Semiparametric Partial Common Principal Component Analysis for Covariance Matrices

BINGKAI WANG

Johns Hopkins Bloomberg School of Public Health, Baltimore, U.S.A.

XI LUO

The University of Texas Health Science Center at Houston, Houston, U.S.A.

YI ZHAO

Indiana University School of Medicine, Indianapolis, U.S.A

BRAIN CAFFO

Johns Hopkins Bloomberg School of Public Health, Baltimore, U.S.A.

Summary

We consider the problem of jointly modeling multiple covariance matrices by partial common principal component analysis (PCPCA), which assumes a proportion of eigenvectors to be shared across covariance matrices and the rest to be individual specific. This paper proposes consistent estimators of shared eigenvectors (called CPCs) in PCPCA as the number of matrices or the number of samples to estimate each matrix goes to infinity. We prove such asymptotic results without making any assumptions on the ranks of the CPC-related eigenvalues. When the number of samples goes to infinity, our result does not require the data to be Gaussian distributed. Furthermore, this paper introduces a sequential testing procedure to identify the number of shared eigenvectors in PCPCA. In simulation studies, our method shows higher accuracy in estimating

CPCs than competing methods. Applied to a motor-task functional magnetic resonance imaging data set, our estimator identifies meaningful brain networks that are consistent with current scientific understandings.

Key words: Consistency; Eigenvectors; Partial common principle components; Semiparametric; Sequential testing.

1. Introduction

Common principal component analysis (CPCA) is an approach that simultaneously models multiple covariance matrices. It extends the idea of principal component analysis by assuming all covariance matrices share the same set of eigenvectors. Since it was first introduced by Flury (1984), CPCA has been extensively applied in various fields including statistics (Gu, 2016; Pepler et al., 2016), finance (Goyal et al., 2008; Xu et al., 2019), and computer science (Ye et al., 2012; Hadjipantelis et al., 2015).

Extensions of CPCA have been investigated from multiple angles. Flury (1987) proposed partial common principal component analysis (PCPCA), where only a proportion of the eigenvectors was assumed to be shared across covariance matrices and the rest to be individually specific. Another direction relaxed the Gaussianity assumption in CPCA, resulting in asymptotic theory for non-Guassian distributions (Boik, 2002; Hallin et al., 2010). Other extensions include, Bayesian approaches (Hoff, 2009), algorithm acceleration (Browne and McNicholas, 2014) and modifications for high-dimensional data (Franks and Hoff, 2016). Among these extensions of CPCA, PCPCA continues to be appealing, as it relaxes the assumption of a completely common eigenspace across matrices while partially preserving the straightforward interpretation of common principal components. Related work on this topic includes Krzanowski (1984); Schott (1999); Boik (2002); Lock et al. (2013); Pepler et al. (2016).

In spite of these extensions, some questions related to PCPCA remain unanswered. Given the

number of CPCs, how one can identify common components from a pool of principal components requires further investigation. Most of the literature simply assumes that the CPCs are those eigenvectors associated with the largest eigenvalues across all covariance matrices (Flury, 1987; Schott, 1999; Pepler et al., 2016). However, the CPCs may be associated with small eigenvalues, or the corresponding eigenvalue of the CPC ranks differently across matrices. This question becomes more challenging if the number of the CPCs is unknown or the data is not Gaussian distributed. Regarding these, Pepler et al. (2016) developed a non-parametric method to select the CPCs in a special case of two covariance matrices. With multiple asymmetric matrices as the response, Lock et al. (2013) proposed a linear model to identify latent factors that explain the joint and individual data variation (JIVE) and Zhou et al. (2016) generalized JIVE to a common and individual feature extraction (CIFE) framework. None of these methods, however, studied the asymptotic properties.

In this paper, we propose a semiparametric PCPCA approach, which can consistently estimate CPCs, without making any assumptions on the ranks of CPC-related eigenvalues. Our method of estimating CPCs builds on an idea from Krzanowski (1984), where a semiparametric approach was proposed in the context of CPCA. We extend this idea to semiparametric PCPCA and provide asymptotic results for our methods as the number of matrices, or the number of samples to estimate each matrix, goes to infinity (both with fixed dimension). If the number of samples goes to infinity, our results do not require the data to be Gaussian distributed. When the number of the CPCs is unknown, we develop a sequential testing procedure, which effectively controls the type I error for Gaussian distributed data. As shown in the simulation study, our method outperforms existing methods in estimating CPCs in a variety of scenarios.

In the next section, we introduce the PCPCA model. In Section 3, we present our proposed semiparametric method to identify the common principal components. We compare our method with Flury's method through simulation studies in Section 4. An application to an fMRI data set is provided in Section 5. Section 6 summarises this paper and discusses future directions.

2. Model and assumptions

We consider a data set, $\{\mathbf{y}_{it}\}$, for $t \in \{1, \ldots, T\}$ and $i \in \{1, \ldots, n\}$, where $\mathbf{y}_{it} \in \mathbb{R}^p$ are independent and identically distributed random samples from a p-dimensional distribution with mean zero and covariance matrix $\mathbf{\Sigma}_i$. In our application example, $\mathbf{y}_{it}$ is a sample of brain fMRI measurements of p regions from subject i at time point t. We assume that $\mathbf{\Sigma}_i$ satisfies the following partial common principal component (PCPC) model:

$$\mathbf{\Sigma}_i = \sum_{j=1}^{k} \lambda_{ij} \boldsymbol{\gamma}_j \boldsymbol{\gamma}_j^\top + \sum_{l=1}^{p-k} \lambda_{i(l+k)} \mathbf{r}_{il} \mathbf{r}_{il}^\top, \tag{2.1}$$

where $\{\lambda_{ij}\}_{j=1}^p$ are the eigenvalues of covariance matrix $\mathbf{\Sigma}_i$. The $\boldsymbol{\gamma}_j$, for $j = 1, \ldots, k$, are the common eigencomponents across subjects. Let $\mathbf{\Gamma} = (\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_k) \in \mathbb{R}^{p \times k}$ $(k \leqslant p)$ be the orthonormal matrix of the common components. The $\mathbf{r}_{il}$, for $l = 1, \ldots, p-k$, are individual specific eigencomponents of subject i. Let $\mathbf{R}_i = (\mathbf{r}_{i1}, \ldots, \mathbf{r}_{i,p-k}) \in \mathbb{R}^{p \times (p-k)}$ be the orthonormal matrix of these components. We assume that $\mathbf{R}_i$ is orthogonal to $\mathbf{\Gamma}$, i.e., $\mathbf{\Gamma}^\top \mathbf{R}_i = \mathbf{0}$. Let $\mathbf{\Lambda}_i = \mathrm{diag}\{\lambda_{i1}, \ldots, \lambda_{ik}\}$ and $\mathbf{\Psi}_i = \sum_{l=1}^{p-k} \lambda_{i(l+k)} \mathbf{r}_{il} \mathbf{r}_{il}^\top$. Then, the PCPC model (2.1) can be reformulated as:

$$\mathbf{\Sigma}_i = \mathbf{\Gamma} \mathbf{\Lambda}_i \mathbf{\Gamma}^\top + \mathbf{\Psi}_i.$$

First proposed by Flury (1987), the PCPC model has an interpretation analogous to CPCA. A CPC, defined by $\boldsymbol{\gamma}_j$, is shared across all subjects. For example, in our application, a CPC represents a functional brain network in the sense that it represents correlations in functional brain measures consistent across subjects. The corresponding diagonal entry of $\mathbf{\Lambda}_i$ is interpreted as the variation of the CPC in subject i. On the other hand, $\mathbf{\Psi}_i$ is the subject-specific model components, which varies across subjects. A toy example of the PCPC model with $p = 4$ and $k = 2$ is shown in Figure 1.

