

# A test of weak separability for multi-way functional data, with application to brain connectivity studies

BY BRIAN LYNCH AND KEHUI CHEN

*Department of Statistics, University of Pittsburgh,  
Pittsburgh, Pennsylvania 15260, U.S.A.*

bcl28@pitt.edu khchen@pitt.edu

5

## SUMMARY

This paper concerns the modeling of multi-way functional data where double or multiple indices are involved. We introduce a concept of weak separability. The weakly separable structure supports the use of factorization methods that decompose the signal into its spatial and temporal components. The analysis reveals interesting connections to the usual strongly separable covariance structure, and provides insights into tensor methods for multi-way functional data. We propose a formal test for the weak separability hypothesis, where the asymptotic null distribution of the test statistic is a chi-square type mixture. The method is applied to study brain functional connectivity derived from source localized magnetoencephalography signals during motor tasks.

10

15

*Some key words:* asymptotics; functional principal component; hypothesis testing; marginal kernel; separable covariance; spatio-temporal data; tensor product.

## 1. INTRODUCTION

Traditional functional data analysis usually concerns data recorded over a continuum, such as growth curves. Dense and regularly-observed functional data can be recorded in a matrix with dimension  $n \times T$ , where  $n$  is the number of subjects and  $T$  is the number of grid points observed for each subject. Multi-way functional data refers to an extension where multiple indices are involved and data can be recorded in a tensor with dimension at least three. Examples include brain imaging data where for each subject  $i = 1, \dots, n$ , we have observations  $X_i(s, t)$ , with a spatial index  $s \in \mathbb{R}^d$  and a time index  $t \in \mathbb{R}^1$ . Other examples include repeatedly or longitudinally observed functional data, such as data obtained from tracking apps where subjects' 24-hour profiles of activities are recorded every day. This type of data can be represented by  $X_i(s, t)$ , where  $s$  denotes the day and  $t$  denotes the time within a day.

20

25

As multi-way functional data become more common with modern techniques, the modeling of this type of data attracts increasing interest. Assume the individual observations  $X_i(s, t)$  are independent and identically distributed realizations of a random process  $X \in L^2(\mathcal{S} \times \mathcal{T})$ ,  $s \in \mathcal{S} \subseteq \mathbb{R}^{d_1}$ ,  $t \in \mathcal{T} \subseteq \mathbb{R}^{d_2}$ , with mean  $\mu$  and continuous covariance operator  $C$ . When we can do so without confusion, we use the same symbol for the covariance operator and its kernel function. A well-established tool in functional data analysis is functional principal component analysis. When applied to the multi-way process  $X$ , functional principal component analysis is based on the Karhunen–Loève representation  $X(s, t) = \mu(s, t) + \sum_{l=1}^{\infty} Z_l h_l(s, t)$ , where  $Z_l$  ( $l = 1, 2, \dots$ ) are the (random) uncorrelated coefficients, and  $h_l(s, t)$  ( $l = 1, 2, \dots$ ) are the eigenfunctions of the covariance operator  $C$ .

30

35

40 To alleviate the difficulties associated with modeling the  $(2d_1 + 2d_2)$ -dimensional full covariance function  $C(s, t, u, v)$  and characterizing the  $(d_1 + d_2)$ -dimensional eigenfunctions, one usually seeks dimension reduction through factorization of the signal into its spatial and temporal components. Chen et al. (2017) proposed product functional principal component analysis,

$$X(s, t) = \mu(s, t) + \sum_{k=1}^{\infty} \sum_{j=1}^{\infty} \chi_{jk} \psi_j(s) \phi_k(t), \quad (1)$$

45 where  $\psi_j(s)$  ( $j = 1, 2, \dots$ ) and  $\phi_k(t)$  ( $k = 1, 2, \dots$ ) are the eigenfunctions of the marginal covariance operators in  $L^2(\mathcal{S})$  and  $L^2(\mathcal{T})$ , with corresponding marginal kernels

$$C_{\mathcal{S}}(s, u) = \int_{\mathcal{T}} C(s, t; u, t) dt, \quad C_{\mathcal{T}}(t, v) = \int_{\mathcal{S}} C(s, t; s, v) ds. \quad (2)$$

Here  $C_{\mathcal{S}}(s, u) = \sum_{j=1}^{\infty} \lambda_j \psi_j(s) \psi_j(u)$  and  $C_{\mathcal{T}}(t, v) = \sum_{k=1}^{\infty} \gamma_k \phi_k(t) \phi_k(v)$ , where  $\lambda_1 \geq \lambda_2 \geq \dots$  and  $\gamma_1 \geq \gamma_2 \geq \dots$  are the eigenvalues. The  $\chi_{jk} = \int_{\mathcal{T}} \int_{\mathcal{S}} \{X(s, t) - \mu(s, t)\} \psi_j(s) \phi_k(t) ds dt$  are the marginal projection scores. To be precise, we should first consider expanding  $X(s, t)$  in terms of completed versions of the bases of marginal eigenfunctions, but since it can be shown 50 that the scores  $\chi_{jk}$  associated with the extra functions needed to complete the bases are 0, the expansion of  $X(s, t)$  in Equation (1) holds.

The above product functional principal component analysis representation is the same as the Karhunen–Loève representation if one makes the separable covariance assumption  $C(s, t; u, v) = aC_1(s, u)C_2(t, v)$ , which we call strong separability in contrast to the weak separability that will be proposed in this paper. However, if strong separability is not assumed, the 55 marginal eigenfunctions no longer carry optimal efficiency guarantees (Aston et al., 2012), and can only be proven to have near-optimality under appropriate assumptions (Chen et al., 2017). Moreover, unlike the  $Z_l$  in multi-way functional principal component analysis, the scores  $\chi_{jk}$  are not guaranteed to be mutually uncorrelated.

60 Factorization of the signal into its spatial ( $s$ ) and temporal ( $t$ ) components, justified using a vague notion of spatial-temporal separability, is a common strategy used in many methods in image analysis and multi-way functional data analysis (Zhang & Zhou, 2005; Lu et al., 2006; Huang et al., 2009; Chen & Müller, 2012; Hung et al., 2012; Allen et al., 2014; Chen et al., 2015, 2017). Despite their empirical success, the rigorous characterization of this separable feature is 65 still mainly restricted to the scope of strong separability, i.e., when the covariance  $C(s, t; u, v)$  is separable. There is a large amount of literature on strong separability in related fields (Lu & Zimmerman, 2005; Fuentes, 2006; Srivastava et al., 2009; Hoff et al., 2011; Horváth & Kokoszka, 2012). Tests for strong separability in functional data settings have been proposed recently (Aston et al., 2017; Constantinou et al., 2017).

70 In this paper, we propose a new concept of weak separability for the process  $X$ , which can be rigorously tested. We show that under weak separability the eigenfunctions of the full covariance  $C(s, t; u, v)$  can be written as tensor products of the marginal eigenfunctions, i.e.,  $\psi_j \otimes \phi_k$ . This means the Karhunen–Loève representation is the same as the product representation in Equation (1), just as if we had strong separability, and to perform functional principal component 75 analysis we only need to calculate the marginal covariances  $C_{\mathcal{S}}(s, u)$  and  $C_{\mathcal{T}}(t, v)$  instead of the full covariance  $C(s, t; u, v)$ . The analysis reveals that if  $C(s, t; u, v)$  is separable, then the process  $X$  is weakly separable, but the converse is not necessarily true. Indeed, weak separability is a much weaker assumption than separable covariance.

80 We develop a test for weak separability based on the empirical correlations between the estimated scores  $\hat{\chi}_{i,jk}$  and  $\hat{\chi}_{i,j'k'}$   $\{i = 1, \dots, n; (j, k) \neq (j', k')\}$ . Although the  $\hat{\chi}_{i,jk}$  are  $n^{1/2}$ -

consistent estimators of the  $\chi_{i,jk}$ , the test statistics based on the  $\hat{\chi}_{i,jk}$  have different null distributions from their counterparts using the  $\chi_{i,jk}$  due to non-negligible estimation errors. The proofs involve expansions of the differences between the estimated marginal eigenfunctions and their true values, i.e.,  $\hat{\psi}_j - \psi_j$  and  $\hat{\phi}_k - \phi_k$ , as well as multi-way tensor products with indices  $(j, k, j', k')$ . A series of careful derivations are carried out to characterize the asymptotic null distribution of the test statistic, which is found to be a  $\chi^2$  type mixture. No Gaussian assumption on  $X$  is imposed. We apply the testing procedure to brain imaging data, where frequency and time-based functional connectivity is constructed from source localized magnetoencephalography signals. The test result supports the use of product functional principal component analysis methods and reveals interesting features about brain connectivity over time and frequency.

