen from (1) or (3) that the ρ-dispersion operator is self-adjoint. Moreover, from the spectral decomposition found in Proposition 2, it will follow that the ρ-dispersion operator is positive semidefinite. Although many results derived in this paper are valid for a wide class of functions ρ, the choice $\rho(u) = u^q$ for some $q > 0$ is especially attractive as the resulting centre is scale invariant and the dispersion is scale equivariant. For general ρ, it would be more appropriate to use a suitably studentized version of the objective functions; to this end, one can insert a preliminary estimator of the trace into the objective function.

We now provide explicit formulae for two main choices of the ρ-function.

When choosing $\rho(u) = u^2$, the score determining the ρ-dispersion operator equals $\mathscr{G}(\mathsf{P}; \mathscr{R}, \mu) = E_{\mathsf{P}}[2\{\mathscr{R} - \mathscr{P}(X; \mu)\}]$. Thus, $\mathscr{R}(\mathsf{P})$ can be found explicitly as $\mathscr{R}(\mathsf{P}) = E_{\mathsf{P}}\{\mathscr{P}(X; \mu)\}$. As the score for the ρ-centre is $G(\mathsf{P}; \mu) = E_{\mathsf{P}}\{2(\mu - X)\}$, the solution is $\mu(\mathsf{P}) = E_{\mathsf{P}}(X)$. Hence, the dispersion operator is the usual covariance operator.

The choice $\rho(u) = u$ is expected to place less emphasis on influential observations and result in more resistant procedures. The corresponding score operators for the dispersion and centre are

$$\mathscr{G}(\mathsf{P}; \mathscr{R}, \mu) = E_{\mathsf{P}}\left\{\frac{\mathscr{R} - \mathscr{P}(X; \mu)}{\|\mathscr{R} - \mathscr{P}(X; \mu)\|}\right\}, \quad G(\mathsf{P}; \mu) = E_{\mathsf{P}}\left(\frac{\mu - X}{\|\mu - X\|}\right).$$

The parameter $\mu(\mathsf{P})$ has been studied by a number of authors under different names in the multivariate as well as functional settings. In the multivariate context Chaudhuri (1996) calls $\mu(\mathsf{P})$ the geometric median; other authors (Serfling, 2004; Sirkiä et al., 2009) use the name spatial median and some authors (Huber & Ronchetti, 2009; Fritz et al., 2012) use the term L^1-centre or L^1-median. In the functional setting, $\mu(\mathsf{P})$ was studied by Locantore et al. (1999) and by Gervini (2008), who calls it the functional or spatial median. We use the term spatial median for $\mu(\mathsf{P})$ and, similarly, we call $\mathscr{R}(\mathsf{P})$ the spatial dispersion operator. To clarify the terminology, we recall that

$$\mathscr{S}(\mathsf{P}) = E_{\mathsf{P}}\left\{\frac{(X - \mu) \otimes (X - \mu)}{\|X - \mu\|^2}\right\}$$

is called the spherical covariance operator (Locantore et al., 1999). Unlike the parameters under the L^2-type loss function, the spatial median and spatial dispersion are not available explicitly. Their empirical counterparts $\hat{\mu} = \mu(\hat{\mathsf{P}})$ and $\hat{\mathscr{R}} = \mathscr{R}(\hat{\mathsf{P}})$ can, however, be obtained numerically, employing a Newton–Raphson algorithm, as explained in the Appendix.

The score function $\rho'(u) = qu^{q-1}$ corresponding to $\rho(u) = u^q$ is unbounded unless $q = 1$. Therefore, the estimator of the spatial dispersion operator, $q = 1$, is resistant, whereas other choices are nonresistant due to the effect of outliers, $q > 1$, or inliers, $q < 1$.

Although the dispersion operator is in general different from the covariance operator unless $\rho(u) = u^2$, it carries useful information on second-order properties of the distribution. There is an interesting link between the spectra of the dispersion and covariance operator. Let X admit the Karhunen–Loève expansion $X = \mu + \sum_{k=1}^{\infty} \lambda_k^{1/2} \beta_k \varphi_k$, where $\beta_1, \beta_2, \ldots$ are zero-mean unit-variance uncorrelated random variables, $\{\lambda_k : k \geqslant 1\}$ are the nonincreasing nonnegative eigenvalues and $\{\varphi_k : k \geqslant 1\}$ are the complete orthonormal eigenfunctions of the covariance operator $\mathscr{C}(\mathsf{P}) = E_{\mathsf{P}}\{(X - \mu) \otimes (X - \mu)\} = \sum_{k=1}^{\infty} \lambda_k \varphi_k \otimes \varphi_k$. We now investigate the eigendecomposition of the theoretical ρ-dispersion operator $\mathscr{R}(\mathsf{P})$ defined via M-estimation as the solution to (3). The main result is as follows.

PROPOSITION 2. *Assume that the Fourier coefficient sequence $\{\beta_k\}_{k=1}^{\infty}$ has a joint distribution that is invariant under the change of the sign of any component. Then, the dispersion operator $\mathscr{R}(\mathsf{P})$ has the same eigenfunctions as the covariance operator $\mathscr{C}(\mathsf{P})$, i.e., there exists a nonnegative sequence $\{\delta_k\}_{k=1}^{\infty}$ such that $\mathscr{R}(\mathsf{P}) = \sum_{k=1}^{\infty} \delta_k \varphi_k \otimes \varphi_k$. Furthermore, the eigenvalues $\delta_1, \delta_2, \ldots$ satisfy the conditions*

$$\delta_k = \lambda_k \frac{E\left(\frac{\rho'[\{\sum_i (\delta_i - \lambda_i \beta_i^2)^2 + \sum_{i \neq l} \lambda_i \lambda_l \beta_i^2 \beta_l^2\}^{1/2}]}{\{\sum_i (\delta_i - \lambda_i \beta_i^2)^2 + \sum_{i \neq l} \lambda_i \lambda_l \beta_i^2 \beta_l^2\}^{1/2}} \beta_k^2\right)}{E\left(\frac{\rho'[\{\sum_i (\delta_i - \lambda_i \beta_i^2)^2 + \sum_{i \neq l} \lambda_i \lambda_l \beta_i^2 \beta_l^2\}^{1/2}]}{\{\sum_i (\delta_i - \lambda_i \beta_i^2)^2 + \sum_{i \neq l} \lambda_i \lambda_l \beta_i^2 \beta_l^2\}^{1/2}}\right)} \quad (k = 1, 2, \ldots).$$

A similar result relating the covariance operator and the spherical covariance operator $\mathscr{S}(\mathsf{P})$ was obtained by Gervini (2008, Theorem 3) who showed that, under the assumption of exchangeability of the coefficient sequence, both operators have the same eigenfunctions in the same order; see also Marden (1999) and Boente & Fraiman (1999). Our proposition shows that the ρ-dispersion operator also has the same set of eigenfunctions. We conjecture that, potentially under further assumptions, the order of the eigenfunctions is also the same; computational experiments back this conjecture. Gervini (2008) assumed that the Karhunen–Loève expansion has only finitely many terms, i.e., that the distribution is concentrated on a finite-dimensional subspace, whereas our results hold even for processes with infinite series expansions. On the other hand, Gervini (2008) needed no moment assumptions, whereas we need to assume finite second moments: without moment assumptions the convergence of an infinite Karhunen–Loève series is not guaranteed, while a finite sum is always well defined regardless of the properties of the random summands.

2·3. *The two-sample test*

Having defined the notion of a dispersion operator, we now construct a two-sample second-order test based upon it. Let $X_1, \ldots, X_{n_1}$ and $Y_1, \ldots, Y_{n_2}$ be two independent random samples from distributions $\mathsf{P}_1, \mathsf{P}_2$ on $\mathcal{H}$, whose ρ-centres are $\mu(\mathsf{P}_1)$, $\mu(\mathsf{P}_2)$ and ρ-dispersion operators are $\mathscr{R}(\mathsf{P}_1), \mathscr{R}(\mathsf{P}_2)$. The goal is to test the null hypothesis $H_0\colon \mathscr{R}(\mathsf{P}_1) = \mathscr{R}(\mathsf{P}_2)$ against the general alternative $H_1\colon \mathscr{R}(\mathsf{P}_1) \neq \mathscr{R}(\mathsf{P}_2)$. Note that $\mu(\mathsf{P}_1)$, $\mu(\mathsf{P}_2)$ can be equal or different, as neither H_0 nor H_1 specifies their relation. We propose to employ the general idea of score tests, that is, to base the test on the estimating score for the general model, without assuming H_0, evaluated at the null estimate of the parameter.

As the centres $\mu(\mathsf{P}_1)$, $\mu(\mathsf{P}_2)$ are not restricted under the null hypothesis, they can be estimated separately by minimizing $L(\hat{\mathsf{P}}_1; \mu_1)$, $L(\hat{\mathsf{P}}_2; \mu_2)$, i.e., by solving $G(\hat{\mathsf{P}}_1; \mu_1) = 0$, $G(\hat{\mathsf{P}}_2; \mu_2) = 0$, respectively. Denote $\mu(\hat{\mathsf{P}}_j)$ by $\hat{\mu}_j$ $(j = 1, 2)$. On the other hand, the null estimator of the dispersion is based on both samples. As we now have two samples, we need to extend our notation to cover situations with two distributions, empirical or theoretical, mixed at proportions a and $1 - a$ for $a \in (0, 1)$. We denote

$$M(\mathsf{P}_1, \mathsf{P}_2, a; \mathscr{R}_1, \mathscr{R}_2, \mu_1, \mu_2) = aM(\mathsf{P}_1; \mathscr{R}_1, \mu_1) + (1 - a)M(\mathsf{P}_2; \mathscr{R}_2, \mu_2).$$

The common null value $\mathscr{R}$ of the dispersion operator is estimated by $\hat{\mathscr{R}}$, which minimizes $M(\hat{\mathsf{P}}_1, \hat{\mathsf{P}}_2, a_n; \mathscr{R}, \mathscr{R}, \hat{\mu}_1, \hat{\mu}_2)$ where $a_n = n_1/n$ with $n = n_1 + n_2$. Equivalently, $\hat{\mathscr{R}}$ solves $\mathscr{G}(\hat{\mathsf{P}}_1, \hat{\mathsf{P}}_2, a_n; \mathscr{R}, \hat{\mu}_1, \hat{\mu}_2) = \mathscr{O}$, the null estimating equation, where $\mathscr{G}(\mathsf{P}_1, \mathsf{P}_2, a; \mathscr{R}, \mu_1, \mu_2) = a\mathscr{G}(\mathsf{P}_1; \mathscr{R}, \mu_1) + (1 - a)\mathscr{G}(\mathsf{P}_2; \mathscr{R}, \mu_2)$.