Our goal is to both find k and estimate $\mathbf{\Gamma}$ consistently, as either $n \to \infty$ or $T \to \infty$, with p fixed. When estimating $\mathbf{\Gamma}$, existing methods, such as Schott (1999); Crainiceanu et al. (2011); Pepler et al. (2016), assumed that CPCs are associated with the largest eigenvalues across all

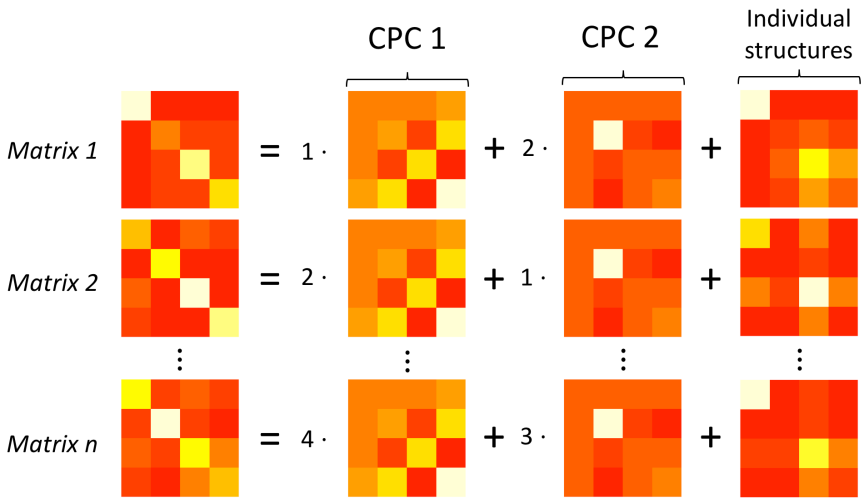

Figure 1. An example of the PCPC model. Each covariance matrix consists of two CPCs and an individual structure. Each CPC has rank 1 and norm 1.

covariance matrices. This is a restrictive assumption, since the corresponding eigenvalue of a CPC may rank consistently low or differently across matrices. For instance, our toy example in Figure 1 shows that CPCs are associated with small eigenvalues in matrices 1 and 2, but with large eigenvalues in matrix n. In addition, in many scientific applications, there is no priori reason to assume that the variation explained by the common components dominates the variation explained by the subject-specific components. For this reason, we do not make any assumptions regarding the rank of CPC-related eigenvalues.

The PCPC model shares some common features with existing partial information decomposition methods, but there exist major differences. Crainiceanu et al. (2011) provided a population value decomposition (PVD) model where common eigenvectors are extracted from concatenated individual eigenvectors. This procedure presumes that common inter-subject components are associated with largest individual eigenvalues. Moreover, the approach does not consider group level diagonalization as a goal. Lock et al. (2013) introduced the JIVE model, which decomposes information from multiple data sources into common components and individual components. Zhou et al. (2016) generalized the JIVE model by a common and individual feature extraction

(CIFE) framework, which has the same objective function as JIVE. Unlike the PCPC model, the common components identified by JIVE and CIFE are not unique, which can make them hard to interpret in practice. Furthermore, PVD, JIVE and CIFE are empirical methods with no asymptotic guarantees.

More recently, Wang et al. (2019) proposed a common reducing subspace model, which assumes $\mathbf{\Sigma}_i = \mathbf{\Gamma}\mathbf{\Omega}_0\mathbf{\Gamma}^\top + \tilde{\mathbf{\Gamma}}\mathbf{\Omega}_i\tilde{\mathbf{\Gamma}}^\top$, where $(\mathbf{\Gamma}, \tilde{\mathbf{\Gamma}}) \in \mathbb{R}^{p\times p}$ forms an eigenbasis and $\mathbf{\Omega}_0 \in \mathbb{R}^{k\times k}, \mathbf{\Omega}_i \in \mathbb{R}^{(p-k)\times(p-k)}, i = 1, \ldots, n$ are positive-definite matrices. This model can be reformulated as $\mathbf{\Sigma}_i = \mathbf{\Gamma}\mathbf{\Lambda}\mathbf{\Gamma}^\top + \mathbf{\Psi}_i$, where $\mathbf{\Lambda} \in \mathbb{R}^{k\times k}$ is a positive-definite diagonal matrix shared across i and $\mathbf{\Psi}_i \in \mathbb{R}^{p\times p}$ is a semi-positive matrix orthogonal to $\mathbf{\Gamma}$ with rank $p-k$. Compared with the PCPC model (2.1), this model requires $\mathbf{\Lambda}_i \equiv \mathbf{\Lambda}$, and is hence a special case of the PCPC model.

To achieve the identifiability of $\mathbf{\Gamma}$ and the consistency of our proposed estimator, for the PCPC model (2.1), we impose the following assumptions for the asymptotics when $n \to \infty$.

Assumption A (for $n \to \infty$):

1. T and p are fixed with $T > 0$ and $p > 1$.

2. Each $(\lambda_{i1}, \ldots, \lambda_{ik})$, $i \in \{1, \ldots, n\}$, is an independent and identically distributed random sample from a distribution with finite mean $(\lambda_1^*, \ldots, \lambda_k^*)$ and finite variance. Furthermore, elements of $(\lambda_{i1}, \ldots, \lambda_{ik})$ are independent to each other.

3. Each $\mathbf{\Psi}_i$, $i \in \{1, \ldots, n\}$, is an independent and identically distributed random sample from a distribution with finite mean $\mathbf{\Psi}^*$ and finite second-order moment. Both $\mathbf{\Psi}_i$ and $\mathbf{\Psi}^*$ are symmetric semi-positive definite matrices with rank $p-k$ and are orthogonal to $\mathbf{\Gamma}$.

4. For each $i \in \{1, \ldots, n\}$, $\mathbf{y}_{it}$ is normally distributed given $\mathbf{\Sigma}_i$.

To the best of our knowledge, we are the first to provide asymptotic results as the number of matrices goes to infinity. Different from the literature where n is fixed (Flury, 1987; Boik, 2002; Pepler et al., 2016; Wang et al., 2019), Assumptions A (b) and (c) assume $(\lambda_{i1}, \ldots, \lambda_{ik})$ and $\mathbf{\Psi}_i$ are

random variables instead of fixed parameters, since otherwise, the number of parameters would explode as n increases. The Gaussian assumption in Assumption A (d) is made for convenience and is stronger than required for our results. For the proof, we only need the fourth-order moment of $\mathbf{y}_{it}$ to be the same as the fourth-order moment of a Gaussian distribution, with mean 0 and covariance $\mathbf{\Sigma}_i$.