## 2. WEAK SEPARABILITY: CONCEPTS AND PROPERTIES

For  $\mathcal{S} \subseteq \mathbb{R}^{d_1}$  and  $\mathcal{T} \subseteq \mathbb{R}^{d_2}$ , we consider the space of square integrable surfaces  $L^2(\mathcal{S} \times \mathcal{T})$  with the standard inner product  $\langle f, g \rangle = \int_{\mathcal{T}} \int_{\mathcal{S}} f(s, t)g(s, t)dsdt$  and the corresponding norm  $\|\cdot\|$ . The data can be viewed as realizations of a random element  $X \in L^2(\mathcal{S} \times \mathcal{T})$ , which we assume has well defined mean function  $\mu$  and covariance operator  $C$ . We assume the covariance is continuous, and  $\mathcal{S}$  and  $\mathcal{T}$  are compact. Unless otherwise noted, these assumptions are used in all the lemmas and theorems.

For orthonormal bases  $f_j$  ( $j = 1, 2, \dots$ ) in  $L^2(\mathcal{S})$  and  $g_k$  ( $k = 1, 2, \dots$ ) in  $L^2(\mathcal{T})$ , the product functions  $f_j(s)g_k(t)$  ( $j = 1, 2, \dots; k = 1, 2, \dots$ ) form an orthonormal basis of  $L^2(\mathcal{S} \times \mathcal{T})$ . We can then have

$$X(s, t) = \mu(s, t) + \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \tilde{\chi}_{jk} f_j(s)g_k(t),$$

where  $\tilde{\chi}_{jk} = \int_{\mathcal{T}} \int_{\mathcal{S}} \{X(s, t) - \mu(s, t)\} f_j(s)g_k(t)dsdt$ .

**DEFINITION 1 (WEAK SEPARABILITY).**  $X(s, t)$  is weakly separable if there exist orthonormal bases  $f_j$  ( $j = 1, 2, \dots$ ) and  $g_k$  ( $k = 1, 2, \dots$ ) such that  $\text{cov}(\tilde{\chi}_{jk}, \tilde{\chi}_{j'k'}) = 0$  for  $j \neq j'$  or  $k \neq k'$ , i.e., the scores  $\tilde{\chi}_{jk}$  ( $j = 1, 2, \dots; k = 1, 2, \dots$ ) are uncorrelated with each other.

In the following, we list several important properties of weak separability, which make this concept attractive in many applications. Detailed proofs are given in the supplementary material.

**LEMMA 1.** If  $X$  is weakly separable, the pair of bases  $f_j$  ( $j = 1, 2, \dots$ ) and  $g_k$  ( $k = 1, 2, \dots$ ) that satisfies weak separability is unique up to a sign, and  $f_j(s) \equiv \psi_j(s)$  and  $g_k(t) \equiv \phi_k(t)$ , where  $\psi_j(s)$  and  $\phi_k(t)$  are the eigenfunctions of the marginal kernels  $C_{\mathcal{S}}(s, u)$  and  $C_{\mathcal{T}}(t, v)$  as defined in Equation (2). Moreover,

$$C(s, t; u, v) = \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \eta_{jk} \psi_j(s) \phi_k(t) \psi_j(u) \phi_k(v), \quad (3)$$

where  $\eta_{jk} = \text{var}(\langle X - \mu, \psi_j \otimes \phi_k \rangle)$ , and the convergence is absolute and uniform.

Lemma 1 shows that natural basis functions for the factorized spatial and temporal effects are eigenfunctions of the marginal kernels. Under weak separability, the eigenfunctions of the covariance  $C(s, t; u, v)$  can be written as tensor products of the marginal eigenfunctions, i.e.,  $\psi_j \otimes \phi_k$ , which could result in a substantial dimension reduction in applications. Lemma 1 also allows us to test the weak separability assumption (see Section 3).

LEMMA 2. *Strong separability, defined as  $C(s, t; u, v) = aC_1(s, u)C_2(t, v)$  with identifiability constraints  $\int_{\mathcal{S}} C_1(s, s)ds = 1$  and  $\int_{\mathcal{T}} C_2(t, t)dt = 1$ , implies weak separability of  $X$ . And up to a constant scaling,  $C_1$  and  $C_2$  are the same as the marginal kernels.*

Lemma 2 shows that strong separability is a special case of weak separability, and the following Lemma 3 further illustrates that weak separability is much more flexible than strong separability.

LEMMA 3. *Define the array  $V = (\eta_{jk}, j = 1, 2, \dots; k = 1, 2, \dots)$ . Strong separability is weak separability with the additional assumption that  $\text{rank}(V) = 1$ . Moreover, under strong separability  $V = a\Lambda\Gamma^T$ , where  $\Lambda = (\lambda_1, \lambda_2, \dots)^T$  and  $\Gamma = (\gamma_1, \gamma_2, \dots)^T$  are the eigenvalues of the marginal kernels, and  $a = 1 / \int_{\mathcal{T}} \int_{\mathcal{S}} C(s, t; s, t)dsdt$  is a normalization constant.*

When the covariance  $C$  is not strongly separable but the process  $X$  is weakly separable, we can show that the covariance function is a sum of  $L$  separable components,  $C(s, t; u, v) = \sum_{l=1}^L a^l C_{\mathcal{S}}^l(s, u)C_{\mathcal{T}}^l(t, v)$ , where  $L \geq \text{rank}(V) > 1$  is the nonnegative rank, defined as  $\text{rank}_+(V) = \min\{\ell : V = V_1 + \dots + V_\ell; V_i \geq 0, \text{rank}(V_i) = 1, \text{ for all } i\}$ , where  $V_i \geq 0$  means that  $V_i$  is entry-wise nonnegative. In applications where one relies on the separable structure of the covariance for ease of computation and interpretation, for example in applications involving the inverse of the covariance, it is not clear whether and how one can modify the concept to work under the weak separability assumption ( $L$  additive separable terms). We defer this to future research.

### 3. TEST OF WEAK SEPARABILITY

#### 3.1. Background

Assume we have a sample of independent and identically distributed smooth processes  $X_i(s, t) \sim X(s, t)$ , and the marginal projection scores  $\chi_{i,jk} = \int_{\mathcal{T}} \int_{\mathcal{S}} \{X_i(s, t) - \mu(s, t)\} \psi_j(s) \phi_k(t) dsdt$ , where  $\psi_j(s)$  and  $\phi_k(t)$  are the eigenfunctions of the marginal covariances. By the definition of weak separability and Lemma 1, testing weak separability is the same as testing the covariance structure of the marginal projection scores, i.e.,  $H_0 : \text{cov}(\chi_{jk}, \chi_{j'k'}) = 0$  for  $j \neq j'$  or  $k \neq k'$ .

The problem of testing covariance structure is a classic problem in multivariate analysis. Suppose we have independent and identically distributed copies of a  $p$ -variate random variable, with mean  $\mu$  and covariance matrix  $\Sigma$ , and we want to test the null hypothesis that  $\Sigma$  is diagonal. Under the traditional multivariate setting where  $p$  is fixed and does not increase with  $n$ , likelihood ratio methods can be used to test the diagonality of  $\Sigma$  (Anderson, 1984). The high-dimensional problem has been studied in the context that  $p/n \rightarrow \gamma \in (0, \infty)$  or even for  $p$  much larger than  $n$  (Ledoit & Wolf, 2002; Liu et al., 2008; Cai et al., 2011; Lan et al., 2015). If we were to observe the sample values  $\chi_{i,jk}$ , a sensible test statistic could be based on the off-diagonal terms of the empirical covariance, i.e.,  $n^{-1/2} \sum_{i=1}^n \chi_{i,jk} \chi_{i,j'k'}$ . However, unlike in the traditional covariance testing problem, we do not directly observe the sample values  $\chi_{i,jk}$ . Instead they are estimated from the sample curves  $X_i(s, t)$  ( $i = 1, \dots, n$ ) as

$$\hat{\chi}_{i,jk} = \int_{\mathcal{T}} \int_{\mathcal{S}} \{X_i(s, t) - \bar{X}(s, t)\} \hat{\psi}_j(s) \hat{\phi}_k(t) dsdt,$$

where  $\bar{X}(s, t) = (1/n) \sum_{i=1}^n X_i(s, t)$ , and  $\hat{\psi}_j$  and  $\hat{\phi}_k$  are eigenfunctions of the estimated marginal covariances  $\hat{C}_{\mathcal{S}}(s, u) = (1/n) \sum_{i=1}^n \int_{\mathcal{T}} \{X_i(s, t) - \bar{X}(s, t)\} \{X_i(u, t) - \bar{X}(u, t)\} dt$  and  $\hat{C}_{\mathcal{T}}(t, v) = (1/n) \sum_{i=1}^n \int_{\mathcal{S}} \{X_i(s, t) - \bar{X}(s, t)\} \{X_i(s, v) - \bar{X}(s, v)\} ds$ . In practice, if the

data for each subject are observed on arbitrarily dense and equally spaced grid points, and recorded in matrices  $X_i$  ( $i = 1, \dots, n$ ), the above estimators can be simplified as  $\hat{C}_S = (1/n) \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})^T$ , and  $\hat{C}_T = (1/n) \sum_{i=1}^n (X_i - \bar{X})^T (X_i - \bar{X})$ . The data cannot immediately be written as matrices if the argument  $s$  has dimension greater than 1, but as long as the observations are dense in  $\mathcal{S}$  one can vectorize them along a certain ordering of  $s$ , compute the marginal covariances, and reorganize back accordingly.