Using the reparameterization $\mathscr{R} = (\mathscr{R}_1 + \mathscr{R}_2)/2$, $\mathscr{T} = (\mathscr{R}_1 - \mathscr{R}_2)/2$, we have $\mathscr{R}_1 = \mathscr{R} + \mathscr{T}$, $\mathscr{R}_2 = \mathscr{R} - \mathscr{T}$ and we need to test $H_0\colon \mathscr{T} = \mathscr{O}$ against $H_1\colon \mathscr{T} \neq \mathscr{O}$. For the test, we need the score in the general model

$$\frac{\partial}{\partial(\mathscr{R}, \mathscr{T})^{\mathrm{T}}} M(\hat{\mathsf{P}}_1, \hat{\mathsf{P}}_2, a_n; \mathscr{R} + \mathscr{T}, \mathscr{R} - \mathscr{T}, \hat{\mu}_1, \hat{\mu}_2) = \begin{pmatrix} \mathscr{G}(\hat{\mathsf{P}}_1, \hat{\mathsf{P}}_2, a_n; \mathscr{R}, \hat{\mu}_1, \hat{\mu}_2) \\ \mathscr{B}(\hat{\mathsf{P}}_1, \hat{\mathsf{P}}_2, a_n; \mathscr{R}, \hat{\mu}_1, \hat{\mu}_2) \end{pmatrix}$$

where $\mathscr{B}(\mathsf{P}_1, \mathsf{P}_2, a; \mathscr{R}, \mu_1, \mu_2) = a\mathscr{G}(\mathsf{P}_1; \mathscr{R}, \mu_1) - (1 - a)\mathscr{G}(\mathsf{P}_2; \mathscr{R}, \mu_2)$. The score test is based on this general score at the null estimator. When evaluated at $(\mathscr{R}, \mathscr{T}) = (\hat{\mathscr{R}}, \mathscr{O})$, the score is zero in the first component. Thus, the test can be based on the second component $\mathscr{B}(\hat{\mathsf{P}}_1, \hat{\mathsf{P}}_2, a_n; \hat{\mathscr{R}}, \hat{\mu}_1, \hat{\mu}_2)$.

When the null hypothesis holds, the score operator $\mathscr{B}(\hat{\mathsf{P}}_1, \hat{\mathsf{P}}_2, a_n; \hat{\mathscr{R}}, \hat{\mu}_1, \hat{\mu}_2)$ is expected to be close to the zero operator, otherwise it should be far from the zero operator. To perform the test, we need to measure the distance of $\mathscr{B}(\hat{\mathsf{P}}_1, \hat{\mathsf{P}}_2, a_n; \hat{\mathscr{R}}, \hat{\mu}_1, \hat{\mu}_2)$ from the zero operator and assess the significance of the resulting test statistic.

One way to measure the distance of the score operator from zero is to use its Hilbert–Schmidt norm. A drawback of this approach is that the resulting statistic does not have a tractable asymptotic distribution. The score operator turns out to be asymptotically Gaussian, but its Hilbert–Schmidt norm is not asymptotically distribution-free. In the context of comparison of covariance operators, Boente et al. (2011) use a simulation procedure to approximate the distribution of the statistic.

Another idea is to mimic the standard procedure from settings where the parameter of interest is Euclidean. In such settings, the difference of the score vector from zero is measured with the help of a quadratic form involving the score vector and the inverse of its covariance matrix. The quadratic statistic is usually asymptotically chi-square distributed and the null hypothesis is then rejected when the value of the statistic is significantly large. In the functional context, the score $\mathscr{B}(\hat{\mathsf{P}}_1, \hat{\mathsf{P}}_2, a_n; \hat{\mathscr{R}}, \hat{\mu}_1, \hat{\mu}_2)$ is infinite dimensional. Due to the noninvertibility of its covariance operator, one cannot construct a quadratic statistic. We overcome this problem by regularizing the score operator using spectral truncation.

The test object $\mathscr{B}(\hat{\mathsf{P}}_1, \hat{\mathsf{P}}_2, a_n; \hat{\mathscr{R}}, \hat{\mu}_1, \hat{\mu}_2)$ is an element of the space of operators $\mathrm{HS}(\mathcal{H}, \mathcal{H})$. Recall that $\mathrm{HS}(\mathcal{H}, \mathcal{H})$ is a Hilbert space with inner product defined as

$$\langle \mathscr{A}_1, \mathscr{A}_2 \rangle = \sum_{k=1}^{\infty} \langle \mathscr{A}_1 e_k, \mathscr{A}_2 e_k \rangle = \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \langle e_j, \mathscr{A}_1 e_k \rangle \langle e_j, \mathscr{A}_2 e_k \rangle, \quad \mathscr{A}_1, \mathscr{A}_2 \in \mathrm{HS}(\mathcal{H}, \mathcal{H}),$$

where $\{e_k : k = 1, 2, \dots\}$ is an arbitrary complete orthonormal basis of $\mathcal{H}$. For any complete orthonormal basis $\{\mathscr{E}_k : k = 1, 2, \dots\}$ of $\mathrm{HS}(\mathcal{H}, \mathcal{H})$, an operator $\mathscr{A} \in \mathrm{HS}(\mathcal{H}, \mathcal{H})$ and the square of its Hilbert–Schmidt norm can be written as

$$\mathscr{A} = \sum_{k=1}^{\infty} \langle \mathscr{A}, \mathscr{E}_k \rangle \mathscr{E}_k, \quad \|\mathscr{A}\|^2 = \sum_{k=1}^{\infty} \langle \mathscr{A}, \mathscr{E}_k \rangle^2.$$

Instead of this infinite series, one can use a truncated version. If $\mathscr{U} \subset \mathrm{HS}(\mathcal{H}, \mathcal{H})$ is a suitably chosen finite-dimensional linear subspace with an orthonormal basis $\{\mathscr{U}_1, \dots, \mathscr{U}_L\}$, then instead of $\|\mathscr{B}(\hat{\mathsf{P}}_1, \hat{\mathsf{P}}_2, a_n; \hat{\mathscr{R}}, \hat{\mu}_1, \hat{\mu}_2)\|^2$ one can use

$$\begin{aligned} \|\pi_{\mathscr{U}} \mathscr{B}(\hat{\mathsf{P}}_1, \hat{\mathsf{P}}_2, a_n; \hat{\mathscr{R}}, \hat{\mu}_1, \hat{\mu}_2)\|^2 &= \|\mathscr{B}(\hat{\mathsf{P}}_1, \hat{\mathsf{P}}_2, a_n; \hat{\mathscr{R}}, \hat{\mu}_1, \hat{\mu}_2)\pi_{\mathscr{U}}\|^2 \\ &= \sum_{l=1}^{L} \langle \mathscr{B}(\hat{\mathsf{P}}_1, \hat{\mathsf{P}}_2, a_n; \hat{\mathscr{R}}, \hat{\mu}_1, \hat{\mu}_2), \mathscr{U}_l \rangle^2, \end{aligned}$$

where $\pi_{\mathscr{U}}$ is the projection onto the subspace $\mathscr{U}$. That is, the test can be based on a score vector with components

$$S_l = \langle \mathscr{B}(\hat{\mathsf{P}}_1, \hat{\mathsf{P}}_2, a_n; \hat{\mathscr{R}}, \hat{\mu}_1, \hat{\mu}_2), \mathscr{U}_l \rangle \quad (l = 1, \dots, L). \tag{4}$$

One particular way of choosing the basis elements $\mathscr{U}_l$ is to derive them from a basis of the Hilbert space $\mathcal{H}$. If U is a K-dimensional linear subspace of $\mathcal{H}$ with an orthonormal basis $\{u_1, \dots, u_K\}$,

then one may use the $L = K(K+1)/2$ orthonormal operators of the form

$$\mathscr{U}_{jk} = \begin{cases} u_j \otimes u_j & (j = k), \\ (u_j \otimes u_k + u_k \otimes u_j)/2^{1/2} & (j < k). \end{cases} \tag{5}$$

There is yet another way of motivating the above truncation. Instead of measuring the difference of $\mathscr{B}(\hat{\mathsf{P}}_1, \hat{\mathsf{P}}_2, a_n; \hat{\mathscr{R}}, \hat{\mu}_1, \hat{\mu}_2)$ from zero on the entire Hilbert space $\mathcal{H}$, we can measure how it differs from the zero operator when attention is restricted to the linear subspace U. More precisely, instead of $\mathscr{B}(\hat{\mathsf{P}}_1, \hat{\mathsf{P}}_2, a_n; \hat{\mathscr{R}}, \hat{\mu}_1, \hat{\mu}_2)$, we use the operator $\pi_U \mathscr{B}(\hat{\mathsf{P}}_1, \hat{\mathsf{P}}_2, a_n; \hat{\mathscr{R}}, \hat{\mu}_1, \hat{\mu}_2)\pi_U$, where π_U is the projection operator on U. Its squared Hilbert–Schmidt norm

$$\|\pi_U \mathscr{B}(\hat{\mathsf{P}}_1, \hat{\mathsf{P}}_2, a_n; \hat{\mathscr{R}}, \hat{\mu}_1, \hat{\mu}_2)\pi_U\|^2 = \sum_{j=1}^{K}\sum_{k=1}^{K} \langle u_j, \mathscr{B}(\hat{\mathsf{P}}_1, \hat{\mathsf{P}}_2, a_n; \hat{\mathscr{R}}, \hat{\mu}_1, \hat{\mu}_2)u_k\rangle^2$$

is a truncated version of

$$\|\mathscr{B}(\hat{\mathsf{P}}_1, \hat{\mathsf{P}}_2, a_n; \hat{\mathscr{R}}, \hat{\mu}_1, \hat{\mu}_2)\|^2 = \sum_{j=1}^{\infty}\sum_{k=1}^{\infty} \langle e_j, \mathscr{B}(\hat{\mathsf{P}}_1, \hat{\mathsf{P}}_2, a_n; \hat{\mathscr{R}}, \hat{\mu}_1, \hat{\mu}_2)e_k\rangle^2,$$

where $\{e_j : j = 1, 2, \dots\}$ is any complete orthonormal basis of $\mathcal{H}$. The resulting scores

$$S_{jk} = \langle u_j, \mathscr{B}(\hat{\mathsf{P}}_1, \hat{\mathsf{P}}_2, a_n; \hat{\mathscr{R}}, \hat{\mu}_1, \hat{\mu}_2)u_k\rangle \quad (1 \leqslant j \leqslant k \leqslant K)$$

are equivalent to (4) with $\mathscr{U}_l$ of the form (5).