In some cases, n is small but T is large. For example, in fMRI data analysis, the number of subjects may be small, but subjects may have long fMRI scans. In other measures with rapid sampling, such as electroencephalograms, this is frequently the case. For such data sets, we prove a similar asymptotic theory as $T \to \infty$ with n and p fixed. This asymptotic theory requires the following assumptions.

Assumption B (for $T \to \infty$):

1. n and p are fixed with $n > 1$ and $p > 1$.
2. For each $i \in \{1, \ldots, n\}$, $(\lambda_{i1}, \ldots, \lambda_{ik})$ and $\mathbf{\Psi}_i$ are fixed.
3. The eigenvalues of $\sum_{i=1}^{n} \mathbf{\Sigma}_i/n$ are distinct.
4. The fourth-order moment of $\mathbf{y}_{it}$ is bounded for $i = 1, \ldots, n$.

Assumption B (b) implies that the asymptotics are conditional on $\mathbf{\Sigma}_i, i = 1, \ldots, n$. The reason to pursue conditional asymptotics is that n is fixed and inference on these specific n distributions is of interest. Unlike Assumption A where a Gaussian distribution is assumed, Assumption B is semiparametric, since it does not put constraints on higher-order moments, except that the fourth-order moment is bounded. Compared with existing asymptotic results for $T \to \infty$, Assumption B is weaker. Flury (1987) and Schott (1999) both assumed that $\{\mathbf{y}_{it}\}$ are normally distributed. Boik (2002) provided asymptotic results for non-normal data, but modeled eigenvalues as known smooth functions of parameters. Though Pepler et al. (2016) and Hallin et al. (2010) relaxed Assumptions B (c) and (d), the former work only focused on the case when $n = 2$ and the latter was for CPCA.

3. Estimation

In this section, we introduce our estimation procedure under two scenarios: (a) the number of the CPCs is known and (b) the number of the CPCs is unknown. When the number of CPCs in known, we prove that our proposed estimator of the common eigenvectors is consistent. When the number of CPCs is unknown, we provide a sequential testing procedure to find the number of CPCs.

3.1 *The number of the CPCs is known*

When the number of CPCs is known, we propose to estimate $\mathbf{\Gamma}$ in two steps: first getting p CPC candidates for $\mathbf{\Gamma}$, denoted as $\hat{\mathbf{\Gamma}}_{\text{candi}} \in \mathbb{R}^{p\times p}$, and then selecting k columns from $\hat{\mathbf{\Gamma}}_{\text{candi}}$ as $\hat{\mathbf{\Gamma}} \in \mathbb{R}^{p\times k}$.

In the first step, $\hat{\mathbf{\Gamma}}_{\text{candi}}$ is calculated as the eigenvectors of $\bar{\mathbf{S}} = \sum_{i=1}^{n} \mathbf{S}_i/n$, where $\mathbf{S}_i = \sum_{t=1}^{T} \mathbf{y}_{it}\mathbf{y}_{it}^{\top}/T$ is the sample covariance matrix of subject i. The columns of $\hat{\mathbf{\Gamma}}_{\text{candi}}$ are ordered in a way that the corresponding eigenvalue of each eigenvector is decreasing. We first define the consistency of an eigenvector estimator.

Definition 1 Let $\{\mathbf{x}_n\}$ denote a series of random vectors in $\mathbb{R}^p$ with ℓ_2-norm 1; that is $\|\mathbf{x}_n\|_2 = 1$ for all n. Let $\mathbf{x}$ be a vector in $\mathbb{R}^p$ such that $\|\mathbf{x}\|_2 = 1$. As $n \to \infty$, $\mathbf{x}_n$ converges to $\mathbf{x}$ if

$$|\langle \mathbf{x}_n, \mathbf{x}\rangle| \xrightarrow{P} 1,$$

where $\langle \cdot, \cdot \rangle$ is the inner product defined in $\mathbb{R}^p$ and $\xrightarrow{P}$ denotes convergence in probability.

Definition 2 As $n \to \infty$ (fixed T and p) or $T \to \infty$ (fixed n and p), $\hat{\boldsymbol{\gamma}}$ is a **consistent** estimator of $\boldsymbol{\gamma}$ if

$$|\langle \hat{\boldsymbol{\gamma}}, \boldsymbol{\gamma}\rangle| \xrightarrow{P} 1,$$

where $\|\hat{\boldsymbol{\gamma}}\|_2 = 1$ and $\|\boldsymbol{\gamma}\|_2 = 1$.

Under Definition 2, the following theorem shows that k out of p columns of $\hat{\mathbf{\Gamma}}_{\text{candi}}$ are consistent estimators of the columns of $\mathbf{\Gamma}$ as n or T goes to infinity.

THEOREM 3.1 Assume the PCPC model (2.1) holds.

1. Under Assumption A, for any column $\boldsymbol{\gamma}_j$ of $\mathbf{\Gamma}$ $(j = 1, \ldots, k)$, there exists a column of $\hat{\mathbf{\Gamma}}_{\text{candi}}$ that is consistent with $\boldsymbol{\gamma}_j$ as $n \to \infty$. Explicitly, let $\mathbf{e}_l \in \mathbb{R}^p$ denote a p-dimensional vector with the l-th entry one and rest zero, then, there exists $l(j) \in \{1, \ldots, p\}$, such that

$$|\langle \boldsymbol{\gamma}_j, \hat{\mathbf{\Gamma}}_{\text{candi}} \mathbf{e}_{l(j)} \rangle| \xrightarrow{P} 1.$$

2. Under Assumption B, for any column $\boldsymbol{\gamma}_j$ of $\mathbf{\Gamma}$ $(j = 1, \ldots, k)$, there exists a column of $\hat{\mathbf{\Gamma}}_{\text{candi}}$ that is consistent to $\boldsymbol{\gamma}_j$ as $T \to \infty$.

Theorem 3.1 implies that, by properly ordering the columns of $\hat{\mathbf{\Gamma}}_{\text{candi}}$, we can achieve that the j-th column of $\hat{\mathbf{\Gamma}}_{\text{candi}}$ converges in probability to $\boldsymbol{\gamma}_j$ for $j = 1, \ldots, k$. To find this ordering, we define a deviation from commonality metric for each column of $\hat{\mathbf{\Gamma}}_{\text{candi}}$:

$$\text{Err}\left(\{\mathbf{y}_{it}\}, \hat{\mathbf{\Gamma}}_{\text{candi}}, j\right) = \frac{1}{n(p-1)} \sum_{l=1, l \neq j}^{p} \frac{\sum_{i=1}^{n} (\hat{\boldsymbol{\gamma}}_j^\top \mathbf{S}_i \hat{\boldsymbol{\gamma}}_l)^2}{(\hat{\boldsymbol{\gamma}}_j^\top \bar{\mathbf{S}} \hat{\boldsymbol{\gamma}}_j)(\hat{\boldsymbol{\gamma}}_l^\top \bar{\mathbf{S}} \hat{\boldsymbol{\gamma}}_l)}, \tag{3.2}$$

where $\bar{\mathbf{S}} = \sum_i \mathbf{S}_i / n$ is the average sample covariance matrix and $\hat{\boldsymbol{\gamma}}_j$ is the j-th column of $\hat{\mathbf{\Gamma}}_{\text{candi}}$. On the right-hand side of Equation (3.2), the numerator captures the sum of squared (j, l) element in $\hat{\mathbf{\Gamma}}_{\text{candi}}^\top \mathbf{S}_i \hat{\mathbf{\Gamma}}_{\text{candi}}$ for $j \neq l$, which is close to zero if $\hat{\boldsymbol{\gamma}}_j$ is a CPC and large otherwise. The denominator is a normalizing term that eliminates the effect of magnitude difference in the eigenvalues. The following theorem shows that $\text{Err}\left(\{\mathbf{y}_{it}\}, \hat{\mathbf{\Gamma}}_{\text{candi}}, j\right)$ can be used to order the columns of $\hat{\mathbf{\Gamma}}_{\text{candi}}$.