Although we can prove that the  $\hat{\chi}_{i,jk}$  are  $n^{1/2}$ -consistent estimators of the  $\chi_{i,jk}$ , test statistics based on  $n^{-1/2} \sum_{i=1}^n \hat{\chi}_{i,jk} \hat{\chi}_{i,j'k'}$  have different null distributions from their counterparts using the  $\chi_{i,jk}$ , and in the following we derive the asymptotic distribution of the former.

### 3.2. The test statistic and its properties

Let  $H$  be a real separable Hilbert space, with inner product  $\langle \cdot, \cdot \rangle$ . Following standard definitions, we denote the space of bounded linear operators on  $H$  as  $\mathcal{B}(H)$ , the space of Hilbert–Schmidt operators on  $H$  as  $\mathcal{B}_{HS}(H)$ , and the space of trace-class operators on  $H$  as  $\mathcal{B}_{Tr}(H)$ . For any trace-class operator  $T \in \mathcal{B}_{Tr}(H)$ , we define its trace by  $tr(T) = \sum_{i=1,2,\dots} \langle T e_i, e_i \rangle$ , where  $e_i$  ( $i = 1, 2, \dots$ ) is an orthonormal basis of  $H$ , and it is easy to see that this definition is independent of the choice of basis.

For  $H_1$  and  $H_2$  two real separable Hilbert spaces, we use  $\otimes$  as the standard tensor product, i.e., for  $x_1 \in H_1$  and  $x_2 \in H_2$ ,  $(x_1 \otimes x_2)$  is the operator from  $H_2$  to  $H_1$  defined by  $(x_1 \otimes x_2)y = \langle x_2, y \rangle x_1$  for any  $y \in H_2$ . With a bit of abuse of notation, we let  $H = H_1 \otimes H_2$  denote the tensor product Hilbert space, which contains all finite sums of  $x_1 \otimes x_2$ , with inner product  $\langle x_1 \otimes x_2, y_1 \otimes y_2 \rangle = \langle x_1, y_1 \rangle \langle x_2, y_2 \rangle$ , for  $x_1, y_1 \in H_1$  and  $x_2, y_2 \in H_2$ . For  $C_1 \in \mathcal{B}(H_1)$  and  $C_2 \in \mathcal{B}(H_2)$ , we let  $C_1 \otimes C_2$  denote the unique bounded linear operator on  $H_1 \otimes H_2$  satisfying  $C_1 \otimes C_2(x_1 \otimes x_2) = C_1 x_1 \otimes C_2 x_2$  for all  $x_1 \in H_1, x_2 \in H_2$ .

We define

$$T_n(j, k, j', k') = n^{-1/2} \sum_{i=1}^n \hat{\chi}_{i,jk} \hat{\chi}_{i,j'k'} \quad \{(j, k) \neq (j', k')\}, \quad (4)$$

and  $\mathcal{Z}_n = n^{1/2}(C_n - C)$ , where the sample covariance operator is defined as  $C_n = (1/n) \sum_{i=1}^n (X_i - \bar{X}) \otimes (X_i - \bar{X})$ .

The following two conditions are needed for the main theorem below and the corollary following it:

*Condition 1.* For some orthonormal basis  $e_j$  ( $j = 1, 2, \dots$ ) of  $L^2(\mathcal{S} \times \mathcal{T})$ ,  $\sum_{j=1,2,\dots} \{\mathbb{E}(\langle X, e_j \rangle^4)\}^{1/4} < \infty$ .

*Condition 2.* For some integers  $P$  and  $K$ , we have  $\delta_P = \min_{j=1,\dots,P} (\lambda_j - \lambda_{j+1}) > 0$  and  $\delta_K = \min_{k=1,\dots,K} (\gamma_k - \gamma_{k+1}) > 0$ .

*Remark 1.* According to Proposition 5 of Mas (2006), Condition 1 implies that  $\mathcal{Z}_n$  converges to a Gaussian random element in  $\mathcal{B}_{Tr}\{L^2(\mathcal{S} \times \mathcal{T})\}$ .

**THEOREM 1.** *Assume Conditions 1 and 2 hold, and that  $X$  is weakly separable. For  $j, j' = 1, \dots, P$  and  $k, k' = 1, \dots, K$  as defined in Condition 2, we have*

(i) for  $j \neq j'$  and  $k \neq k'$ ,

$$T_n(j, k, j', k') = tr [\{(\psi_j \otimes \psi_{j'}) \otimes (\phi_k \otimes \phi_{k'})\} \mathcal{Z}_n] + o_p(1),$$

(ii) for  $j = j'$  and  $k \neq k'$ ,

$$\begin{aligned} T_n(j, k, j, k') &= \text{tr} \left[ \{(\psi_j \otimes \psi_j) \tilde{\otimes} (\phi_k \otimes \phi_{k'})\} \mathcal{Z}_n \right] \\ &\quad + \text{tr} \left( [Id_1 \tilde{\otimes} \{\eta_{jk'}(\gamma_k - \gamma_{k'})^{-1} \phi_k \otimes \phi_{k'}\}] \mathcal{Z}_n \right) \\ &\quad + \text{tr} \left( [Id_1 \tilde{\otimes} \{\eta_{jk}(\gamma_{k'} - \gamma_k)^{-1} \phi_{k'} \otimes \phi_k\}] \mathcal{Z}_n \right) + o_p(1), \end{aligned}$$

195 (iii) for  $j \neq j'$  and  $k = k'$ ,

$$\begin{aligned} T_n(j, k, j', k) &= \text{tr} \left[ \{(\psi_j \otimes \psi_{j'}) \tilde{\otimes} (\phi_k \otimes \phi_k)\} \mathcal{Z}_n \right] \\ &\quad + \text{tr} \left( [\{\eta_{jk}(\lambda_{j'} - \lambda_j)^{-1} \psi_{j'} \otimes \psi_j\} \tilde{\otimes} Id_2] \mathcal{Z}_n \right) \\ &\quad + \text{tr} \left( [\{\eta_{j'k}(\lambda_j - \lambda_{j'})^{-1} \psi_j \otimes \psi_{j'}\} \tilde{\otimes} Id_2] \mathcal{Z}_n \right) + o_p(1), \end{aligned}$$

where  $Id_1$  and  $Id_2$  are identity operators on  $L^2(\mathcal{S})$  and  $L^2(\mathcal{T})$ , respectively.

200 *Remark 2.* Since  $n^{1/2} \text{tr} \{(\psi_j \otimes \psi_{j'}) \tilde{\otimes} (\phi_k \otimes \phi_{k'}) C\}$  is zero under the null hypothesis, the first term in each case of the above theorem is the same as  $n^{1/2} \text{tr} \{(\psi_j \otimes \psi_{j'}) \tilde{\otimes} (\phi_k \otimes \phi_{k'}) C_n\} = n^{-1/2} \sum_{i=1}^n \chi_{i,jk} \chi_{i,j'k'}$ , i.e., the counterpart of  $T_n$  as if we had the true marginal projection scores. The second and third terms, if they exist, are non-negligible estimation errors.

205 *COROLLARY 1.* Assume Conditions 1 and 2 hold, and that  $X$  is weakly separable. For different sets of  $(j, k, j', k')$ ,  $j, j' = 1, \dots, P$ ;  $k, k' = 1, \dots, K$ , satisfying  $(j, k) \neq (j', k')$ , the  $T_n(j, k, j', k')$ 's are asymptotically jointly Gaussian with mean zero and covariance structure  $\Theta$ . The formula for  $\Theta$  is given in the proof.

### 3.3. Tests based on $\chi^2$ type mixtures

210 *LEMMA 4.* For  $j \neq j'$ ,  $\sum_{k=1}^{\infty} \text{E}(\chi_{jk} \chi_{j'k}) = 0$ , and for  $k \neq k'$ ,  $\sum_{j=1}^{\infty} \text{E}(\chi_{jk} \chi_{jk'}) = 0$ . This also holds in the empirical version such that for  $j \neq j'$ ,  $\sum_{k=1}^{\infty} T_n(j, k, j', k) = 0$ , and for  $k \neq k'$ ,  $\sum_{j=1}^{\infty} T_n(j, k, j, k') = 0$ .

The above lemma does not assume weak separability. Recall the fact that principal component scores in traditional functional principal component analysis are uncorrelated. This lemma is a generalized result for the marginal projection scores.