It is natural to use the basis operators of the form (5) with $u_1, \dots, u_K$ being the first K eigenfunctions of the dispersion operator $\mathscr{R}$ because, in light of Mercer's theorem, they carry the main portion of information about the dispersion operator. In practice, the eigenfunctions of $\mathscr{R}$ are not known, so one uses the eigenfunctions of the pooled sample estimator $\hat{\mathscr{R}}$. The number of components K can be selected as the minimal number for the cumulative proportion of dispersion explained by the subspace to exceed a certain threshold, e.g., 80% of the trace of the corresponding pooled sample dispersion operator. The proportion of dispersion, corresponding to the eigenvalues of the dispersion operator, is in general not equivalent to the proportion of variability, corresponding to the eigenvalues of the covariance operator.

To construct the test statistic, instead of simply summing squares of the terms S_l of the form (4), one combines them in a quadratic form reflecting their covariance structure.

The formal test will be based on the asymptotic distribution of the test statistic. Let n_1, n_2 be such that $n_1 \to \infty$, $n_2 \to \infty$ and $a_n = n_1/n \to a \in (0, 1)$. Assume that $\|G(\mathsf{P}_j; \mu)\|^2$, $\|\mathscr{G}(\mathsf{P}_j; \mathscr{R}, \mu)\|^2$ $(j = 1, 2)$ are finite. Let the function $\rho\colon \mathbb{R}_0^+ \to \mathbb{R}_0^+$ be twice differentiable, strictly increasing, and convex with $\rho(0) = 0$. Assume that the laws P_1, P_2 satisfy the conditions of Corollary 1 and the expectations $E_{\mathsf{P}_j}\{\rho'(\|X - \mu\|)^2\}$, $E_{\mathsf{P}_j}[\rho'\{\|\mathscr{P}(X; \mu) - \mathscr{R}\|\}^2]$, $E_{\mathsf{P}_j}\{\rho''(\|X - \mu\|)\}$, $E_{\mathsf{P}_j}[\rho''\{\|\mathscr{P}(X; \mu) - \mathscr{R}\|\}]$ and

$$E_{\mathsf{P}_j}\left\{\frac{\rho'(\|X - \mu\|)}{\|X - \mu\|}\right\}, \quad E_{\mathsf{P}_j}\left[\frac{\rho'\{\|\mathscr{P}(X; \mu) - \mathscr{R}\|\}}{\|\mathscr{P}(X; \mu) - \mathscr{R}\|}\right] \quad (j = 1, 2)$$

are finite. Assume that the derivatives $\mathscr{D}(\mathsf{P}_j; \mu)$, $\mathfrak{D}(\mathsf{P}_j; \mathscr{R}, \mu)$, $\mathbb{D}(\mathsf{P}_j; \mathscr{R}, \mu)$ given in (A1)–(A3) in the Appendix exist for $j = 1, 2$.

Let S be a score vector of length L of the form (4) for some linearly independent operators $\mathscr{U}_l = \mathscr{U}_l^{(n)}$. Let the operators $\mathscr{U}_l$ be either nonrandom, independent of n, or convergent in probability to some nonrandom limits, up to a possible sign ambiguity in the sense that there

exist some operators $\mathscr{U}_l^\infty$ such that $|\langle \mathscr{U}_l^{(n)}, \mathscr{U}_l^\infty \rangle|$ converges to 1. In this set-up, we have the following theorem.

THEOREM 1. *Under the null hypothesis $H_0 : \mathscr{R}(\mathsf{P}_1) = \mathscr{R}(\mathsf{P}_2)$, the score $n^{1/2}\mathscr{B}(\hat{\mathsf{P}}_1, \hat{\mathsf{P}}_2, a_n; \hat{\mathscr{R}}, \hat{\mu}_1, \hat{\mu}_2)$ converges weakly to a mean zero Gaussian random operator with covariance operator, which can be consistently estimated by $\mathfrak{W}(\hat{\mathsf{P}}_1, \hat{\mathsf{P}}_2, a_n; \hat{\mathscr{R}}, \hat{\mu}_1, \hat{\mu}_2)$ given in (A5) in the Appendix. The asymptotic distribution of the score vector $n^{1/2}S$ is L-variate zero-mean Gaussian with a covariance matrix that is consistently estimated by a matrix W with entries $W_{j,l} = \langle \mathscr{U}_j, \mathfrak{W}(\hat{\mathsf{P}}_1, \hat{\mathsf{P}}_2, a_n; \hat{\mathscr{R}}, \hat{\mu}_1, \hat{\mu}_2)\mathscr{U}_l \rangle$ $(j, l = 1, \ldots, L)$. The test statistic $T = nS^{\mathrm{T}}W^{-1}S$ asymptotically follows a χ^2 distribution with L degrees of freedom.*

We now deal with the two main cases, spatial and L^2-type, explicitly. In the spatial case, $\rho(u) = u$, we test the null hypothesis that the spatial dispersion operators are equal in both samples. The score operator takes the form

$$\mathscr{B}(\hat{\mathsf{P}}_1, \hat{\mathsf{P}}_2, a_n; \hat{\mathscr{R}}, \hat{\mu}_1, \hat{\mu}_2) = \frac{1}{n}\sum_{i=1}^{n_1} \frac{\hat{\mathscr{R}} - \mathscr{P}(X_i; \hat{\mu}_1)}{\|\hat{\mathscr{R}} - \mathscr{P}(X_i; \hat{\mu}_1)\|} - \frac{1}{n}\sum_{i=1}^{n_2} \frac{\hat{\mathscr{R}} - \mathscr{P}(Y_i; \hat{\mu}_2)}{\|\hat{\mathscr{R}} - \mathscr{P}(Y_i; \hat{\mu}_2)\|}.$$

The Fréchet derivatives $\mathscr{D}(\mathsf{P}; \mu)$, $\mathfrak{D}(\mathsf{P}; \mathscr{R}, \mu)$ involved in the covariance operator of the score are

$$\mathscr{D}(\mathsf{P}; \mu) = E_{\mathsf{P}}\left[\frac{1}{\|X - \mu\|}\left\{\mathscr{I} - \frac{(X-\mu)\otimes(X-\mu)}{\|X-\mu\|^2}\right\}\right],$$

$$\mathfrak{D}(\mathsf{P}; \mathscr{R}, \mu) = E_{\mathsf{P}}\left(\frac{1}{\|\mathscr{P}(X;\mu) - \mathscr{R}\|}\left[\mathfrak{I} - \frac{\{\mathscr{P}(X;\mu) - \mathscr{R}\}\otimes\{\mathscr{P}(X;\mu) - \mathscr{R}\}}{\|\mathscr{P}(X;\mu) - \mathscr{R}\|^2}\right]\right),$$

and the derivative $\mathbb{D}(\mathsf{P}; \mathscr{R}, \mu)$ evaluated at $f \in \mathcal{H}$ is

$$\mathbb{D}(\mathsf{P}; \mathscr{R}, \mu)f = E_{\mathsf{P}}\left[\frac{-\mathbb{Q}(X;\mu)f}{\|\mathscr{P}(X;\mu) - \mathscr{R}\|} + \frac{\langle \mathscr{P}(X;\mu) - \mathscr{R}, \mathbb{Q}(X;\mu)f\rangle}{\|\mathscr{P}(X;\mu) - \mathscr{R}\|^3}\{\mathscr{P}(X;\mu) - \mathscr{R}\}\right].$$