THEOREM 3.2 For all $j \in \{1, \ldots, k\}$, let $\hat{\boldsymbol{\gamma}}_{j_n}$ be the estimate of $\boldsymbol{\gamma}_j$ in $\hat{\mathbf{\Gamma}}_{\text{candi}}$, where $j_n = \arg\max_{l \in \{1, \ldots, p\}} |\langle \hat{\boldsymbol{\gamma}}_l, \boldsymbol{\gamma}_j \rangle|$. Assume the PCPC model (2.1) holds.

1. Under Assumption A, as $n \to \infty$, we have

$$\mathrm{Err}\left(\{\mathbf{y}_{it}\}, \hat{\mathbf{\Gamma}}_{\mathrm{candi}}, j_n\right) \xrightarrow{P} \frac{1}{T}.$$

   In addition, let $L_n = \{1, \ldots, p\} \setminus \{1_n, \ldots, k_n\}$, then there exits a positive constant C $(0 < C < \infty)$ independent of n, such that, as $n \to \infty$,

$$\min_{l \in L_n} \mathrm{Err}\left(\{\mathbf{y}_{it}\}, \hat{\mathbf{\Gamma}}_{\mathrm{candi}}, l\right) \xrightarrow{P} \frac{1}{T} + C.$$

   Here $l \neq j_n$ for $j = 1, \ldots, k$.

2. Under Assumption B, as $T \to \infty$, we have

$$\mathrm{Err}\left(\{\mathbf{y}_{it}\}, \hat{\mathbf{\Gamma}}_{\mathrm{candi}}, j_n\right) \xrightarrow{a.s.} 0 \quad \text{and} \quad \min_{l \in L_n} \mathrm{Err}\left(\{\mathbf{y}_{it}\}, \hat{\mathbf{\Gamma}}_{\mathrm{candi}}, l\right) \xrightarrow{a.s.} \tilde{C},$$

   where $\tilde{C}$ is a positive constant $(0 < \tilde{C} < \infty)$ independent of T and $\xrightarrow{a.s.}$ denotes convergence almost surely.

In Theorem 3.2, the asymptotic results of $n \to \infty$ and $T \to \infty$ are different, which results from the different assumptions made in the two cases. When $n \to \infty$, the commonality metric is related to the fourth-order moment of $\mathbf{y}_{it}$, which yields a positive probability limit for a CPC estimate. If a column of $\hat{\mathbf{\Gamma}}_{\mathrm{candi}}$ is not an estimate of any CPC, then it is related to $\mathbf{\Psi}_i$ and the commonality metric is larger due to the variability of $\mathbf{\Psi}_i$. When $T \to \infty$, we have $\hat{\boldsymbol{\gamma}}_j^\top \mathbf{S}_i \hat{\boldsymbol{\gamma}}_l \xrightarrow{a.s.} 0$ for $l \neq j$ if and only if $\hat{\boldsymbol{\gamma}}_j$ is a CPC estimate, making the commonality metric converge to 0 only for a CPC estimate. Despite these differences, CPC estimates in both cases have the least deviation from commonality metric among all columns of $\hat{\mathbf{\Gamma}}_{\mathrm{candi}}$ asymptotically, which is essential for identifying CPC estimates from $\hat{\mathbf{\Gamma}}_{\mathrm{candi}}$.

Given Theorem 3.2, in practice, we can rank the columns of $\hat{\mathbf{\Gamma}}_{\mathrm{candi}}$ in an increasing order of $\mathrm{Err}\left(\{\mathbf{y}_{it}\}, \hat{\mathbf{\Gamma}}_{\mathrm{candi}}, j\right)$ for $j = 1, \ldots, p$ and consider the first k columns as $\hat{\mathbf{\Gamma}}$. When n and T are small, it is possible that some columns of $\hat{\mathbf{\Gamma}}$ have a large deviation from commonality metric and are not "close" to any CPC. This bias, however, will disappear as $n \to \infty$ or $T \to \infty$, as guaranteed by Theorems 3.1 and 3.2. The estimating procedure is summarized in Algorithm 1.

Algorithm 1 An algorithm to estimate CPCs in model (2.1) when k is known.

Input: A Data set $\{\mathbf{y}_{it}\}$, $t = 1, \ldots, T$, $i = 1, \ldots, n$, and $k \in \{1, \ldots, p-2, p\}$.

1. Calculate the sample covariance matrix $\mathbf{S}_i = \sum_{t=1}^{T} \mathbf{y}_{it}\mathbf{y}_{it}^\top / T$ for each i.

2. Perform eigen-decomposition on $\bar{\mathbf{S}} = \sum_{i=1}^{n} \mathbf{S}_i / n$ and obtain the estimated eigenvectors denoted as $\hat{\mathbf{\Gamma}}_{\text{candi}}$.

3. Reorder the columns of $\hat{\mathbf{\Gamma}}_{\text{candi}}$ such that $\text{Err}\left(\{\mathbf{y}_{it}\}, \hat{\mathbf{\Gamma}}_{\text{candi}}, j\right)$ is increasing in j, and let $\hat{\mathbf{\Gamma}}$ be the first k columns of $\hat{\mathbf{\Gamma}}_{\text{candi}}$.

Output: A $p \times k$ orthonormal matrix $\hat{\mathbf{\Gamma}}$.

With the number of CPCs given, we generalize the results of Flury (1987) in three directions. First, Algorithm 1 can consistently estimate a CPC as $n \to \infty$, a case Flury (1987) did not cover. Second, when $T \to \infty$, Theorems 3.1 and 3.2 relax the Gaussian assumption made by Flury (1987). Third, Theorems 3.1 and 3.2 guarantee the identification of the CPCs without making assumptions on the ranks of CPC-related eigenvalues.