215 Due to this linear relationship between the different terms of  $T_n$ , the asymptotic covariance  $\Theta$  will be degenerate, and thus the statistic we consider is the sum of squares of the terms of  $T_n$  without normalizing by the covariance. In practice, for suitably chosen  $P_n$  and  $K_n$ , we use the statistic defined as

$$S_n = \sum_{j, j'=1, \dots, P_n; k, k'=1, \dots, K_n; (j, k) < (j', k')} \{T_n(j, k, j', k')\}^2,$$

220 where  $(j, k) < (j', k')$  means  $(j - 1) * K_n + k < (j' - 1) * K_n + k'$ .

Take  $T_n$  to be a long vector of length  $m = P_n K_n (P_n K_n - 1) / 2$  created by stacking all of the  $T_n(j, k, j', k')$   $\{j, j' = 1, \dots, P_n; k, k' = 1, \dots, K_n; (j, k) < (j', k')\}$ . Then by Corollary 1,  $T_n \sim N_m(0, \Theta)$  under  $H_0$ , where we now take  $\Theta$  to be a covariance matrix. Define the spectral decomposition of  $\Theta$  as  $\Theta = UQU^T$ , where  $Q$  is diagonal with diagonal entries  $\sigma_1, \dots, \sigma_m$ , which are the eigenvalues of  $\Theta$  ordered from largest to smallest, and  $U = [u_1 \dots u_m]$ , where the  $u_i$  are orthonormal column vectors. By Lemma 4, some of the  $\sigma_i$  are 0. Since  $S_n = \|T_n\|^2 = \|U^T T_n\|^2$  and  $U^T T_n \sim N_m(0, Q)$ , we can write  $S_n = \sum_{i=1}^m \sigma_i A_i$  where the  $A_i$  are independent and identically distributed  $\chi_1^2$ , i.e., the null distribution of  $S_n$  is a weighted sum of  $\chi^2$  distributions, which we call a  $\chi^2$  type mixture.

The Welch–Satterthwaite approximation for a  $\chi^2$  type mixture (Zhang, 2013) approximates  $S_n \sim \beta \chi_d^2$  and determines  $\beta$  and  $d$  from matching the first 2 cumulants (the mean and the variance). This results in  $\beta = \text{tr}(\Theta^2)/\text{tr}(\Theta)$  and  $d = \{\text{tr}(\Theta)\}^2/\text{tr}(\Theta^2)$ . By using a plug-in estimator of  $\Theta$ , we can approximate the P-value for our test as an upper tail probability of  $\beta \chi_d^2$ . When the first  $(P_n, K_n)$  terms do not satisfy weak separability, we have  $S_n \rightarrow \infty$  in probability by noticing that for at least one set of  $(j, k, j', k')$ , the first term in Equation (1) in the proof of Theorem 1 is on the order of  $n^{1/2}$ .

The consistent selection of  $(P_n, K_n)$  for hypothesis testing is a challenging problem. The optimal choice of  $(P_n, K_n)$  needs to be defined according to the problem at hand and subsequent analysis of interest. Here we focus on the subspace where the subsequent product functional principal component analysis is going to be carried out. A criterion we will use to evaluate a given choice of  $(P_n, K_n)$  is the fraction of variance explained by the first  $P_n$  and  $K_n$  components, defined as

$$\text{FVE}(P_n, K_n) = \frac{\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{P_n} \sum_{k=1}^{K_n} \hat{\chi}_{i,jk}^2}{\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \hat{\chi}_{i,jk}^2}.$$

This definition can be justified by noting its relation to the normalized mean squared  $L^2$  loss of the truncated process  $\tilde{X}(s, t) = \mu(s, t) + \sum_{j=1}^{P_n} \sum_{k=1}^{K_n} \chi_{jk} \psi_j(s) \phi_k(t)$ . In particular,

$$\frac{\text{E}(\|X - \tilde{X}\|^2)}{\text{E}(\|X - \mu\|^2)} = 1 - \frac{\sum_{j=1}^{P_n} \sum_{k=1}^{K_n} \eta_{jk}}{\sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \eta_{jk}}.$$

The latter term is approximated by our definition of fraction of variance explained. The above equality only relies on the orthogonality of the eigenfunctions, not the weak separability assumptions. Thus, it still makes sense to consider this definition of fraction of variance explained even when  $H_0$  is not true.

We also define the marginal fractions of variance explained as  $\text{FVE}_S(P_n) = \sum_{j=1}^{P_n} \hat{\lambda}_j / \sum_{j=1}^{\infty} \hat{\lambda}_j$  and  $\text{FVE}_T(K_n) = \sum_{k=1}^{K_n} \hat{\gamma}_k / \sum_{k=1}^{\infty} \hat{\gamma}_k$ , where the  $\hat{\lambda}_j$  are the eigenvalues of  $\hat{C}_S$  and the  $\hat{\gamma}_k$  are the eigenvalues of  $\hat{C}_T$ . In practice the infinite sums in the denominators of  $\text{FVE}(P_n, K_n)$ ,  $\text{FVE}_S(P_n)$ , and  $\text{FVE}_T(K_n)$  will have to be replaced with the largest number of terms that can reasonably be considered nonzero.

Noting that  $\sum_{j=1}^{\infty} \text{E}(\chi_{jk}^2) = \gamma_k$ ,  $\sum_{k=1}^{\infty} \text{E}(\chi_{jk}^2) = \lambda_j$  and  $\sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \text{E}(\chi_{jk}^2) = \sum_{j=1}^{\infty} \lambda_j = \sum_{k=1}^{\infty} \gamma_k$ , we have

$$\text{FVE}(P_n, K_n) \gtrsim \text{FVE}_S(P_n) + \text{FVE}_T(K_n) - 1,$$

subject to estimation error (to see, for example, that  $\sum_{k=1}^{\infty} \text{E}(\chi_{jk}^2) = \lambda_j$ , take  $j = j'$  in the proof of Lemma 4, with no need to assume weak separability). Therefore, we propose the following fraction of variance explained procedure: First choose  $P_n$  and  $K_n$  such that the marginal fractions of variance explained are at least 90%. If this choice results in  $\text{FVE}(P_n, K_n) \geq 90\%$ , use these values of  $P_n$  and  $K_n$ . If not, use the values of  $P_n$  and  $K_n$  that have marginal fractions of variance explained at least 95%, in which case  $\text{FVE}(P_n, K_n)$  is expected to be above 90%.

### 3.4. Bootstrap approximation

As an alternative to asymptotic approximation, we can also consider a bootstrap approach to approximate the distribution of the test statistic. Theorem 1 provides theoretical support for the use of the following empirical bootstrap procedure (Van Der Vaart & Wellner, 1996). Our simulations show that the asymptotic approximation based on the  $\chi^2$  type mixture has very satisfactory

265 performance. We still present the bootstrap approximation here since it is generally applicable to similar tests where the asymptotic null distributions do not have closed form.

At each step, draw a random sample from the data  $X_1, \dots, X_n$  with replacement. Denote this sample as  $X_1^*, \dots, X_n^*$ . Let

$$\hat{\chi}_{i,jk}^* = \int_{\mathcal{T}} \int_{\mathcal{S}} \{X_i^*(s, t) - \bar{X}^*(s, t)\} \hat{\psi}_j^*(s) \hat{\phi}_k^*(t) ds dt,$$

270 where  $\bar{X}^*$  is the sample mean of the  $X_i^*$ , and the  $\hat{\psi}_j^*$  and  $\hat{\phi}_k^*$  are the eigenfunctions of the estimated marginal covariances of the  $X_i^*$ . The signs of the  $\hat{\psi}_j^*$  and  $\hat{\phi}_k^*$  are chosen to minimize  $\|\hat{\psi}_j^* - \hat{\psi}_j\|$  and  $\|\hat{\phi}_k^* - \hat{\phi}_k\|$ , respectively. Let

$$T_n^*(j, k, j', k') = n^{-1/2} \sum_{i=1}^n \hat{\chi}_{i,jk}^* \hat{\chi}_{i,j'k'}^*.$$

The empirical bootstrap test statistic is calculated as

$$S_n^* = \sum_{j, j'=1, \dots, P_n; k, k'=1, \dots, K_n; (j, k) < (j', k')} \{T_n^*(j, k, j', k') - T_n(j, k, j', k')\}^2.$$

This procedure is repeated  $B$  times, and the P-value is approximated as the proportion of bootstrap test statistics  $S_n^*$  that are larger than the test statistic  $S_n$ .

275 Theorem 3.9.13 in Van Der Vaart & Wellner (1996) can be used to prove the validity of the bootstrap procedure, i.e., the conditional random laws (given data) of  $S_n^*$  are asymptotically consistent almost surely for estimating the laws of  $S_n$ , under the null hypothesis. By Theorem 1, we have that under the null hypothesis,  $T_n$  can be written as  $\Phi'_P\{n^{1/2}(\mathcal{P}_n - P)\} + o(1)$  and  $T_n^* - T_n$  can be written as  $\Phi'_P\{n^{1/2}(\mathcal{P}_n^* - \mathcal{P}_n)\} + o(1)$ , where  $\Phi'_P$  is a linear continuous mapping that depends on the three different cases in Theorem 1. Thus, Theorem 3.9.13 applies.