When the L^2 approach, $\rho(u) = u^2$, is employed, the hypothesis to be tested states that the covariance operators in both samples are equal. The null estimator of $\mathscr{R}$ takes the form $\hat{\mathscr{R}} = a_n\hat{\mathscr{R}}_1 + (1 - a_n)\hat{\mathscr{R}}_2$, that is, the pooled covariance estimator. The test score operator equals

$$\mathscr{B}(\hat{\mathsf{P}}_1, \hat{\mathsf{P}}_2, a_n; \hat{\mathscr{R}}, \hat{\mu}_1, \hat{\mu}_2) = a_n 2(\hat{\mathscr{R}} - \hat{\mathscr{R}}_1) - (1 - a_n)2(\hat{\mathscr{R}} - \hat{\mathscr{R}}_2) = 4a_n(1 - a_n)(\hat{\mathscr{R}}_2 - \hat{\mathscr{R}}_1),$$

which is a multiple of the difference of the empirical covariance operators. So, the test is equivalent to a Wald-type test proposed by Panaretos et al. (2010). This is different from the spatial test for which the score does not simplify to the difference of the spatial dispersions, so the score test differs from the Wald test. To compute the covariance operator of the test score, we first notice that $\mathbb{D}(\mathsf{P}; \mathscr{R}, \mu) = -2\,E_{\mathsf{P}}\{\mathbb{Q}(X; \mu)\}$ equals zero at $\mu = \mu(\mathsf{P}) = E_{\mathsf{P}}(X)$; see (A4) in the Appendix. Consequently, the fact that the centres of the two distributions must be estimated does not affect the asymptotic distribution, as could be expected. Also, $\mathfrak{D}(\mathsf{P}; \mathscr{R}, \mu) = 2\mathfrak{I}$. Hence, after straightforward calculations, the estimator of the covariance operator of the test operator is

$$\mathfrak{W}(\hat{\mathsf{P}}_1, \hat{\mathsf{P}}_2, a_n; \hat{\mathscr{R}}, \hat{\mu}_1, \hat{\mu}_2) = 4a_n(1 - a_n)\{(1 - a_n)\mathfrak{J}(\hat{\mathsf{P}}_1; \hat{\mathscr{R}}, \hat{\mu}_1) + a_n\mathfrak{J}(\hat{\mathsf{P}}_2; \hat{\mathscr{R}}, \hat{\mu}_2)\}$$

$$= 16a_n(1-a_n) \times \left[(1-a_n)\frac{1}{n_1}\sum_{i=1}^{n_1}\{\mathscr{P}(X_i;\hat{\mu}_1)-\hat{\mathscr{R}}_1\}\otimes\{\mathscr{P}(X_i;\hat{\mu}_1)-\hat{\mathscr{R}}_1\} + a_n\frac{1}{n_2}\sum_{i=1}^{n_2}\{\mathscr{P}(Y_i;\hat{\mu}_2)-\hat{\mathscr{R}}_2\}\otimes\{\mathscr{P}(Y_i;\hat{\mu}_2)-\hat{\mathscr{R}}_2\}\right].$$

In Panaretos et al. (2010), the limiting covariance of the L^2 score for the Wald-type test was investigated in the special case of Gaussian data and a simpler formula was found.

3. A simulation study

In order to investigate the performance of the testing procedure introduced in § 2·3, we generate random samples of size n_1, n_2 of curves of the form

$$X(t) = \mu_1(t) + \sum_{k=1}^{10}\lambda_{1k}^{1/2}a_{1k}2^{1/2}\sin\{2\pi k(t+\gamma_{1k})\} + \sum_{k=1}^{10}\nu_{1k}^{1/2}b_{1k}2^{1/2}\cos\{2\pi k(t+\delta_{1k})\},$$

$$Y(t) = \mu_2(t) + \sum_{k=1}^{10}\lambda_{2k}^{1/2}a_{2k}2^{1/2}\sin\{2\pi k(t+\gamma_{2k})\} + \sum_{k=1}^{10}\nu_{2k}^{1/2}b_{2k}2^{1/2}\cos\{2\pi k(t+\delta_{2k})\},$$

where the coefficients a_{jk}, b_{jk} are mutually independent random variables with zero-mean and unit variance. Three symmetric coefficient distributions are considered: normal, uniform and t_5, all scaled to have unit variance. As the test procedures are invariant with respect to the location shift of one or both samples, we set $\mu_1(t) = \mu_2(t) = 0$. Unless stated otherwise, we set $\gamma_{jk} = \delta_{jk} = 0$ in all situations. We perform the nonresistant L^2 test and the proposed spatial dispersion test at the nominal level $\alpha = 0{\cdot}05$. The sample sizes are $n_1 = n_2 = 50$. The basis of the subspace for dimension reduction consists of several leading eigenfunctions of the pooled sample estimator of the dispersion operator; that is, the pooled sample empirical covariance for the L^2 test and the pooled sample empirical spatial dispersion for the spatial test. The number of components K included in the basis is selected as the minimal number needed to explain at least 80% of the dispersion.

We first study the behaviour of the test procedures under the null hypothesis. We set $\lambda_{1k} = \lambda_{2k} = k^{-3}$ and $\nu_{1k} = \nu_{2k} = (1/3)^k$.

We begin with uncontaminated samples to verify that the tests maintain the prescribed nominal level. The first row of Table 1 shows that, in general, the asymptotic distribution approximates the distribution of both test statistics reasonably well. The asymptotic approximation for the L^2 method is slightly less accurate and tends to be liberal for distributions with light tails, i.e., normal and uniform.

Next we simulate datasets contaminated by atypical observations. Mean contamination, i.e., observations whose mean is different from the mean of the central distribution, usually impacts the level more seriously than pure covariance contamination, i.e., observations with the same mean but different covariance structure. Thus, we focus on mean contamination, i.e., outliers, in the study of the resistance of the level. In one or both samples, m_j out of n_j observations were replaced by observations that have mean function μ_j^{cont} instead of μ_j and the same covariance structure as the original distribution. We consider various distances of the contamination distribution from the central distribution and various contamination proportions, as indicated in Tables 1

Table 1. *Empirical rejection probabilities* (%) *at the nominal level* $\alpha = 5\%$ *under the null hypothesis. Samples of size* $n_1 = n_2 = 50$ *are contaminated by* m_1, m_2 *observations with mean functions* μ_1^{cont}, μ_2^{cont}, *respectively, and the same covariance structure as the central distribution. Estimates are based on* 2000 *simulation runs*

				Normal		t_5		Uniform	
m_1	$\mu_1^{\text{cont}}(t)$	m_2	$\mu_2^{\text{cont}}(t)$	L^2	Spatial	L^2	Spatial	L^2	Spatial
0		0		7·1	5·0	5·4	5·3	7·8	4·6
5	1	5	$1{\cdot}5 - 3\sin(\pi t)$	9·2	6·6	8·2	6·4	10·0	4·6
5	1·5	5	$1{\cdot}5 - 3\sin(\pi t)$	14·4	6·4	14·6	6·8	14·6	4·6
5	2·5	5	$1{\cdot}5 - 3\sin(\pi t)$	22·9	6·0	23·0	7·2	23·0	5·1
5	1	5	$2 - 4\sin(\pi t)$	11·2	7·2	10·3	7·7	11·7	5·2
5	1·5	5	$2 - 4\sin(\pi t)$	18·8	7·2	19·8	7·8	20·0	5·4
5	2·5	5	$2 - 4\sin(\pi t)$	30·4	7·2	32·4	8·2	30·8	6·4
5	1	5	$2{\cdot}5 - 5\sin(\pi t)$	14·1	8·2	14·0	8·0	15·0	6·4
5	1·5	5	$2{\cdot}5 - 5\sin(\pi t)$	25·9	8·2	25·4	8·4	27·8	6·5
5	2·5	5	$2{\cdot}5 - 5\sin(\pi t)$	41·8	8·3	46·4	9·0	42·4	7·2
5	1	0		7·4	6·0	6·4	5·4	8·6	5·0
5	1·5	0		12·6	5·9	11·2	5·7	13·4	4·6
5	2·5	0		19·0	6·1	17·8	6·0	17·8	4·7
0		5	$1{\cdot}5 - 3\sin(\pi t)$	9·0	6·0	7·2	6·6	9·8	5·6
0		5	$2 - 4\sin(\pi t)$	12·3	6·8	10·8	7·7	13·0	6·6
0		5	$2{\cdot}5 - 5\sin(\pi t)$	16·4	7·6	14·4	8·7	16·8	7·6

Table 2. *Empirical rejection probabilities* (%) *at the nominal level* $\alpha = 5\%$ *under the null hypothesis. Samples of size* $n_1 = n_2 = 50$ *are contaminated by* m_1, m_2 *observations with mean functions* $\mu_1^{\text{cont}}(t) = 1{\cdot}5$, $\mu_2^{\text{cont}}(t) = 2 - 4\sin(\pi t)$, *respectively, and the same covariance structure as the central distribution. Estimates are based on* 2000 *simulation runs*

	$m_1 = m, m_2 = 0$		$m_1 = 0, m_2 = m$		$m_1 = m_2 = m$	
m	L^2	Spatial	L^2	Spatial	L^2	Spatial
0	7·1	5·0	7·1	5·0	7·1	5·0
1	7·0	5·4	6·7	5·1	7·2	5·6
2	6·8	5·0	7·5	5·4	7·8	5·6
3	6·9	5·3	8·7	5·6	8·4	6·2
4	8·4	6·2	10·7	6·2	11·2	6·4
5	12·6	5·9	12·3	6·8	18·8	7·2
6	24·8	6·5	14·8	7·5	39·2	8·1
7	57·8	7·4	17·2	8·6	71·6	10·2
8	89·2	7·9	20·8	9·2	93·0	17·6
9	99·0	11·9	24·7	11·4	99·0	28·2
10	99·8	18·4	28·2	13·6	100·0	42·7

and 2. We consider only atypical observations that are not very far from the central distribution. These are the most insidious because they are often hidden in the main, apparently typical part of the dataset, do not stand out and thus are not easily identified visually, yet they often have a devastating impact on the behaviour of the nonresistant test. To illustrate this, we plot in Fig. 1 typical simulated samples with $m_1 = 5$, $\mu_1^{\text{cont}}(t) = 1{\cdot}5$ and $m_2 = 5$, $\mu_2^{\text{cont}}(t) = 2 - 4\sin(\pi t)$. When looking at the plots, one would be unable to identify atypical observations, if they were not highlighted. Visually, many of them do not seem to be very different from most curves, whereas some curves from the central distribution could be considered unusual.