3.2 *The number of CPCs is unknown*

Based on the idea of Schott (1999), we use a sequential hypothesis testing approach to find k. For $j = 0, 1, \ldots, p-2, p$, we sequentially perform the following testings

$$\mathrm{H}_{0,j} : k = j \quad \leftrightarrow \quad \mathrm{H}_{1,j} : k \geqslant j + 1.$$

Starting from $j = 0$, if $\mathrm{H}_{0,j}$ is rejected, then we proceed to test $\mathrm{H}_{0,j+1}$; otherwise we estimate $\hat{k} = j$. Before the first test, we order the columns of $\hat{\mathbf{\Gamma}}_{\text{candi}}$ such that $\text{Err}\left(\{\mathbf{y}_{it}\}, \hat{\mathbf{\Gamma}}_{\text{candi}}, j\right)$ is increasing in j. When testing the j-th hypothesis, we simulate the distribution of $\text{Err}\left(\{\mathbf{y}_{it}\}, \hat{\mathbf{\Gamma}}_{\text{candi}}, j+1\right)$ under $\mathrm{H}_{0,j}$, denoted as $\hat{F}_{j+1}$, and reject $H_{0,j}$ if $\text{Err}\left(\{\mathbf{y}_{it}\}, \hat{\mathbf{\Gamma}}_{\text{candi}}, j+1\right)$ is smaller than the α-quantile of $\hat{F}_{j+1}$. The logic of this rejection rule is that $\text{Err}\left(\{\mathbf{y}_{it}\}, \hat{\mathbf{\Gamma}}_{\text{candi}}, j+1\right)$ is small under

$\mathrm{H}_{1,j}$, but the α-quantile of $\hat{F}_{j+1}$ is generally large, since $\hat{\boldsymbol{\gamma}}_{j+1}$ is not a common eigenvector under the null hypothesis. Adjusting for multiple testing is unnecessary here, since the family-wise type I error is $\mathbb{P}(\hat{k} \geqslant k_0 + 1 | k = k_0) = \alpha$, if the truth is $k = k_0$.

Given $\boldsymbol{\Gamma}$ and $\{\lambda_{i1}, \ldots, \lambda_{ip}\}, i = 1, \ldots, n$, defined in the PCPC model (2.1), we calculate $\hat{F}_{j+1}$ by repeating the following steps for m times. In practice, we can approximate $\boldsymbol{\Gamma}$ using $\hat{\boldsymbol{\Gamma}}$ output in Algorithm 1 and estimate $\{\lambda_{i1}, \ldots, \lambda_{ip}\}$ by $\{\hat{\boldsymbol{\gamma}}_1^\top \mathbf{S}_i \hat{\boldsymbol{\gamma}}_1, \ldots, \hat{\boldsymbol{\gamma}}_p^\top \mathbf{S}_i \hat{\boldsymbol{\gamma}}_p\}$ for $i = 1, \ldots, n$, where $\hat{\boldsymbol{\gamma}}_j$ is the j-th column of $\hat{\boldsymbol{\Gamma}}_{\text{candi}}$.

1. For each $i = 1, \ldots, n$, independently and uniformly generate $\mathbf{R}_i^{(\text{sim})}$ from the sample space

$$\{\mathbf{R}_i^{(\text{sim})} \in \mathbb{R}^{p \times (p-j)} : \mathbf{R}_i^{(\text{sim})\top} \mathbf{R}_i^{(\text{sim})} = \mathbf{I}_{n-j}, \mathbf{R}_i^{(\text{sim})\top} \boldsymbol{\Gamma} = \mathbf{0}\}.$$

2. Construct $\boldsymbol{\Sigma}_i^{(\text{sim})} = (\boldsymbol{\Gamma}, \mathbf{R}_i^{(\text{sim})}) \operatorname{diag}\{\lambda_{i1}, \ldots, \lambda_{ip}\} (\boldsymbol{\Gamma}, \mathbf{R}_i^{(\text{sim})})^\top$. Generate $\mathbf{y}_{it}^{(\text{sim})}, t = 1, \ldots, T$, from multivariate Gaussian distribution with mean $\mathbf{0}$ and covariance $\boldsymbol{\Sigma}_i^{(\text{sim})}$.

3. Given the data set $\{\mathbf{y}_{it}^{(\text{sim})}\}$, calculate $\hat{\boldsymbol{\Gamma}}_{\text{candi}}^{(\text{sim})}$ as described in Algorithm 1 and output $\operatorname{Err}(\{\mathbf{y}_{it}^{(\text{sim})}\}, \hat{\boldsymbol{\Gamma}}_{\text{candi}}^{(\text{sim})}, j+1)$.

Then $\hat{F}_{j+1} = \sum_{l=1}^m \delta_l / m$, where δ_l denotes a point mass at $\operatorname{Err}(\{\mathbf{y}_{it}^{(\text{sim})}\}, \hat{\boldsymbol{\Gamma}}_{\text{candi}}^{(\text{sim})}, j+1)$ output by the l-th simulation. The following theorem shows that the type I error rate for each test is bounded by α under regularity assumptions.

Theorem 3.3 Assume $\{\mathbf{y}_{it} | \boldsymbol{\Sigma}_i\}$ follows a multivariate Gaussian distribution with mean $\mathbf{0}$ and covariance $\boldsymbol{\Sigma}_i$, and $\mathbf{R}_i$ follows a uniform distribution on its sample space defined in Section 2. Then under $\mathrm{H}_{0,j}$, as $m \to \infty$, $\hat{F}_{j+1}$ converges in distribution to the true distribution of $\operatorname{Err}(\{\mathbf{y}_{it}\}, \hat{\boldsymbol{\Gamma}}_{\text{candi}}, j+1)$ given $\boldsymbol{\Gamma}$ and $\{\lambda_{i1}, \ldots, \lambda_{ip}\}, i = 1, \ldots, n$.

The procedure of finding k and estimate $\boldsymbol{\Gamma}$ is described in Algorithm 2.

Algorithm 2 A two-step algorithm to estimate CPCs in model (2.1) when k is unknown.

Input: A Data set $\{\mathbf{y}_{it}\}$, $t = 1, \ldots, T_i$, $i = 1, \ldots, n$.

Step 1: Get candidates $\hat{\mathbf{\Gamma}}_{\text{candi}}$ for $\mathbf{\Gamma}$.

(i) Calculate the sample covariance matrix $\mathbf{S}_i = \sum_{t=1}^{\top} \mathbf{y}_{it}\mathbf{y}_{it}^{\top}/T$ for each i.

(ii) Perform eigen-decomposition on $\bar{\mathbf{S}} = \sum_{i=1}^{n} \mathbf{S}_i/n$ and obtain the estimated eigenvectors denoted as $\hat{\mathbf{\Gamma}}_{\text{candi}}$.

(iii) Reorder the columns of $\hat{\mathbf{\Gamma}}_{\text{candi}}$ such that $\text{Err}\left(\{\mathbf{y}_{it}\}, \hat{\mathbf{\Gamma}}_{\text{candi}}, j\right)$ is increasing in j.

Step 2: Identify $\hat{\mathbf{\Gamma}}$ from $\hat{\mathbf{\Gamma}}_{\text{candi}}$.

(i) Initialize $\hat{k} = 0$.

(ii) Test the hypothesis $\mathrm{H}_{0,\hat{k}} : k = \hat{k} \leftrightarrow \mathrm{H}_{1,\hat{k}} : k \geqslant \hat{k} + 1$ by a simulation test described in Section 3.2 with significance level $\alpha = 0.05$ and 1,000 simulations.