280 Other than the above non-studentized empirical bootstrap based on  $S_n$ , we have also considered a bootstrap procedure based on a marginally studentized test statistic, in which we divide each term in  $S_n$  by its corresponding estimated variance  $\hat{\theta}(j, k, j', k')$  (which is the plug-in estimate of  $\theta(j, k, j', k')$ , the diagonal entry of  $\Theta$  corresponding to the asymptotic variance of  $T_n(j, k, j', k')$ ). However, we have found this procedure is much more time consuming, and requires substantially higher sample size to achieve high power, in comparison to the non-studentized empirical bootstrap method. This is not unexpected, since the form of  $\theta(j, k, j', k')$  is very complicated and plug-in estimation adds extra variability. Therefore, we do not recommend the marginally studentized empirical bootstrap method.

#### 290 4. NUMERICAL STUDY

In this section, we conduct numerical experiments to check the size and power of the proposed test for weak separability. Following the notation in Section 3, we are basically testing if the  $\text{cov}(\chi_{jk}, \chi_{j'k'})$  are all zero for  $(j, k) \neq (j', k')$ . We consider two different choices for the joint distribution of the  $\chi_{jk}$ . The first is the multivariate normal and the second is the multivariate  $t$  distribution. The diagonal values of  $\Sigma$ , the covariance matrix of the  $\chi_{jk}$ , are determined by the matrix  $V = \{\text{var}(\chi_{jk}), j, k = 1, \dots, 8\}$ . We consider two different choices for  $V$ , which we denote as  $V_1$  and  $V_2$  (specified later). Under  $H_0$  (when all of the off-diagonal values of  $\Sigma$  are 0),  $V_1$  corresponds to a strongly separable covariance structure, while  $V_2$  corresponds to a weakly separable structure that is not strongly separable. To study power, for a given choice of  $V_1$  or



$V_2$ , we take  $\text{cov}(\chi_{i,12}, \chi_{i,21})$  to be the largest positive value such that  $\Sigma$  is positive definite, and we also consider half of this value. Alternatively, we let 3 off-diagonal terms,  $\text{cov}(\chi_{i,12}, \chi_{i,21})$ ,  $\text{cov}(\chi_{i,11}, \chi_{i,22})$ , and  $\text{cov}(\chi_{i,13}, \chi_{i,31})$ , take their largest positive values such that  $\Sigma$  is positive definite. 300

Empirical rejection rates at the .05 significance level from 200 simulations runs for  $n = 50, 100, 500$  are shown in Tables 1 through 4. The fraction of variance explained method described in Section 3.3 ends up with  $P_n = 3$  and  $K_n = 2$  in most trials, and we show results with  $(P_n, K_n)$  chosen by this procedure, as well as directly setting  $(P_n, K_n) = (2, 2), (3, 3)$ , or  $(4, 4)$  for all trials. We see that both the  $\chi^2$  type mixture approximation and the empirical bootstrap procedure are able to control the type I error under all scenarios and achieve very good power as  $n$  or the signal increase, although the empirical bootstrap is slightly less powerful for small  $n$ . Even when the chosen nonzero off-diagonal covariance terms are set to their maximum values, the other off-diagonal covariance terms of  $\Sigma$  are zero, and so the signal is moderate. The test procedures are slightly less powerful in the multivariate  $t$  case for small  $n$ , as the asymptotics likely come into play more quickly for the normal data. The rejection rates are in general stable across different choices of  $(P_n, K_n)$ ; although  $(P_n, K_n) = (2, 2)$  seems to have higher power in some cases, the power stabilizes to a reasonable value for larger  $(P_n, K_n)$ . 305  
310  
315

Table 1. Rejection rates for the  $\chi^2$  type mixture weak separability test procedure, using  $V_1$  and choosing  $(P_n, K_n)$  with the fraction of variance explained procedure or as  $(2, 2)$ ,  $(3, 3)$ , or  $(4, 4)$

Scenario	Normal				Multivariate $t$			
	FVE	(2,2)	(3,3)	(4,4)	FVE	(2,2)	(3,3)	(4,4)
$n = 50$								
$H_0$	0.055	0.020	0.020	0.040	0.025	0.020	0.005	0.045
$\text{cov}(\chi_{12}, \chi_{21}) = 0.065$	0.715	0.785	0.740	0.755	0.440	0.445	0.395	0.370
$\text{cov}(\chi_{12}, \chi_{21}) = 0.13$	1.000	1.000	1.000	1.000	0.935	0.965	0.940	0.940
3 nonzero terms	1.000	1.000	1.000	1.000	0.960	0.990	0.990	0.965
$n = 100$								
$H_0$	0.035	0.075	0.045	0.050	0.025	0.040	0.020	0.010
$\text{cov}(\chi_{12}, \chi_{21}) = 0.065$	0.985	0.985	0.985	0.985	0.800	0.810	0.785	0.710
$\text{cov}(\chi_{12}, \chi_{21}) = 0.13$	1.000	1.000	1.000	1.000	1.000	0.990	0.990	0.990
3 nonzero terms	1.000	1.000	1.000	1.000	1.000	0.990	1.000	1.000
$n = 500$								
$H_0$	0.060	0.060	0.040	0.055	0.020	0.045	0.020	0.045
$\text{cov}(\chi_{12}, \chi_{21}) = 0.065$	1.000	1.000	1.000	1.000	0.995	0.995	1.000	0.990
$\text{cov}(\chi_{12}, \chi_{21}) = 0.13$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3 nonzero terms	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

FVE, fraction of variance explained procedure.

Details of the simulation settings are as follows: We generate independent samples of data  $X_i(s, t) = \sum_{j=1}^8 \sum_{k=1}^8 \chi_{i,jk} \psi_j(s) \phi_k(t)$  ( $i = 1, \dots, n$ ), where the scores  $\chi_{i,jk}$  are mean 0 random variables that we generate directly. We let  $s$  and  $t$  take values from 0 to 1 on an evenly spaced grid of 20 points. For the  $\psi_j$  we use the functions  $\psi_j(s) = -2^{1/2} \cos\{\pi(n+1)s\}$  for  $j$  odd and  $\psi_j(s) = 2^{1/2} \sin(\pi ns)$  for  $j$  even. We define the  $\phi_k$  by taking the first 3 B-spline functions produced by Matlab's `spcol` function using order 4 with knots at 0, 0.5, and 1, combining these with the first 5  $\psi_j$  as defined above, and orthonormalizing using Gram-Schmidt. 320

Let  $\chi_i$  be the vector of  $\chi_{i,jk}$  for  $j, k = 1, \dots, 8$ . We simulate each  $\chi_i$  independently from either  $N(0, \Sigma)$  or the multivariate  $t$  distribution. In the latter case, we first simulate a vector  $x$  of 325

Table 2. Rejection rates for the  $\chi^2$  type mixture weak separability test procedure, using  $V_2$  and choosing  $(P_n, K_n)$  with the fraction of variance explained procedure or as (2, 2), (3, 3), or (4, 4)

Scenario	Normal				Multivariate $t$			
	FVE	(2,2)	(3,3)	(4,4)	FVE	(2,2)	(3,3)	(4,4)
$n = 50$								
$H_0$	0.030	0.025	0.030	0.025	0.015	0.030	0.015	0.005
$\text{cov}(\chi_{12}, \chi_{21}) = 0 \cdot 055$	0.515	0.845	0.440	0.465	0.305	0.555	0.225	0.205
$\text{cov}(\chi_{12}, \chi_{21}) = 0 \cdot 11$	0.995	0.995	0.995	0.990	0.850	0.955	0.825	0.770
3 nonzero terms	1.000	1.000	1.000	1.000	0.965	0.985	0.950	0.970
$n = 100$								
$H_0$	0.045	0.055	0.035	0.040	0.010	0.050	0.040	0.020
$\text{cov}(\chi_{12}, \chi_{21}) = 0 \cdot 055$	0.920	0.990	0.930	0.920	0.625	0.900	0.605	0.500
$\text{cov}(\chi_{12}, \chi_{21}) = 0 \cdot 11$	1.000	1.000	1.000	1.000	0.990	1.000	0.965	0.955
3 nonzero terms	1.000	1.000	1.000	1.000	0.995	1.000	1.000	0.980
$n = 500$								
$H_0$	0.045	0.065	0.025	0.040	0.025	0.065	0.050	0.035
$\text{cov}(\chi_{12}, \chi_{21}) = 0 \cdot 055$	1.000	1.000	1.000	1.000	0.970	1.000	0.995	0.995
$\text{cov}(\chi_{12}, \chi_{21}) = 0 \cdot 11$	1.000	1.000	1.000	1.000	1.000	1.000	0.990	0.995
3 nonzero terms	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

FVE, fraction of variance explained procedure.