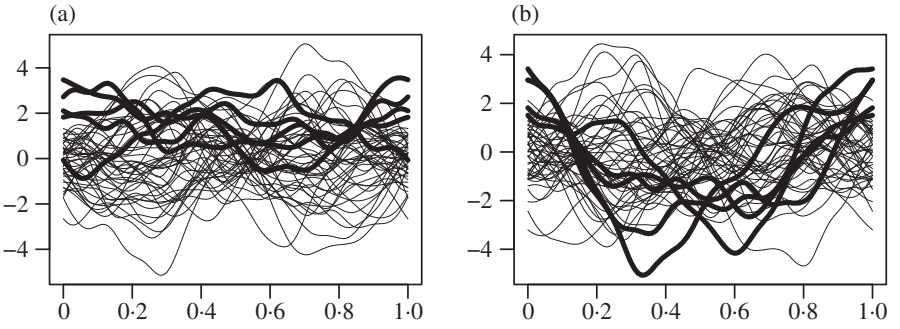

Fig. 1. Simulated contaminated samples. (a) Samples with $m_1 = 5$ atypical observations with $\mu_1^{\text{cont}}(t) = 1{\cdot}5$; (b) Samples with $m_2 = 5$ atypical observations with $\mu_2^{\text{cont}}(t) = 2 - 4\sin(\pi t)$. Atypical obsevations plotted in bold.

Table 1 shows that the proposed spatial test is much more resistant to contamination than the L^2-type test. For instance, notice that for $m_1 = m_2 = 5$, i.e., 10% contamination of both samples, the level of the spatial test in all situations considered is only slightly inflated, while the actual level of the L^2-type test exceeds 40%. Similarly, if one of the samples contains five atypical observations and the other is not contaminated, i.e., 10% contamination of one sample with 5% contamination overall, the spatial test rejects with probability close to the nominal level, while the level of the L^2-type test is as high as 19%. As the magnitude of atypical observations increases, the true level of the L^2 test, unlike that of the spatial one, increases dramatically. Comparing the behaviour of the tests across the various coefficient distributions, we observe no important differences. The higher resistance of the spatial method is also documented in Table 2, where the dependence of the level on the amount of contamination is studied for Gaussian data. The spatial procedure can tolerate much more contamination than can the L^2-type method.

Now we focus on the behaviour of the tests under alternatives. We consider five alternative scenarios. Under all of them, the parameters of the distribution of the first sample are $\lambda_{1k} = k^{-3}$ and $\nu_{1k} = (2/5)^k$. The parameters of the second sample are as follows. Under scenario I, we have $\lambda_{2k} = 1{\cdot}6\lambda_{1k}$ and $\nu_{2k} = 1{\cdot}6\nu_{1k}$ $(k = 1, \ldots, 10)$, so the samples differ only in scale, their covariance structure is otherwise the same. Under scenario II, we use $\lambda_{21} = 1{\cdot}5$, $\nu_{21} = 0{\cdot}8$ and $\lambda_{2k} = \lambda_{1k}$ and $\nu_{2k} = \nu_{1k}$ $(k = 2, \ldots, 10)$, so the covariance operators differ in the two leading eigenvalues, which however correspond to the same eigenfunctions. Scenario III has $\lambda_{2k} = \lambda_{1k}$ $(k = 1, \ldots, 10)$ and $\nu_{21} = 0{\cdot}2$, $\nu_{22} = 0{\cdot}35$ and $\nu_{2k} = \nu_{1k}$ $(k = 3, \ldots, 10)$; here the difference is on the second and third eigenvalues whose corresponding eigenfunctions are the same but in the opposite order. Under scenario IV, we set $\lambda_{22} = \lambda_{13}$, $\lambda_{23} = \lambda_{12}$, $\nu_{22} = \nu_{13}$, $\nu_{23} = \nu_{12}$ and $\lambda_{2k} = \lambda_{1k}$, $\nu_{2k} = \nu_{1k}$ $(k \notin \{2, 3\})$, so the difference occurs further down in the spectrum; eigenfunctions with indices 3, 4, 5, 6 are permuted, the leading two eigen-elements do not differ. Under scenario V, we use $\lambda_{2k} = \lambda_{1k}$, $\nu_{2k} = \nu_{1k}$ and $\gamma_{2k} = \delta_{2k} = 0{\cdot}15$ $(k = 1, \ldots, 10)$; in this case, the whole eigenbases are different but the eigenvalues remain the same in both samples.

First, we compare the power of the proposed spatial method with the L^2-type method for samples without contamination. Table 3 shows that in most cases the power of the spatial test is lower than the power of the L^2-type test for distributions with light tails. The lower efficiency of the spatial method is the price we pay for its increased resistance. Both methods have comparable power in the heavy tailed case under most scenarios. Under scenario IV the spatial method outperforms the L^2-type method. This is due to the automatic selection of K: for instance in the normal case, for the L^2-type test K equals 3 in 91 percent of cases while, for the spatial test, K equals 4 in 96 percent of cases; as the covariance operators differ on the third to sixth eigen-elements, K equal to 4 captures more of the difference.

Table 3. *Empirical rejection probabilities* (%) *at the nominal level* $\alpha = 5\%$ *under various alternative scenarios for samples of size* $n_1 = n_2 = 50$ *without contamination. Estimates are based on* 1000 *simulation runs*

	Normal		t_5		Uniform	
	L^2	Spatial	L^2	Spatial	L^2	Spatial
I	55	40	28	30	93	62
II	53	29	28	22	92	48
III	74	53	36	38	99	85
IV	38	61	24	53	49	73
V	76	58	53	51	96	72

Table 4. *Empirical rejection probabilities* (%) *of the spatial test at the nominal level* $\alpha = 5\%$ *under various alternative scenarios for samples of size* $n_1 = n_2 = 50$ *contaminated by* m_1, m_2 *atypical observations. Estimates are based on* 1000 *simulation runs*

Contamination configuration	m_1	m_2	I	II	III	IV	V
	0	0	40	29	53	61	58
A	5	5	12	16	57	64	59
	5	0	34	25	54	62	58
	0	5	15	16	56	63	61
B	5	5	29	22	36	39	55
	5	0	33	28	46	74	55
	0	5	40	28	49	34	57
C	5	5	24	18	34	39	52
	5	0	32	22	43	50	62
	0	5	31	24	43	49	48

Next, we investigate the impact of contamination on the power of the spatial test; we do not study the L^2-type test as we have seen before that its level is unreliable for contaminated data. The goal is to study if and how contamination can decrease the power. Similarly to the null scenario, here we also observed that mean contamination usually increases the rejection probability. Therefore, it is more interesting to contaminate data with curves with atypical covariance structure. We experimented with many configurations of atypical observations such that it is difficult to identify them visually and found that often even covariance contamination increases the rejection probability. Nevertheless, we were able to find some configurations for which we observed a decrease of the power in some situations. The central distributions follow the same scenarios I–V as before with normally distributed coefficients. Contamination configurations are as follows. Under configuration A, the contamination distribution has $\lambda_{1k}^{\text{cont}} = 1{\cdot}4\lambda_{1k}$, $\nu_{1k}^{\text{cont}} = 1{\cdot}4\nu_{1k}$, $\lambda_{2k}^{\text{cont}} = 0{\cdot}25\lambda_{2k}$ and $\nu_{2k}^{\text{cont}} = 0{\cdot}25\nu_{2k}$ $(k = 1, \ldots, 10)$, other parameters of the contamination distribution are the same as for the central distribution. Under configuration B, we set $\lambda_{1k}^{\text{cont}} = 0{\cdot}3\lambda_{1k}$ and $\lambda_{2k}^{\text{cont}} = 0{\cdot}3\lambda_{2k}$ $(k = 1, \ldots, 10)$, $\nu_{1k}^{\text{cont}} = 0{\cdot}3\nu_{1k}$ and $\nu_{2k}^{\text{cont}} = 0{\cdot}3\nu_{2k}$ $(k = 3, \ldots, 10)$, and $\nu_{11}^{\text{cont}} = \nu_{21}^{\text{cont}} = 1$ and $\nu_{12}^{\text{cont}} = \nu_{22}^{\text{cont}} = 0{\cdot}9$, while other parameters remain unchanged. Under configuration C, atypical observations in the first sample follow the central distribution of the second sample and atypical observations in the second sample follow the central distribution of the first sample.

The simulation results are presented in Table 4. We report only configurations with some detrimental effect on the power, while many configurations not reported here do not have such an effect. Under configuration A, we can see a decrease of the rejection probability for scenarios I and II. Configuration A was specifically designed to decrease the power under scenario I:

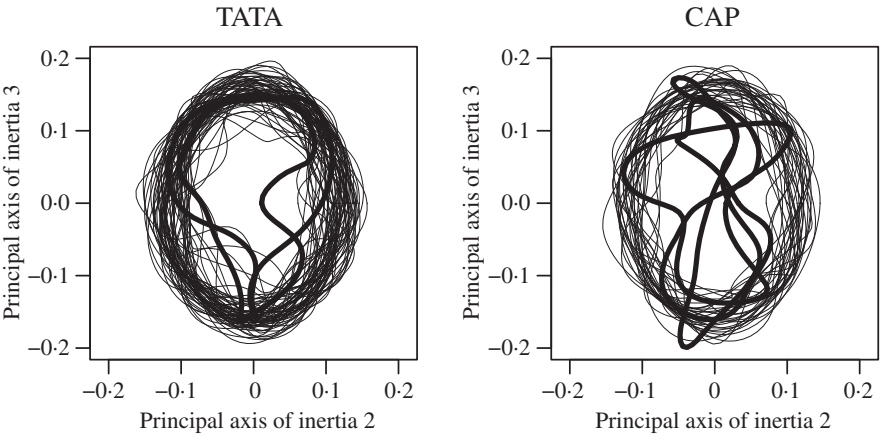

Fig. 2. Projection of DNA minicircle curves on the first principal plane spanned by the second and third principal axis of inertia. Atypical observations plotted in bold.

atypical observations deviate from the central distribution against the direction of the alternative; specifically, both the central and contamination distributions have proportional covariance operators but in the opposite direction. A similar phenomenon is seen for scenario II, where the directions of the alternative and of the contamination distribution are in a similar relationship. On the other hand, we observe no important effect of contamination of type A under scenarios III–V because in these cases atypical observations do not go against the alternative. Under configuration B, the power decreases mainly for scenarios III and IV. Configuration B downweights components other than the first and second cosine component, where it puts higher weight equal for both samples. As these are components carrying an important part of the difference between the covariances, one expects some decrease of the rejection probability, especially under scenarios III and IV. Under configuration C, the two samples are partly mixed, i.e., one sample contaminates the other sample and vice versa. This blurs the difference and somewhat decreases the power under some of the scenarios.