(iii) Based on the testing result:

- if $\mathrm{H}_{\hat{k}}$ is not rejected, return $\hat{k}$ and $\hat{\mathbf{\Gamma}}$ as the first $\hat{k}$ columns of $\hat{\mathbf{\Gamma}}_{\text{candi}}$;
- if $\hat{k} = p - 2$ and $H_{\hat{k}}$ is rejected, return $\hat{k} = p$ and $\hat{\mathbf{\Gamma}} = \hat{\mathbf{\Gamma}}_{\text{candi}}$;
- otherwise, increase $\hat{k}$ by 1 and repeat **Step 2** (ii).

4. Simulation study

In this section, we perform three simulation studies. The first confirms the asymptotic results given by Theorem 3.2. The second tests the performance of the algorithm for finding k. The last compares our proposed method with existing approaches under different scenarios.

4.1 *Design and data generating mechanism*

In the first simulation, we let $p = 20$ and $k = 10$, define $\lambda_j = e^{0.5(j-1)}$ for $j = 1, \ldots, p$, and assume that $\{\mathbf{y}_{it}\}$ follows a multivariate Gaussian distribution and CPCs rank randomly in each covariance matrix. For the study of the asymptotics as $n \to \infty$, we set $T = 50$ and $n = 50, 100, 500, 1000$; and for the study of the asymptotics as $T \to \infty$, we set $n = 50$ and $T = 50, 100, 500, 1000$. For each combination of n and T, we simulate data and compare the distribution of the k-th smallest commonality metric, which is the largest metric of CPC estimates, with the distribution of the $(k+1)$-th smallest commonality metric, which is the smallest metric of non-CPC estimates.

The second simulation has the same setting as the first one except that we fix $n = 100$, $T = 100$ and vary $k = 1, 10, 20$. For each k, we simulate data and run Algorithm 2 to get $\hat{k}$. We use $P(\hat{k} = k)$ to measure the performance of the algorithm for finding k.

The last simulation compares our proposed method (with or without k known) with Flury's method (Flury, 1987) and the PVD method (Crainiceanu et al., 2011) through two scenarios below. There are other partial information decomposition methods, such as JIVE and CIFE described in Section 2, but they do not have a unique CPC estimates, which makes the comparison with these methods via simulation infeasible.

Scenario 1: $\{\mathbf{y}_{it}\}$ follow Gaussian distribution. The CPC-associated eigenvalues rank randomly in each covariance matrix.

Scenario 2: $\{\mathbf{y}_{it}\}$ follow non-Gaussian distribution. CPCs are associated with largest eigenvalues in each covariance matrix.

For each of the above two scenarios, we consider two cases: $p = 20$ and $p = 100$, which represent small-scale and large-scale problem, respectively. When $p = 20$, we let $n = 100$, $T = 100$, $k = 10$ and define $\lambda_j = e^{0.5(j-1)}$ for $j = 1, \ldots, p$. When $p = 100$, we let $n = 1000$, $T = 1000$, $k = 20$ and define $\lambda_j = e^{0.1(j-1)}$ for $j = 1, \ldots, p$.

For each case in each scenario, we run Algorithm 1 if k is known and Algorithm 2 if k is unknown. For Flury's method, we run the algorithm given by Flury and Gautschi (1986) to estimate $\hat{\mathbf{\Gamma}}_{candi}$ and select k columns associated with the largest eigenvalues of $\bar{\mathbf{S}}$. For population covariance decomposition, we use the default setting, that is, first calculating the top k eigenvectors of $\mathbf{S}_i$ (denoted as $\mathbf{U}_i$) and then estimating $\mathbf{\Gamma}$ as the top k eigenvectors of $\mathbf{U} = [\mathbf{U}_1|, \ldots, |\mathbf{U}_n]$. To measure the performance of each method, we define $\sum_{j=1}^{k} \max |\boldsymbol{\gamma}_j^\top \hat{\mathbf{\Gamma}}|/k$ as the accuracy metric of $\hat{\mathbf{\Gamma}}$. This accuracy metric is in the range of $[0, 1]$ with larger values indicating better accuracy.

In simulation 1, simulation 2 and scenario 1 of simulation 3, we simulate the data as follows for 1000 replications. Given p, n, T, k and $\{\lambda_j\}_{j=1}^{p}$, we sample one $\mathbf{\Gamma}$ from the space $\{\mathbf{\Gamma} : \mathbf{\Gamma}^\top \mathbf{\Gamma} = \mathbf{I}_k\}$ as the common eigenvectors, and randomly partition $\{\lambda_j\}_{j=1}^{p}$ into two parts: one with k elements as the common eigenvalues (denoted as $\{\lambda_j^*\}_{j=1}^{k}$) and the other one consisting of $p-k$ elements (denoted as $\{\lambda_j^*\}_{j=k+1}^{p}$). For $i = 1, \ldots, n$, we independently sample λ_{ij} from a chi-squared distribution with degrees of freedom λ_j^* and construct $\mathbf{R}_i = \mathbf{U}_i \mathbf{D}_i \mathbf{U}_i^\top$, where $\mathbf{U}_i$ is an independent sample from the space $\{\mathbf{U} : \mathbf{U}^\top \mathbf{U} = \mathbf{I}_{n-k}, \mathbf{U}^\top \mathbf{\Gamma} = \mathbf{0}_{(n-k)\times k}\}$ and $\mathbf{D}_i = \mathrm{diag}\{\{\lambda_j^*\}_{j=k+1}^{p}\}$. Then we construct $\mathbf{\Sigma}_i = \mathbf{\Gamma}\, \mathrm{diag}\{\{\lambda_{ij}\}_{j=1}^{k}\} \mathbf{\Gamma}^\top + \mathbf{R}_i$ and $\{\mathbf{y}_{it}, t = 1, \ldots, T\}$ are independently sampled from $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_i)$. For scenario 2 of simulation 3, the data generating procedure is similar, except that $\{\lambda_j^*\}_{j=1}^{k} = \{\lambda_j\}_{j=p-k+1}^{p}$ is fixed, and $\mathbf{y}_{it} = \mathbf{\Sigma}_i^{-\frac{1}{2}} \tilde{\mathbf{y}}_{it}$, where $\{\tilde{\mathbf{y}}_{it}, t = 1, \ldots, T\}$ are independently sampled from a multivariate-gamma distribution with mean $\mathbf{0}$, variance $\mathbf{I}_p$ and skewness $10\mathbf{I}_p$.

4.2 *Simulation results*

Simulation results are summarized in Figure 2 and Tables 1 and 2 for simulations 1, 2 and 3 respectively.

Figure 2 shows that the deviation from commonality metric converges to its limit when T is fixed and $n \to \infty$ and when n is fixed and $T \to \infty$. This confirms the results of Theorem 3.2 and indicates that this metric can be used to distinguish CPC estimates and non-CPC estimates

when either n or T is large.

Table 1 displays the probability of finding the correct k under different settings using the algorithm described in Section 3.2. When data is Gaussian distributed, this algorithm can find k with high accuracy whenever k is small, medium or large. Under the non-Gaussian data generating distribution, however, this algorithm performs poorly when k is large. The reason of this result is twofold: first, this algorithm is based on simulations from Gaussian distributed data and is hence not appropriate for non-Gaussian data; second, when k is large, the number of null hypotheses to reject is large, which reduces the overall power. As a result, we recommend using this algorithm for Gaussian distributed data. If k is large, we may not find all CPCs, but the identified CPCs are still valid.