Table 3. Rejection rates for the non-studentized empirical bootstrap weak separability test procedure, using  $V_1$  and choosing  $(P_n, K_n)$  with the fraction of variance explained procedure or as (2, 2), (3, 3), or (4, 4)

Scenario	Normal				Multivariate $t$			
	FVE	(2,2)	(3,3)	(4,4)	FVE	(2,2)	(3,3)	(4,4)
$n = 50$								
$H_0$	0.035	0.025	0.035	0.025	0.010	0.010	0.005	0.005
$\text{cov}(\chi_{12}, \chi_{21}) = 0 \cdot 065$	0.670	0.680	0.650	0.640	0.350	0.375	0.330	0.300
$\text{cov}(\chi_{12}, \chi_{21}) = 0 \cdot 13$	1.000	1.000	1.000	1.000	0.890	0.900	0.885	0.880
3 nonzero terms	1.000	1.000	1.000	1.000	0.920	0.920	0.920	0.915
$n = 100$								
$H_0$	0.070	0.060	0.060	0.060	0.015	0.010	0.015	0.015
$\text{cov}(\chi_{12}, \chi_{21}) = 0 \cdot 065$	0.990	0.995	0.985	0.985	0.735	0.785	0.700	0.680
$\text{cov}(\chi_{12}, \chi_{21}) = 0 \cdot 13$	1.000	1.000	1.000	1.000	0.970	0.970	0.965	0.960
3 nonzero terms	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$n = 500$								
$H_0$	0.065	0.050	0.060	0.060	0.055	0.050	0.055	0.055
$\text{cov}(\chi_{12}, \chi_{21}) = 0 \cdot 065$	1.000	1.000	1.000	1.000	0.995	0.995	0.995	0.995
$\text{cov}(\chi_{12}, \chi_{21}) = 0 \cdot 13$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3 nonzero terms	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

FVE, fraction of variance explained procedure.

length 64 from  $N(0, \Sigma)$ . One standard definition of a multivariate  $t$  vector is  $x/(u/v)^{1/2}$ , where  $u$  is a chi-square random variable with  $v$  degrees of freedom that is independent of  $x$ . However, we use  $x/\{u/(v-2)\}^{1/2}$  as our multivariate  $t$  vector so that its covariance matrix is  $\Sigma$ . We take  $v = 6$  in our simulations. For each of 200 trials, we simulate data  $X_i(s, t)$  ( $i = 1, \dots, n$ ) in the manner described above, estimate the marginal projection scores, calculate the test statistic,

Table 4. Rejection rates for the non-studentized empirical bootstrap weak separability test procedure, using  $V_2$  and choosing  $(P_n, K_n)$  with the fraction of variance explained procedure or as  $(2, 2)$ ,  $(3, 3)$ , or  $(4, 4)$

Scenario	Normal				Multivariate $t$			
	FVE	(2,2)	(3,3)	(4,4)	FVE	(2,2)	(3,3)	(4,4)
$n = 50$								
$H_0$	0.065	0.045	0.050	0.045	0.015	0.020	0.010	0.010
$\text{cov}(\chi_{12}, \chi_{21}) = 0 \cdot 055$	0.420	0.785	0.390	0.380	0.220	0.395	0.160	0.150
$\text{cov}(\chi_{12}, \chi_{21}) = 0 \cdot 11$	0.970	1.000	0.965	0.965	0.730	0.865	0.690	0.675
3 nonzero terms	1.000	1.000	1.000	1.000	0.895	0.925	0.885	0.885
$n = 100$								
$H_0$	0.025	0.010	0.020	0.020	0.035	0.035	0.030	0.020
$\text{cov}(\chi_{12}, \chi_{21}) = 0 \cdot 055$	0.950	1.000	0.955	0.955	0.585	0.855	0.575	0.515
$\text{cov}(\chi_{12}, \chi_{21}) = 0 \cdot 11$	1.000	1.000	1.000	1.000	0.980	0.995	0.980	0.975
3 nonzero terms	1.000	1.000	1.000	1.000	0.985	0.990	0.985	0.985
$n = 500$								
$H_0$	0.030	0.065	0.030	0.030	0.010	0.035	0.005	0
$\text{cov}(\chi_{12}, \chi_{21}) = 0 \cdot 055$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$\text{cov}(\chi_{12}, \chi_{21}) = 0 \cdot 11$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3 nonzero terms	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

FVE, fraction of variance explained procedure.

and obtain P-values from the test procedures as described in Section 3, using  $B = 1000$  for the bootstrap procedure.

We choose  $V_1$  and  $V_2$  to both give  $\lambda_j = \exp\{1 \cdot 2(9 - j)\} / \left\{ \sum_{j'=1}^8 \exp(1 \cdot 2j') \right\}$  ( $j = 1, \dots, 8$ ) and  $\gamma_k = \exp\{1 \cdot 6(9 - k)\} / \left\{ \sum_{k'=1}^8 \exp(1 \cdot 6k') \right\}$  ( $k = 1, \dots, 8$ ) as the eigenvalues of the marginal covariances  $C_S$  and  $C_T$ .  $V_1$  is defined as the rank 1 matrix computed from the outer product of the vectors of  $\lambda_j$  and  $\gamma_k$ , while  $V_2$  is a rank 2 matrix with first 2 rows multiples of each other and rows 3 through 8 multiples of each other.

## 5. APPLICATION TO BRAIN CONNECTIVITY STUDIES

Brain imaging analysis is an area where functional data increasingly arise. An important goal in brain imaging studies is to analyze functional connectivity between different regions of the brain. We focus on magnetoencephalography (MEG), which measures neuronal activity by recording magnetic fields generated within the brain. We use MEG data collected by the Human Connectome Project, a study that has compiled a large amount of high quality multi-modal neural data, much of which is freely accessible at <https://db.humanconnectome.org> (Van Essen et al., 2013; WU-Minn HCP, 2017). We will focus on the motor task data, particularly the trials where subjects moved their right hand. The signal for each trial is recorded from -1.2 to 1.2 seconds in intervals of about 2 ms, where time 0 corresponds to the start of the motion. In the preprocessed sensor-level data (see WU-Minn HCP (2017) for preprocessing details), there are 61 subjects with motor data, and the subjects have an average of 75.38 trials. Our connectivity analysis will focus on two regions of interest: the left primary motor cortex and the right inferior parietal lobule. These regions of interest are spatially separated, likely activated during the task, and potentially functionally connected.

As the MEG sensors are distant from the brain, directly using their signals to represent regions of interest can lead to spurious connectivity measurements. This is due to the volume conduction/field spread problem, in which each sensor picks up the activity of several sources, as well as the common input problem, in which a common source provides input to a pair of signals that do not directly interact (Larson-Prior et al., 2013; Bastos & Schoffelen, 2015). For these reasons, we will use source reconstruction to estimate the signals arising from the cortical surface. Source reconstruction is common in MEG analysis, but it is an inverse problem on which constraints must be placed to obtain a unique solution (Pizzella et al., 2014). The source reconstruction method we use is minimum norm estimation as implemented in the Matlab package FieldTrip (Lin et al., 2004; Oostenveld et al., 2010).

MEG signals are inherently oscillatory, and synchronization at certain frequency ranges of the activity in different regions has been shown to be related to tasks performed by the brain (Pizzella et al., 2014). To study how frequency-based coupling between regions of interest changes over the course of a task, we calculate the time-frequency representations of their signals based on Morlet wavelets, using FieldTrip’s `ft_freqanalysis` function. The time-frequency representation of a signal is its representation at time  $t$  and frequency  $s$  as a complex number  $A(s, t)e^{iB(s, t)}$ , where  $A(s, t)$  is the amplitude and  $B(s, t)$  is the phase.

Given time-frequency representations  $A_{1,k}(s, t)e^{iB_{1,k}(s, t)}$  and  $A_{2,k}(s, t)e^{iB_{2,k}(s, t)}$  for two signals recorded in trial  $k$ ,  $k = 1, \dots, n_T$ , we use the phase locking value (Lachaux et al., 1999) to measure their connectivity, calculated as

$$\text{PLV}(s, t) = (1/n_T) \left| \sum_{k=1}^{n_T} e^{i\{B_{1,k}(s, t) - B_{2,k}(s, t)\}} \right|.$$

The phase locking value takes values from 0 to 1, with 1 indicating complete phase synchrony over trials and 0 indicating no phase synchrony. The phase locking value, like many connectivity measures, is based on an analogue of the cross-correlation function called the coherence, but the phase locking value disregards the amplitudes and considers only the magnitude of the average of the phase differences as unit vectors in the complex plane. The phase locking value has gained popularity due to the belief that phase differences reveal more about functional connectivity than changes in amplitude (Lachaux et al., 1999; Aydore et al., 2013; Bastos & Schoffelen, 2015).