4. An illustration: DNA minicircle data

We illustrate the proposed methods on a dataset consisting of reconstructed three-dimensional electron microscope images of loops called minicircles obtained from short strands of DNA (Amzallag et al., 2006). The dataset contains 99 DNA minicircles of two types, TATA, 65 observations, and CAP, 34 observations, with identical base-pair sequences, except for a short subsequence where they differ. The main question is whether this difference affects the flexibility properties of the DNA minicircles. One way to formalize the flexibility properties is through the fluctuation pattern around the mean minicircle shape. This naturally leads one to consider two-sample second-order functional comparisons. DNA minicircles are closed curves in $\mathbb{R}^3$. In the original dataset, each curve was randomly rotated and shifted in $\mathbb{R}^3$ and had no starting point and no orientation. In Panaretos et al. (2010), an alignment procedure based on the moment of inertia tensor was used as a means of alignment of the curves in a common coordinate system. Figure 2 shows projections of aligned curves on the plane spanned by the two principal axes of inertia.

Using inverse weights induced by Gervini's (2008) spatial median, Panaretos et al. (2010) identified five unusual curves, possible outliers, and removed them from the analysis of the covariance structure. These atypical curves, plotted in thick lines in Fig. 2, are visibly different from the remaining curves. Panaretos et al. (2010) analysed the data without the atypical observations using a test comparing empirical covariance operators under the assumption that

the curves are Gaussian. Under this assumption, they observed significant differences at the 5% level. These differences were highly significant with a numerically zero p-value, when the comparison was restricted to the eigenvalues of the covariance operators; the corresponding empirical eigenfunctions suggested that the eigenfunction structure of the two operators was very similar.

Taking advantage of the results in the present paper, we may run an L^2-type test without assuming normality. When doing so, with the atypical observations still removed, the p-value of the L^2-type score test of the equality of covariance operators equals 0·023 with the dimension of the subspace on which the test operator is projected equal to $K = 6$, suggesting persistence of the effect, independently of a Gaussian assumption. Instead of removing apparently atypical observations manually, one might also wish to run an analysis on the complete dataset. However, the performance of L^2-type procedures was seen to be highly unstable in the presence of atypical observations, such as the ones in the present dataset, see Tables 1 and 2. By contrast, the spatial dispersion test was seen to maintain a level close to nominal in our simulations, especially in outlier scenarios similar to the one in the minicircle data. There may be further influential observations lurking in the sample. For this reason, we applied the score test based on the spatial dispersion operator, using the full minicircle dataset. In contrast to the other procedures, this yielded the p-value 0·353 indicative of a lack of significant differences in the spatial dispersions. The value of K was selected as the minimal number of components needed to explain 80% of the trace of the underlying null dispersion estimator. No further outliers were detected by the resistant test. The discordance between the L^2 and spatial tests is probably due to the reduced efficiency of the resistant procedure when the two samples share common eigenfunctions, as seems to be the case in the minicircle dataset; recall that the dispersion operator shares the same eigenfunctions with the covariance operator, possibly up to order. It was seen in our simulations that, in general, though the level of the spatial test was conserved, in the presence of influential observations its power was appreciably reduced when differences were only in the eigenvalues, i.e., under scenarios I and II in Table 4, as compared to scenarios where differences exist between the eigenfunctions, too, i.e., scenarios III–V in Table 4. Moreover the present framework does not immediately yield a special version of the test that would concentrate only on the eigenvalue structure; the complete structure of the operator is taken into account.

ACKNOWLEDGEMENT

We thank the editor, associate editor, and two anonymous referees for their extensive, constructive, and in-depth comments and suggestions. This research was supported in part by the European Research Council.

SUPPLEMENTARY MATERIAL

Supplementary material available at *Biometrika* online includes proofs of Proposition 1, Corollary 1, Proposition 2, Theorem 1 and a technical lemma needed in the proof of Theorem 1.

APPENDIX

Computation

Assume that the observations $X_i \in \mathcal{H}$ are represented as linear combinations of some known fixed basis elements ψ_j, that is, $X_i = \sum_{j=1}^{p} \xi_{ij} \psi_j$. This representation is usually obtained by a least squares procedure, possibly with smoothing, from some form of discrete original observations of X_i. The exact form of the original data depends on the particular application. For instance, when $\mathcal{H}$ is a functional, L^2, space indexed by one-dimensional time, the original data usually consist of observations $X_i(t_k)$ $(k = 1, \ldots, m)$ for a grid

of points $t_1 < \cdots < t_m$. Now suppose that the original data are observed discretely but exactly, i.e., without noise; later we explain how to handle noisy discrete observations.

The methods proposed in this paper have the advantage that all required quantities and operations can be expressed in terms of basis coefficients; thus, from the computational point of view the task is multivariate. To estimate the centre, it is enough to find the vector of coefficients m_j in its basis expansion $\mu = \sum_{j=1}^{p} m_j \psi_j$. Similarly, for the dispersion operator, we need to find the matrix of coefficients $R_{jj'}$ in the expansion

$$\mathscr{R} = \sum_{j=1}^{p} \sum_{j'=1}^{p} R_{jj'} \psi_j \otimes \psi_{j'}.$$

For simplicity, we first assume that the basis $\psi_1, \ldots, \psi_p$ is orthonormal. Then, the norm in the objective function for μ is simply the norm of the coefficient vector, i.e., $\|X_i - \mu\|^2 = \|\xi_i - m\|^2 = \sum_{j=1}^{p} (\xi_{ij} - m_j)^2$, and the score operator $G(\hat{\mathsf{P}}; \mu)$ is equivalent to the p-vector

$$\frac{1}{n} \sum_{i=1}^{n} \frac{\rho'(\|\xi_i - m\|)}{\|\xi_i - m\|} (m - \xi_i).$$

The Hilbert–Schmidt norm in the objective function for $\mathscr{R}$ is the Frobenius norm of the coefficient matrix, i.e.,

$$\|\mathscr{P}(X_i; \mu) - \mathscr{R}\|^2 = \|(\xi_i - m)(\xi_i - m)^{\mathrm{T}} - R\|^2 = \sum_{j=1}^{p} \sum_{j'=1}^{p} \{(\xi_{ij} - m_j)(\xi_{ij'} - m_{j'}) - R_{jj'}\}^2,$$

and the score operator $\mathscr{G}(\hat{\mathsf{P}}; \mathscr{R}, \mu)$ is equivalent to the $p \times p$ matrix

$$\frac{1}{n} \sum_{i=1}^{n} \frac{\rho'\{\|(\xi_i - m)(\xi_i - m)^{\mathrm{T}} - R\|\}}{\|(\xi_i - m)(\xi_i - m)^{\mathrm{T}} - R\|} \{R - (\xi_i - m)(\xi_i - m)^{\mathrm{T}}\}.$$

For the two-sample test, the operator $\mathscr{B}(\hat{\mathsf{P}}_1, \hat{\mathsf{P}}_2, a_n; \hat{\mathscr{R}}, \hat{\mu}_1, \hat{\mu}_2)$ and the basis elements $\mathscr{U}_l$ for dimension reduction are equivalent to matrices, and the score components S_l are computed as their inner products. Similarly, all quantities involved in the covariance matrix of the score vector are computed in a multivariate setting. When the basis $\psi_1, \ldots, \psi_p$ is not orthonormal, one simply multiplies each coefficient vector ξ_i by the matrix $A^{1/2}$ where A has entries $a_{jj'} = \langle \psi_j, \psi_{j'} \rangle$, and performs all computations, i.e., estimation of the centre and dispersion, eigen-decomposition and the two-sample test, with these transformed multivariate inputs. This corresponds to switching from the original basis to the orthonormal basis $A^{-1/2}(\psi_1, \ldots, \psi_p)^{\mathrm{T}}$. If needed, the centre and the eigenfunctions can then be obtained in the original basis by multiplying their coefficient vectors by $A^{-1/2}$ and in the dispersion by multiplying its coefficient matrix by $A^{-1/2}$ from both sides. We refer to Ramsay & Silverman (2005, § 8.4.2) for a detailed explanation of a similar problem of computing functional principal components from coefficients with respect to a general non-orthonormal basis.

To estimate the centre and dispersion one solves the corresponding multivariate optimization problem. If $\rho(u) = u^2$, the solutions are the sample mean and covariance matrix of the coefficient vectors; otherwise an iterative procedure is used. We use the Broyden–Fletcher–Goldfarb–Shanno quasi-Newton method implemented in the R package (R Development Core Team, 2012) in the function optim, initialized by the componentwise median of ξ_i for the centre and the componentwise median of $(\xi_i - m)(\xi_i - m)^{\mathrm{T}}$ for the dispersion. This numerical procedure was reliable and reasonably fast in our experiments. This is in agreement with a detailed study of the numerical performance of various algorithms for the spatial median presented by Fritz et al. (2012).

In functional settings one can directly use the functional values on a grid of points instead of computing with basis coefficients. The basis approach is slightly more general than the discretization approach because it can be used for any separable Hilbert space, not only a functional space, and in the functional

case it does not require a common grid for all functions. Standard software for functional data analysis, such as the fda package in R, uses basis representations of data.

In many applications, the original functional values on a grid of points are observed with noise. In such situations, some degree of smoothing is necessary for the reconstruction of the underlying functional data. Ramsay & Silverman (2005, Chapter 5) describe how roughness penalties can be used to compute the basis coefficients of the functions. After this preliminary step, our methods can be applied to the reconstructed curve, i.e., their basis coefficients, as described above.