Table 2 gives the comparison of our proposed method with k known or unknown to Flury's method and PVD. In both scenarios, our proposed method outperforms Flury's method and PVD, even when the true number of CPCs is unknown. In Scenario 1, since both Flury's method and PVD assume CPCs are associated with the largest eigenvalues for each matrix, their accuracy is much lower than our proposed method. In Scenario 2, all four methods have modest accuracy, but our proposed method with k unknown, Flury's method and PVD have lower accuracy due to the non-Gaussian assumption. In contrast, our proposed method with k known remains highly accurate, since it is semiparametric.

Table 1. The probability of finding the correct k using the algorithm described in Section 3.2.

Data generating distribution	$k = 1$	$k = 10$	$k = 20$
Gaussian	0.94	0.89	1.00
Non-Gaussian	0.47	0.02	0.00

5. Task fMRI data example

We apply the proposed semiparametric PCPC method to the Human Connectome Project (HCP) motor-task fMRI data. The HCP project studies the brain connectome, both structural and func-

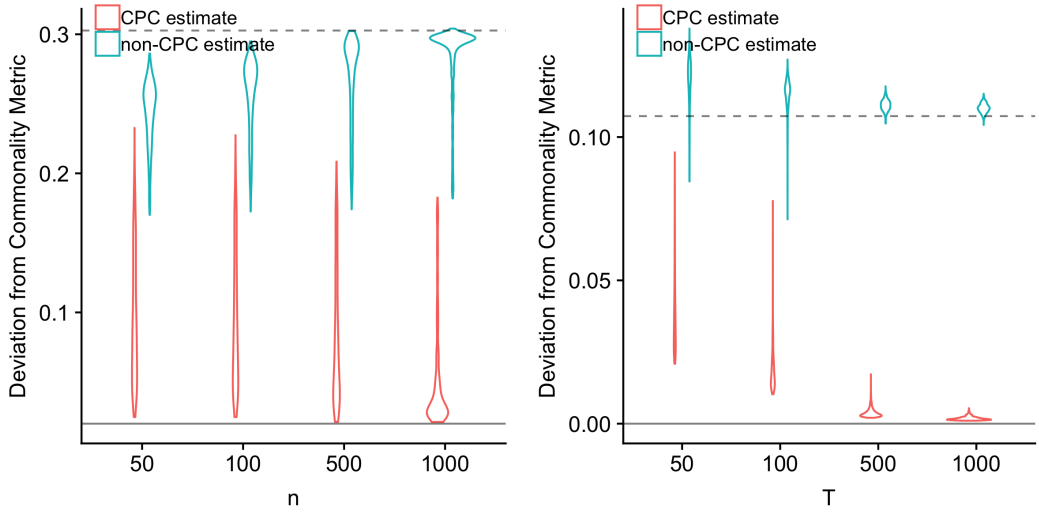

Figure 2. Distribution of the "Deviation from commonality" metric (3.2) as n (left panel) or T (right panel) goes to infinity for the last CPC estimate and the first non-CPC estimate. The solid line is the probability limit for the CPC estimate and the dashed line is the probability limit for the non-CPC estimate calculated from Theorem 3.2. The left panel demonstrates that the metric converges in probability to its limit, while the right panel shows that the metric converges almost surely.

Table 2. Accuracy of methods in estimating CPC under different scenarios. Semi-1: the proposed semi-parametric method with k known. Semi-2: the proposed semiparametric method with k unknown. Flury: the Flury's method by Flury (1987). PVD: population value decomposition by Crainiceanu et al. (2011).

		Semi-1	Semi-2	Flury	PVD
Scenario 1	$p = 20$	0.99	0.96	0.26	0.50
	$p = 100$	0.99	0.93	0.24	0.20
Scenario 2	$p = 20$	0.98	0.95	0.88	0.94
	$p = 100$	1.00	0.95	0.91	0.95

tional, of healthy adults. The data set includes $n = 136$ healthy young adults from the most recent S1200 release. Adapted from the experimental design in Buckner et al. (2011) and Thomas Yeo et al. (2011), the task fMRI consists of ten task blocks including two tongue movement blocks, four hand movement blocks (two left and two right) and four foot movement blocks (two left and two right), as well as three 15-second fixation blocks. In each movement task block, a three-second visual cue was first presented followed by a 12-second movement. Participants were instructed to follow the visual cue to either move their tongue, or tap their left/right fingers, or squeeze their left/right toes to map the corresponding motor areas. The tasks were randomly intermixed.

Once the ordering was fixed, the task onsets are nearly consistent across participants. The fMRI data were collected for $T = 284$ time points (with TR $= 0.72$ second) and preprocessed following the HCP minimal preprocessing pipeline (Glasser et al., 2013). We extracted blood-oxygen-level dependent (BOLD) signals from $p = 35$ functional brain regions in the sensorimotor network (Power et al., 2011) and averaged over voxels within the 5 mm radius. According to the Doornik-Hansen test for multivariate normality by Doornik and Hansen (2008), the BOLD signals are not Gaussian distributed (p-value 10^{-7}). Since Flury's method and PVD are not able to identify the CPCs associated with small eigenvalues and require prespecified k, we present the result from the proposed semiparametric method only. Results of Flury's method and PVD letting $k = p$ are given in the Supplementary Material, which differs from the results of the semiparametric method.

Among 35 CPC candidates, Algorithm 2 identifies 31 as CPCs, which explains 80% of the total variance of the average covariance matrix. Figure 1 in the Supplementary Material summarises the results of sequential testings. To explore the relationship between the identified CPCs and the motor tasks, we plot the average time course of each CPC estimate (i.e., $\sum_{i=1}^{n} \boldsymbol{\gamma}_j^\top \mathbf{y}_{it}/n$ for $j = 1, \ldots, 31$) and compare it with task time bins. We also visualize brain regions with loading magnitude greater than 0.1 in a brain map. As a result, at least ten of the identified CPCs are related with tasks (no statistical test is performed) and a list of identified brain networks are provided in Table 3. Figure 3 presents an example of task-related CPC (CPC 29). In Figure 3, the average time course suggests an brain network of right hand movement and left foot movement, which is confirmed by the brain map. In this component, brain areas associated with motor control of the right hand yield high negative loadings (blue regions on the left hemisphere of the brain in Figure 3); and regions associated with motor control of the left foot yield high positive loadings (red regions on the right hemisphere of the brain in Figure 3). The lateral separation of the brain in terms of the loading sign suggests that during these motor tasks, the associated left and right hemispheres are functionally negatively correlated. Figure 4 presents an example of the

CPC that is not related to the tasks. Even though the time course does not show a clear pattern, this CPC is concentrated in a region of the brain, which is modularized as a tongue region by Power et al. (2011). For the four components that are not identified as CPCs, two appear task-related and two do not. Since Algorithm 2 can be conservative when the true number of CPCs is large and the distribution is non-Gaussian based on simulation, it is possible that some of these four CPC candidates are actually CPC. For all 35 CPC candidates, the average time course and the brain maps are provided in the Supplementary Material.

Table 3. Task-related brain networks identified by Algorithm 2.