Because we calculate the time-frequency representation using wider time windows for lower frequencies, we are limited in how low of frequencies we can consider, and our preliminary results for power show a lack of activity above 50 Hz. Thus, we calculate the time-frequency representation from 8 to 50 Hz, corresponding to the alpha to gamma low frequency bands. In each trial, the motion usually lasts no longer than about 0.75 seconds. The signal at a time period shortly before time 0 is of interest, as it can represent brain activity when subjects have received the movement cue but have not yet reacted to it. However, the trials are not disjoint, so the signal at times further before 0 overlaps with the signal from the previous trial. Thus, in our analysis we will consider times for each trial on the range of -0.25 to 0.75 seconds. Calculating phase locking value between the two source-reconstructed signals corresponding to our regions of interest, the data we analyze is  $\text{PLV}_i(s, t)$ , where  $i = 1, \dots, 61$ ;  $8 \leq s \leq 50$ ; and  $-0.25 \leq t \leq 0.75$ .

Figure 1 shows the average of the phase locking value matrices over all subjects, as well as slightly smoothed phase locking value matrices for 3 randomly selected subjects. The level of activity seems to vary between subjects. The average phase locking value displays higher synchrony near the beginning of the movement (time 0) in the alpha and beta bands, and the individual subjects’ plots also show higher values near time 0. However, the average has small

values overall, which indicates high variability between subjects, and points to the need to study covariance structure and modes of variation.

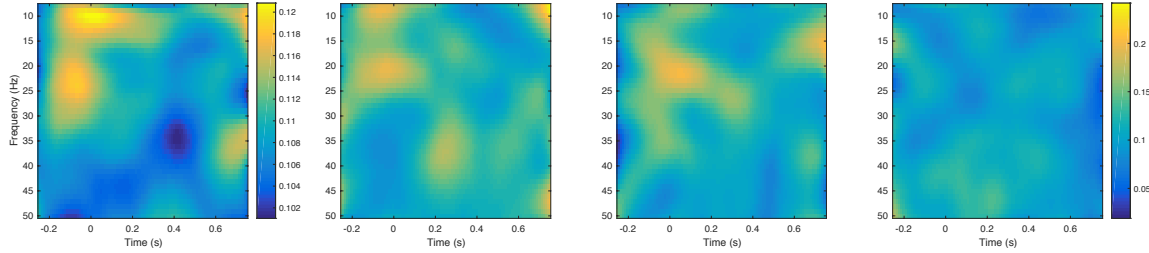


Fig. 1. Plots of the average source-level phase locking value (left), and the source-level phase locking value for 3 subjects. The latter plots use the same color scale, which is different than that used for the average.

For the phase locking value data described above, using the fraction of variance explained procedure described in Section 3-3, we choose the number of components to be  $P_n = 7$  and  $K_n = 7$ . We apply the weak separability test using both the  $\chi^2$  type mixture approximation (P-value = 0.5293) and the empirical bootstrap (P-value = 0.9260). The weak separability test does not reject the null hypothesis of weak separability, which supports the use of functional principal component analysis based on products of marginal eigenfunctions. We also apply the strong separability test of Aston et al. (2017) via their R package covsep (Tavakoli, 2016). The resulting P-values are  $1 \cdot 198 \times 10^{-4}$  for their chi-square approximation and 0.08 for their non-studentized empirical bootstrap method.

Product functional principal component analysis represents the data  $PLV_i(s, t)$  with products  $\psi_j(s)\phi_k(t)$  of the marginal eigenfunctions, where  $\psi_j(s)$  represents the frequency component and  $\phi_k(t)$  represents the time component. The 3 estimated eigenfunction products that account for the most variance are  $\hat{\psi}_1(s)\hat{\phi}_1(t)$ ,  $\hat{\psi}_2(s)\hat{\phi}_1(t)$ , and  $\hat{\psi}_3(s)\hat{\phi}_1(t)$ . The variance explained by  $\hat{\psi}_1(s)\hat{\phi}_1(t)$  is by far the largest. The estimated marginal eigenvalues  $\hat{\lambda}_j$  and  $\hat{\gamma}_k$  are plotted in Figure 2. We see the first eigenvalue dominates the others, and there is also a slight drop between the second two  $\hat{\lambda}_j$  and the rest, reflecting the fact that  $\hat{\psi}_2(s)\hat{\phi}_1(t)$  and  $\hat{\psi}_3(s)\hat{\phi}_1(t)$  are the products that explain the second and third highest amounts of variance, respectively.

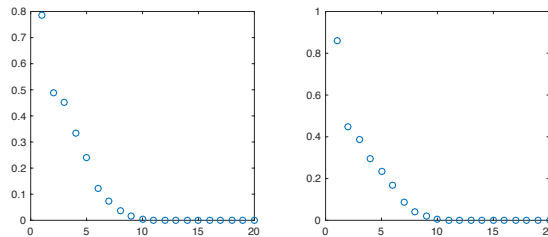


Fig. 2. Plots of the first 20 estimated marginal eigenvalues  $\hat{\lambda}_j$  (left, for frequency) and  $\hat{\gamma}_k$  (right, for time).

The product functions  $\hat{\psi}_1(s)\hat{\phi}_1(t)$ ,  $\hat{\psi}_2(s)\hat{\phi}_1(t)$ , and  $\hat{\psi}_3(s)\hat{\phi}_1(t)$  are plotted in Figure 3. These products capture modes of variation mainly around -0.2 to 0.2 s, from when the subject receives the cue to move to when they just start moving. This variation can be seen more clearly in the first temporal eigenfunction  $\hat{\phi}_1(t)$  (shown on the bottom right of Figure 4, which plots the individual

420 marginal eigenfunctions), which peaks slightly after 0 s.  $\hat{\psi}_1(s)\hat{\phi}_1(t)$  shows that, within this time range, subjects generally vary in synchrony from the alpha band to the beginning of the gamma low band, peaking within the beta band around 20–30 Hz.  $\hat{\psi}_2(s)\hat{\phi}_1(t)$  shows a contrast between the beta low band and gamma low band. That is, subjects with higher  $\chi_{21}$  values have lower synchrony in the beta low band and higher synchrony in the gamma low band.  $\hat{\psi}_3(s)\hat{\phi}_1(t)$  shows a contrast between the alpha band and the beta high band.

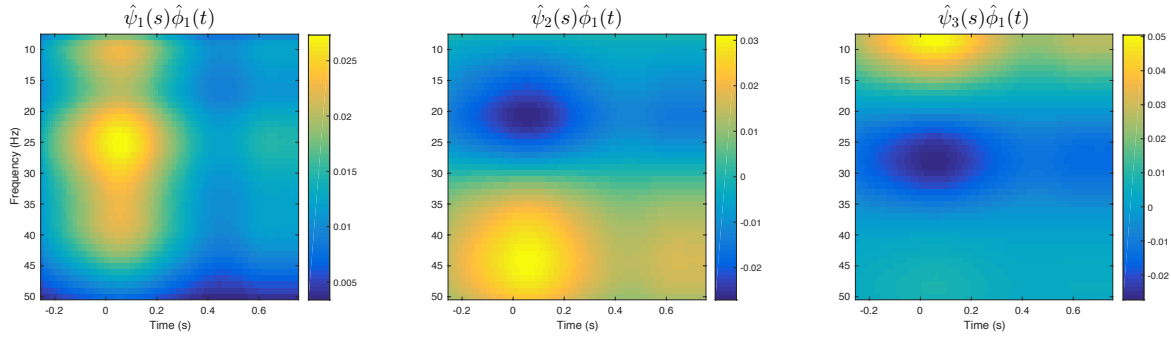


Fig. 3. Plots of the products of the estimated eigenfunctions that explain the most variance.

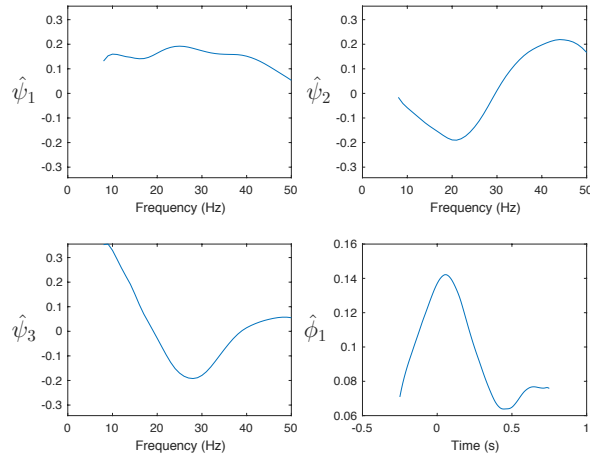


Fig. 4. Plots of the estimated eigenfunctions  $\hat{\psi}_j(s)$  and  $\hat{\phi}_k(t)$  whose products explain the most variance.