In the case of the spatial median, Gervini (2008, pp. 589–590) proposes an alternative method to deal with noise in discretely observed functions. Rather than on denoising and reconstructing the curves, his procedure is based on removing the bias, which is due to the errors, in the norm in the objective function with the help of a consistent estimate of the variance of the errors. He uses this idea in connection with numerical integration on a grid, but it can be adapted to the basis approach as well. However, this method is less practical for second-order problems, as one would also need to estimate higher order moments of the errors and use convoluted formulae to remove the bias from the norm in the objective functional.

Technical material

We now derive several key expressions pertaining to the assumptions, statement and discussion of Theorem 1. We use the script font, e.g., $\mathscr{D}$, $\mathscr{J}$, $\mathscr{I}$, for linear operators on $\mathcal{H}$, i.e., linear mappings $\mathcal{H} \to \mathcal{H}$, the fraktur font, e.g., $\mathfrak{D}$, $\mathfrak{J}$, $\mathfrak{I}$, $\mathfrak{H}$, $\mathfrak{W}$, for linear operators on Hilbert–Schmidt operators on $\mathcal{H}$, i.e., linear mappings $\mathrm{HS}(\mathcal{H}, \mathcal{H}) \to \mathrm{HS}(\mathcal{H}, \mathcal{H})$, and the blackboard bold font, e.g., $\mathbb{D}$, $\mathbb{J}$, $\mathbb{H}$, $\mathbb{Q}$, for linear operators from $\mathcal{H}$ to Hilbert–Schmidt operators on $\mathcal{H}$, i.e., linear mappings $\mathcal{H} \to \mathrm{HS}(\mathcal{H}, \mathcal{H})$.

First, we introduce certain derivatives in the Fréchet sense as follows. Denote by $\mathscr{I}$ and $\mathfrak{I}$ the identity operators on $\mathcal{H}$ and $\mathrm{HS}(\mathcal{H}, \mathcal{H})$, respectively. The derivative

$$\mathscr{D}(\mathsf{P}; \mu) = \frac{\partial}{\partial \mu} G(\mathsf{P}; \mu) = E_{\mathsf{P}} \left[\frac{\rho'(\|X - \mu\|)}{\|X - \mu\|} \mathscr{I} + \left\{ \frac{\rho''(\|X - \mu\|)}{\|X - \mu\|^2} - \frac{\rho'(\|X - \mu\|)}{\|X - \mu\|^3} \right\} \mathscr{P}(X; \mu) \right] \quad \text{(A1)}$$

is a linear mapping from $\mathcal{H}$ to $\mathcal{H}$. The derivative

$$\begin{aligned} \mathfrak{D}(\mathsf{P}; \mathscr{R}, \mu) = \frac{\partial}{\partial \mathscr{R}} \mathscr{G}(\mathsf{P}; \mathscr{R}, \mu) = E_{\mathsf{P}} \Bigg(& \frac{\rho'\{\|\mathscr{P}(X; \mu) - \mathscr{R}\|\}}{\|\mathscr{P}(X; \mu) - \mathscr{R}\|} \mathfrak{I} \\ & + \left[\frac{\rho''\{\|\mathscr{P}(X; \mu) - \mathscr{R}\|\}}{\|\mathscr{P}(X; \mu) - \mathscr{R}\|^2} - \frac{\rho'\{\|\mathscr{P}(X; \mu) - \mathscr{R}\|\}}{\|\mathscr{P}(X; \mu) - \mathscr{R}\|^3} \right] \mathfrak{P}(X; \mathscr{R}, \mu) \Bigg), \end{aligned} \quad \text{(A2)}$$

where we denote $\mathfrak{P}(x; \mathscr{R}, \mu) = \{\mathscr{P}(x; \mu) - \mathscr{R}\} \otimes \{\mathscr{P}(x; \mu) - \mathscr{R}\}$, is a linear mapping from $\mathrm{HS}(\mathcal{H}, \mathcal{H})$ to $\mathrm{HS}(\mathcal{H}, \mathcal{H})$. We define

$$\mathbb{D}(\mathsf{P}; \mathscr{R}, \mu) = \frac{\partial}{\partial \mu} \mathscr{G}(\mathsf{P}; \mathscr{R}, \mu), \quad \text{(A3)}$$

which is a linear mapping from $\mathcal{H}$ to $\mathrm{HS}(\mathcal{H}, \mathcal{H})$. To compute it, we first compute

$$\mathbb{Q}(x; \mu) = \frac{\partial}{\partial \mu} \mathscr{P}(x; \mu).$$

We consider its value at some $f \in \mathcal{H}$, i.e., we investigate the operator $\mathbb{Q}(x; \mu) f \in \mathrm{HS}(\mathcal{H}, \mathcal{H})$. This is done through its coordinate representation as follows. For any $g_1, g_2 \in \mathcal{H}$, we have

$$\begin{aligned} \langle g_1, \{\mathbb{Q}(x; \mu) f\} g_2 \rangle &= \left\langle g_1, \left[\left\{ \frac{\partial}{\partial \mu} \mathscr{P}(x; \mu) \right\} f \right] g_2 \right\rangle = \left\{ \frac{\partial}{\partial \mu} \langle g_1, \mathscr{P}(x; \mu) g_2 \rangle \right\} f \\ &= \left\{ \frac{\partial}{\partial \mu} (\langle x - \mu, g_1 \rangle \langle x - \mu, g_2 \rangle) \right\} f = -(\langle x - \mu, g_2 \rangle g_1 + \langle x - \mu, g_1 \rangle g_2) f \\ &= -\langle x - \mu, g_2 \rangle \langle g_1, f \rangle - \langle x - \mu, g_1 \rangle \langle g_2, f \rangle. \end{aligned} \quad \text{(A4)}$$

Then, the derivative of $\mathscr{G}(\mathsf{P}; \mathscr{R}, \mu)$ with respect to μ evaluated at $f \in \mathcal{H}$ is

$$\begin{aligned}\mathbb{D}(\mathsf{P}; \mathscr{R}, \mu) f = &- E_{\mathsf{P}} \left[\frac{\rho'\{\|\mathscr{P}(X; \mu) - \mathscr{R}\|\}}{\|\mathscr{P}(X; \mu) - \mathscr{R}\|} \mathbb{Q}(X; \mu) f \right] \\ &- E_{\mathsf{P}} \left(\left[\frac{\rho''\{\|\mathscr{P}(X; \mu) - \mathscr{R}\|\}}{\|\mathscr{P}(X; \mu) - \mathscr{R}\|^2} - \frac{\rho'\{\|\mathscr{P}(X; \mu) - \mathscr{R}\|\}}{\|\mathscr{P}(X; \mu) - \mathscr{R}\|^3} \right] \right. \\ &\left. \times \langle \mathscr{P}(X; \mu) - \mathscr{R}, \mathbb{Q}(X; \mu) f \rangle \{\mathscr{P}(X; \mu) - \mathscr{R}\} \right).\end{aligned}$$

We set

$$\begin{aligned}\mathfrak{D}_0(\mathsf{P}_1, \mathsf{P}_2, a; \mathscr{R}, \mu_1, \mu_2) &= a\mathfrak{D}(\mathsf{P}_1; \mathscr{R}, \mu_1) + (1 - a)\mathfrak{D}(\mathsf{P}_2; \mathscr{R}, \mu_2), \\ \mathfrak{D}_1(\mathsf{P}_1, \mathsf{P}_2, a; \mathscr{R}, \mu_1, \mu_2) &= a\mathfrak{D}(\mathsf{P}_1; \mathscr{R}, \mu_1) - (1 - a)\mathfrak{D}(\mathsf{P}_2; \mathscr{R}, \mu_2).\end{aligned}$$

Next, using the notation $f^{\otimes 2} = f \otimes f$ for $f \in \mathcal{H}$ and $\mathscr{A}^{\otimes 2} = \mathscr{A} \otimes \mathscr{A}$ for $\mathscr{A} \in \mathrm{HS}(\mathcal{H}, \mathcal{H})$, we define

$$\begin{aligned}\mathscr{J}(\mathsf{P}; \mu) &= E_{\mathsf{P}} \left[\left\{ \frac{\rho'(\|X - \mu\|)}{\|X - \mu\|} (\mu - X) - G(\mathsf{P}; \mu) \right\}^{\otimes 2} \right] \\ \mathfrak{J}(\mathsf{P}; \mathscr{R}, \mu) &= E_{\mathsf{P}} \left(\left[\frac{\rho'\{\|\mathscr{P}(X; \mu) - \mathscr{R}\|\}}{\|\mathscr{P}(X; \mu) - \mathscr{R}\|} \{\mathscr{R} - \mathscr{P}(X; \mu)\} - \mathscr{G}(\mathsf{P}; \mathscr{R}, \mu) \right]^{\otimes 2} \right)\end{aligned}$$

and

$$\begin{aligned}\mathbb{J}(\mathsf{P}; \mathscr{R}, \mu) = E_{\mathsf{P}} \left(\left[\frac{\rho'\{\|\mathscr{P}(X; \mu) - \mathscr{R}\|\}}{\|\mathscr{P}(X; \mu) - \mathscr{R}\|} \{\mathscr{R} - \mathscr{P}(X; \mu)\} - \mathscr{G}(\mathsf{P}; \mathscr{R}, \mu) \right] \right. \\ \left. \otimes \left\{ \frac{\rho'(\|X - \mu\|)}{\|X - \mu\|} (\mu - X) - G(\mathsf{P}; \mu) \right\} \right).\end{aligned}$$