## 6. DISCUSSION

Much of the benefit of using product functional principal component analysis under weak separability is related to ease of interpretation and computation; by representing the eigenfunctions as tensor products of the marginal eigenfunctions, one consumes far fewer degrees of freedom and only needs to compute the marginal covariances instead of the full covariance. When the weak separable assumption does not hold, the product functional principal component analysis scores are correlated, and one expects to have to use more terms in product functional principal

430

component analysis than in conventional functional principal component analysis to explain the same amount of variance. Product functional principal component analysis can still be used under the alternative as a dimension reduction approach, but one needs to be aware of the above issues. We believe that the notion of weak separability will inspire new methodological developments for multi-way functional data analysis, such as multi-way regularization on marginal components (Chen & Lei, 2015). 435

Although our data example has  $s$  and  $t$  in  $\mathbb{R}$ , both the theory and the accompanying code for our test work for scenarios where  $d_1 > 1$  or  $d_2 > 1$ . For example, when modeling brain imaging data  $X_i(s, t)$  observed on a dense grid of voxels or dipoles over the cortical surface, in which  $s \in \mathbb{R}^2$  or  $s \in \mathbb{R}^3$ , one can first vectorize along  $s$  by ordering the dipoles from 1 to  $M$ , compute the marginal covariances, perform the hypothesis test of weak separability, and reorganize back to the space  $\mathcal{S}$  for interpretation and visualization. However, in scenarios where data are only very sparsely observed on the domain, where individual-subject smoothing is not appropriate, the problem is much more challenging and beyond the scope of this paper. We will possibly pursue it in future work. 440  
445

#### ACKNOWLEDGEMENT

This work is partially supported by NSF1612458.

Data were provided in part by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University. 450

#### SUPPLEMENTARY MATERIAL

Supplementary material available at *Biometrika* online includes proofs of the lemmas, theorem, and corollary (.pdf file), and a .zip compressed file archive with Matlab functions (.m files) used to implement the test procedures, a readme file, and example code (.m file) demonstrating the test procedures using an example data array  $X$  (.mat file). 455

#### REFERENCES

- ALLEN, G. I., GROSENICK, L. & TAYLOR, J. (2014). A generalized least-square matrix decomposition. *Journal of the American Statistical Association* **109**, 145–159. 460
- ANDERSON, T. W. (1984). *An introduction to multivariate statistical analysis*. Wiley New York, 2nd ed.
- ASTON, J. A., KIRCH, C. et al. (2012). Evaluating stationarity via change-point alternatives with applications to fmri data. *The Annals of Applied Statistics* **6**, 1906–1948.
- ASTON, J. A., PIGOLI, D., TAVAKOLI, S. et al. (2017). Tests for separability in nonparametric covariance operators of random surfaces. *The Annals of Statistics* **45**, 1431–1461. 465
- AYDORE, S., PANTAZIS, D. & LEAHY, R. M. (2013). A note on the phase locking value and its properties. *Neuroimage* **74**, 231–244.
- BASTOS, A. M. & SCHOFFELEN, J.-M. (2015). A tutorial review of functional connectivity analysis methods and their interpretational pitfalls. *Frontiers in systems neuroscience* **9**.
- CAI, T. T., JIANG, T. et al. (2011). Limiting laws of coherence of random matrices with applications to testing covariance structure and construction of compressed sensing matrices. *The Annals of Statistics* **39**, 1496–1525. 470
- CHEN, K., DELICADO, P. & MÜLLER, H.-G. (2017). Modelling function-valued stochastic processes, with applications to fertility dynamics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **79**, 177–196.
- CHEN, K. & LEI, J. (2015). Localized functional principal component analysis. *Journal of the American Statistical Association* **110**, 1266–1275. 475

- CHEN, K. & MÜLLER, H.-G. (2012). Modeling repeated functional observations. *Journal of the American Statistical Association* **107**, 1599–1609.
- 480 CHEN, K., ZHANG, X., PETERSEN, A. & MÜLLER, H.-G. (2015). Quantifying infinite-dimensional data: Functional data analysis in action. *Statistics in Biosciences*, 1–23.
- CONSTANTINOU, P., KOKOSZKA, P. & REIMHERR, M. (2017). Testing separability of space-time functional processes. *Biometrika* **104**, 425–437.
- FUENTES, M. (2006). Testing for separability of spatial–temporal covariance functions. *Journal of statistical planning and inference* **136**, 447–466.
- 485 HOFF, P. D. et al. (2011). Separable covariance arrays via the tucker product, with applications to multivariate relational data. *Bayesian Analysis* **6**, 179–196.
- HORVÁTH, L. & KOKOSZKA, P. (2012). *Inference for Functional Data with Applications*. New York: Springer.
- HUANG, J. Z., SHEN, H. & BUJA, A. (2009). The analysis of two-way functional data using two-way regularized singular value decompositions. *Journal of the American Statistical Association* **104**, 1609–1620.
- 490 HUNG, H., WU, P.-S., TU, I., HUANG, S.-Y. et al. (2012). On multilinear principal component analysis of order-two tensors. *Biometrika* **99**, 569–583.
- LACHAUX, J.-P., RODRIGUEZ, E., MARTINERIE, J., VARELA, F. J. et al. (1999). Measuring phase synchrony in brain signals. *Human brain mapping* **8**, 194–208.
- LAN, W., LUO, R., TSAI, C.-L., WANG, H. & YANG, Y. (2015). Testing the diagonality of a large covariance matrix in a regression setting. *Journal of Business & Economic Statistics* **33**, 76–86.
- 495 LARSON-PRIOR, L. J., OOSTENVELD, R., DELLA PENNA, S., MICHALAREAS, G., PRIOR, F., BABAJANI-FEREMI, A., SCHOFFELEN, J.-M., MARZETTI, L., DE PASQUALE, F., DI POMPEO, F. et al. (2013). Adding dynamics to the human connectome project with meg. *Neuroimage* **80**, 190–201.
- LEDOIT, O. & WOLF, M. (2002). Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *The Annals of Statistics* **30**, 1081–1102.
- 500 LIN, F.-H., WITZEL, T., HÄMÄLÄINEN, M. S., DALE, A. M., BELLIVEAU, J. W. & STUFFLEBEAM, S. M. (2004). Spectral spatiotemporal imaging of cortical oscillations and interactions in the human brain. *Neuroimage* **23**, 582–595.
- LIU, W.-D., LIN, Z., SHAO, Q.-M. et al. (2008). The asymptotic distribution and berry–esseen bound of a new test for independence in high dimension with an application to stochastic optimization. *The Annals of Applied Probability* **18**, 2337–2366.
- 505 LU, H., PLATANIOTIS, K. N. & VENETSANOPOULOS, A. N. (2006). Multilinear principal component analysis of tensor objects for recognition. In *18th International Conference on Pattern Recognition*, vol. 2. IEEE.
- LU, N. & ZIMMERMAN, D. L. (2005). The likelihood ratio test for a separable covariance matrix. *Statistics & probability letters* **73**, 449–457.
- 510 MAS, A. (2006). A sufficient condition for the clt in the space of nuclear operators—application to covariance of random functions. *Statistics & probability letters* **76**, 1503–1509.
- OOSTENVELD, R., FRIES, P., MARIS, E. & SCHOFFELEN, J.-M. (2010). Fieldtrip: open source software for advanced analysis of meg, eeg, and invasive electrophysiological data. *Computational intelligence and neuroscience* **2011**.
- 515 PIZZELLA, V., MARZETTI, L., DELLA PENNA, S., DE PASQUALE, F., ZAPPASODI, F. & ROMANI, G. L. (2014). Magnetoencephalography in the study of brain dynamics. *Functional neurology* **29**, 241.
- SRIVASTAVA, M. S., VON ROSEN, T. & VON ROSEN, D. (2009). Estimation and testing in general multivariate linear models with kronecker product covariance structure. *Sankhyā: The Indian Journal of Statistics, Series A* **71**, 137–163.
- 520 TAVAKOLI, S. (2016). *covsep: Tests for Determining if the Covariance Structure of 2-Dimensional Data is Separable*. R package version 1.0.0.
- VAN DER VAART, A. & WELLNER, J. (1996). *Weak convergence and empirical processes*. Springer Verlag.
- 525 VAN ESSEN, D. C., SMITH, S. M., BARCH, D. M., BEHRENS, T. E., YACOUB, E., UGURBIL, K., CONSORTIUM, W.-M. H. et al. (2013). The wu-minn human connectome project: an overview. *Neuroimage* **80**, 62–79.
- WU-MINN HCP (2017). *1200 Subjects Data Release Reference Manual*.
- ZHANG, D. & ZHOU, Z.-H. (2005). (2d) 2pca: Two-directional two-dimensional pca for efficient face representation and recognition. *Neurocomputing* **69**, 224–231.
- ZHANG, J.-T. (2013). *Analysis of variance for functional data*. CRC Press.

[Received 27 December 2017. Editorial decision on 20 April 2018]