Next, we denote

$$\begin{aligned}\mathfrak{H}_1(\mathsf{P}_1, \mathsf{P}_2, a; \mathscr{R}, \mu_1, \mu_2) &= \mathfrak{I} - \mathfrak{D}_1(\mathsf{P}_1, \mathsf{P}_2, a; \mathscr{R}, \mu_1, \mu_2)\mathfrak{D}_0(\mathsf{P}_1, \mathsf{P}_2, a; \mathscr{R}, \mu_1, \mu_2)^{-1}, \\ \mathbb{H}_1(\mathsf{P}_1, \mathsf{P}_2, a; \mathscr{R}, \mu_1, \mu_2) &= \mathfrak{H}_1(\mathsf{P}_1, \mathsf{P}_2, a; \mathscr{R}, \mu_1, \mu_2)\mathbb{D}(\mathsf{P}_1; \mathscr{R}, \mu_1)\mathscr{D}(\mathsf{P}_1; \mu_1)^{-1}, \\ \mathfrak{H}_2(\mathsf{P}_1, \mathsf{P}_2, a; \mathscr{R}, \mu_1, \mu_2) &= \mathfrak{I} + \mathfrak{D}_1(\mathsf{P}_1, \mathsf{P}_2, a; \mathscr{R}, \mu_1, \mu_2)\mathfrak{D}_0(\mathsf{P}_1, \mathsf{P}_2, a; \mathscr{R}, \mu_1, \mu_2)^{-1}, \\ \mathbb{H}_2(\mathsf{P}_1, \mathsf{P}_2, a; \mathscr{R}, \mu_1, \mu_2) &= \mathfrak{H}_2(\mathsf{P}_1, \mathsf{P}_2, a; \mathscr{R}, \mu_1, \mu_2)\mathbb{D}(\mathsf{P}_2; \mathscr{R}, \mu_2)\mathscr{D}(\mathsf{P}_2; \mu_2)^{-1},\end{aligned}$$

where $\mathfrak{I}$ stands for the identity operator on $\mathrm{HS}(\mathcal{H}, \mathcal{H})$. Finally, we set

$$\mathfrak{W}(\mathsf{P}_1, \mathsf{P}_2, a; \mathscr{R}, \mu_1, \mu_2) = a\mathfrak{W}_1(\mathsf{P}_1, \mathsf{P}_2, a; \mathscr{R}, \mu_1, \mu_2) + (1 - a)\mathfrak{W}_2(\mathsf{P}_1, \mathsf{P}_2, a; \mathscr{R}, \mu_1, \mu_2), \tag{A5}$$

where

$$\begin{aligned}\mathfrak{W}_1(\mathsf{P}_1, \mathsf{P}_2, a; \mathscr{R}, \mu_1, \mu_2) = &\ \mathfrak{H}_1(\mathsf{P}_1, \mathsf{P}_2, a; \mathscr{R}, \mu_1, \mu_2)\mathfrak{J}(\mathsf{P}_1; \mathscr{R}, \mu_1)\mathfrak{H}_1(\mathsf{P}_1, \mathsf{P}_2, a; \mathscr{R}, \mu_1, \mu_2)^* \\ &- \mathfrak{H}_1(\mathsf{P}_1, \mathsf{P}_2, a; \mathscr{R}, \mu_1, \mu_2)\mathbb{J}(\mathsf{P}_1; \mathscr{R}, \mu_1)\mathbb{H}_1(\mathsf{P}_1, \mathsf{P}_2, a; \mathscr{R}, \mu_1, \mu_2)^* \\ &- \mathbb{H}_1(\mathsf{P}_1, \mathsf{P}_2, a; \mathscr{R}, \mu_1, \mu_2)\mathbb{J}(\mathsf{P}_1; \mathscr{R}, \mu_1)^*\mathfrak{H}_1(\mathsf{P}_1, \mathsf{P}_2, a; \mathscr{R}, \mu_1, \mu_2)^* \\ &+ \mathbb{H}_1(\mathsf{P}_1, \mathsf{P}_2, a; \mathscr{R}, \mu_1, \mu_2)\mathscr{J}(\mathsf{P}_1; \mathscr{R}, \mu_1)\mathbb{H}_1(\mathsf{P}_1, \mathsf{P}_2, a; \mathscr{R}, \mu_1, \mu_2)^*\end{aligned}$$

with * denoting adjoint operators, and $\mathfrak{W}_2(\mathsf{P}_1, \mathsf{P}_2, a; \mathscr{R}, \mu_1, \mu_2)$ is defined analogously with $\mathbb{H}_2, \mathfrak{H}_2$ in place of $\mathbb{H}_1, \mathfrak{H}_1$, respectively, and P_2 instead of P_1 in $\mathfrak{J}, \mathbb{J}, \mathscr{J}$.

REFERENCES

ADLER, R. J. (1990). *An Introduction to Continuity, Extrema, and Related Topics for General Gaussian Processes*. Institute of Mathematical Statistics Lecture Notes—Monograph Series, 12. Hayward: Institute of Mathematical Statistics.

AMZALLAG, A., VAILLANT, C., JACOB, M., UNSER, M., BEDNAR, J., KAHN, J. D., DUBOCHET, J., STASIAK, A. & MADDOCKS, J. H. (2006). 3D reconstruction and comparison of shapes of DNA minicircles observed by cryo-electron microscopy. *Nucleic Acids Res.* **34**, e125.

ANDERSON, M. J. (2006). Distance-based tests for homogeneity of multivariate dispersions. *Biometrics* **62**, 245–53.

BALI, L., BOENTE, G., TYLER, D. E. & WANG, J.-L. (2012). Robust functional principal components: A projection-pursuit approach. *Ann. Statist.* **39**, 2852–82.

BENKO, M., HÄRDLE, W. & KNEIP, A. (2009). Common functional principal components. *Ann. Statist.* **37**, 1–34.

BOENTE, G. & FRAIMAN, R. (1999). Comment on a paper by Locantore et al. *Test* **8**, 28–35.

BOENTE, G., RODRIGUEZ, D. & SUED, M. (2011). Testing the equality of covariance operators. In *Recent Advances in Functional Data Analysis and Related Topics*, Ed. F. Ferraty, pp. 49–53. Heidelberg: Physica-Verlag.

BOSQ, D. (2000). *Linear Processes in Function Spaces: Theory and Applications*. New York: Springer.

BOX, G. E. P. (1953). Non-normality and tests on variances. *Biometrika* **40**, 318–35.

CHAUDHURI, P. (1996). On a geometric notion of quantiles for multivariate data. *J. Am. Statist. Assoc.* **91**, 862–72.

DAUXOIS, J., POUSSE, A. & ROMAIN, Y. (1982). Asymptotic theory for the principal component analysis of a vector random function: some applications to statistical inference. *J. Mult. Anal.* **12**, 136–54.

FRITZ, H., FILZMOSER, P. & CROUX, C. (2012). A comparison of algorithms for the multivariate L_1-median. *Comp. Statist.*, to appear. doi: 10.1007/s00180-011-0262-4.

GABRYS, R. & KOKOSZKA, P. (2007). Portmanteau test of independence for functional observations. *J. Am. Statist. Assoc.* **102**, 1338–48.

GERVINI, D. (2006). Free-knot spline smoothing for functional data. *J. R. Statist. Soc.* B **68**, 671–87.

GERVINI, D. (2008). Robust functional estimation using the median and spherical principal components. *Biometrika* **95**, 587–600.

HALL, P. & HOSSEINI-NASAB, M. (2006). On properties of functional principal components analysis. *J. R. Statist. Soc.* B **68**, 109–26.

HALL, P., MÜLLER, H.-G. & WANG, J.-L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *Ann. Statist.* **34**, 1493–517.

HAMPEL, F. R., RONCHETTI, E. M., ROUSSEEUW, P. J. & STAHEL, W. A. (1986). *Robust Statistics*. New York: Wiley.

HORVÁTH, L., HUŠKOVÁ, M. & KOKOSZKA, P. (2010). Testing the stability of the functional autoregressive process. *J. Mult. Anal.* **101**, 352–67.

HUBER, P. J. & RONCHETTI, E. M. (2009). *Robust Statistics*. Hoboken: Wiley.

LAYARD, M. W. J. (1974). A Monte Carlo comparison of tests for equality of convariance matrices. *Biometrika* **61**, 461–5.

LI, G. & CHEN, Z. (1985). Projection-pursuit approach to robust dispersion matrices and principal components: Primary theory and Monte Carlo. *J. Am. Statist. Assoc.* **80**, 759–66.

LOCANTORE, N., MARRON, J. S., SIMPSON, D. G., TRIPOLI, N., ZHANG, J. T. & COHEN, K. L. (1999). Robust principal component analysis for functional data. *Test* **8**, 1–73.

MARDEN, J. I. (1999). Some robust estimates of principal components. *Statist. Prob. Lett.* **43**, 349–59.

O'BRIEN, P. C. (1992). Robust procedures for testing equality of covariance matrices. *Biometrics* **48**, 819–27.

OLSON, C. L. (1974). Comparative robustness of six tests in multivariate analysis of variance. *J. Am. Statist. Assoc.* **69**, 894–08.

PANARETOS, V. M., KRAUS, D. & MADDOCKS, J. H. (2010). Second-order comparison of Gaussian random functions and the geometry of DNA minicircles. *J. Am. Statist. Assoc.* **105**, 670–82.

R DEVELOPMENT CORE TEAM (2012). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0, http://www.R-project.org.

RAMSAY, J. & SILVERMAN, B. W. (2005). *Functional Data Analysis*. New York: Springer.

SERFLING, R. (2004). Nonparametric multivariate descriptive measures based on spatial quantiles. *J. Statist. Plan. Infer.* **123**, 259–78.

SIRKIÄ, S., TASKINEN, S., OJA, H. & TYLER, D. E. (2009). Tests and estimates of shape based on spatial signs and ranks. *J. Nonparam. Statist.* **21**, 155–76.

SUN, Y. & GENTON, M. G. (2011). Functional boxplots. *J. Comp. Graph. Statist.* **20**, 316–334.

TIKU, M. L. & BALAKRISHNAN, N. (1985). Testing the equality of variance-covariance matrices the robust way. *Commun. Statist.* A **14**, 3033–51.

YAO, F. & LEE, T. C. M. (2006). Penalized spline models for functional principal component analysis. *J. R. Statist. Soc.* B **68**, 3–25.

ZHANG, J., PANTULA, S. G. & BOOS, D. D. (1991). Robust methods for testing the pattern of a single covariance matrix. *Biometrika* **78**, 787–95.

[*Received April* 2011. *Revised May* 2012]