CPC No.	Variance explained	Brain network
1	1.1%	Left hand
5	1.5%	Feet
11	2.5%	Left hand
17	1.3%	Right hand
20	2.0%	Right foot
22	2.8%	Feet, right hand
23	2.6%	Feet
28	2.1%	Left hand
29	5.5%	Right hand, left foot
31	20.0%	Hands, feet, mouth

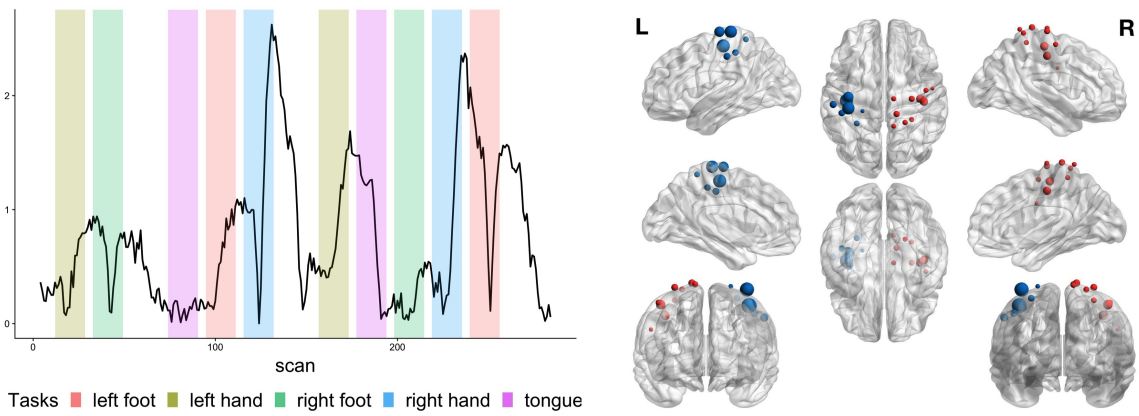

Figure 3. Average time course (left panel) and brain regions (right panel) of CPC 29. In the left panel, each bin represents the time period of a task. In the right panel, each node is a brain region, with size standing for the absolute loading and color representing the sign of the loading (blue for negative and red for positive). Brain regions with absolute loading smaller than 0.1 are not shown in the figure.

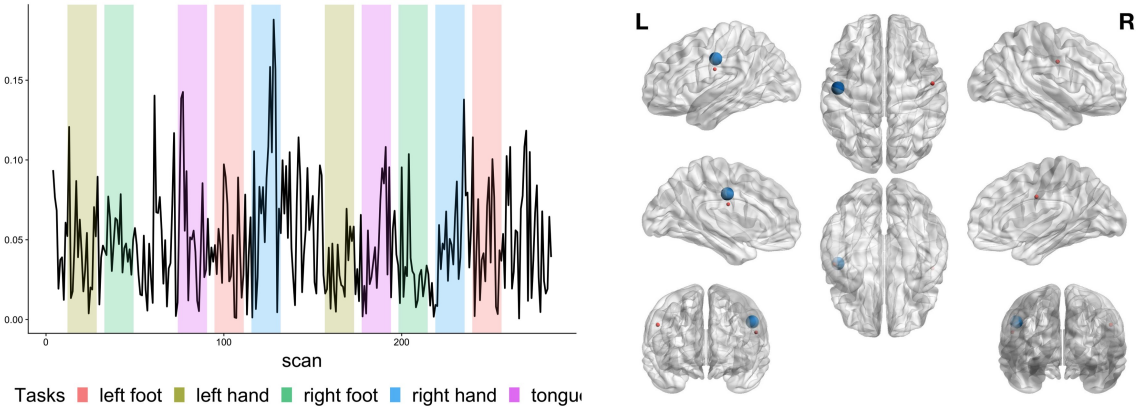

Figure 4. Average time course (left panel) and brain regions (right panel) of CPC 25. In the left panel, each bin represents the time period of a task. In the right panel, each node is a brain region, with size standing for the absolute loading and color representing the sign of the loading (blue for negative and red for positive). Brain regions with absolute loading smaller than 0.1 are not shown in the figure.

6. Discussion

In this paper, we propose a semiparametric PCPCA model and provide algorithms to identify CPCs with or without knowing the true number of CPCs. Furthermore, we prove the asymptotic consistency of our proposed estimators, even when the data generating distribution is non-Gaussian. In simulation studies, our estimator consistently outperforms Flury's method and PVD and shows high accuracy if the number of CPCs is known. Applied to the motor-task fMRI data, our method identifies meaningful brain networks which match the current findings.

In PCPCA, a CPC may not be associated with the largest eigenvalues across all covariance matrices. For this reason, our proposed method allows for arbitrary association between CPC and eigenvalues, which makes the model more flexible. One challenge resulting from this flexibility is to find k, the number of CPCs, since the signal of CPCs can be weak or inseparable from non-common principal components. Our proposed algorithm for finding k performs well under Gaussian distribution, but can be conservative if the underlying distribution is non-Gaussian or k is large. Furthermore, sequential hypothesis testing usually requires huge computational

resources and can be slow for high-dimensional matrices. Hence an efficient and robust method for finding k will be one future direction.

Our proposed method, as well as the literature, assumes p, the dimension of covariance matrices, is fixed. One exciting future direction could be finding solutions to handle data with large p but small n and T.

References

Boik, R. J. (2002). Spectral models for covariance matrices. *Biometrika* **89,** 159–182.

Browne, R. P. and McNicholas, P. D. (2014). Estimating common principal components in high dimensions. *Advances in Data Analysis and Classification* **8,** 217–226.

Buckner, R. L., Krienen, F. M., Castellanos, A., Diaz, J. C., and Yeo, B. T. T. (2011). The organization of the human cerebellum estimated by intrinsic functional connectivity. *Journal of Neurophysiology* **106,** 2322–2345.

Crainiceanu, C. M., Caffo, B. S., Luo, S., Zipunnikov, V. M., and Punjabi, N. M. (2011). Population value decomposition, a framework for the analysis of image populations. *Journal of the American Statistical Association* **106,** 775–790.

Doornik, J. A. and Hansen, H. (2008). An omnibus test for univariate and multivariate normality. *Oxford Bulletin of Economics and Statistics* **70,** 927–939.

Flury, B. and Gautschi, W. (1986). An algorithm for simultaneous orthogonal transformation of several positive definite symmetric matrices to nearly diagonal form. *SIAM Journal on Scientific and Statistical Computing* **7,** 169–184.

Flury, B. K. (1987). Two generalizations of the common principal component model. *Biometrika* **74,** 59–69.

Flury, B. N. (1984). Common principal components in k groups. *Journal of the American Statistical Association* **79,** 892–898.

Franks, A. and Hoff, P. (2016). Shared subspace models for multi-group covariance estimation. *arXiv* .

Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J. R., Essen, D. C. V., and Jenkinson, M. (2013). The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage* **80,** 105 – 124.

Goyal, A., Pérignon, C., and Villa, C. (2008). How common are common return factors across the NYSE and Nasdaq? *Journal of Financial Economics* **90,** 252 – 271.

Gu, F. (2016). Analysis of correlation matrices using scale-invariant common principal component models and a hierarchy of relationships between correlation matrices. *Structural Equation Modeling: A Multidisciplinary Journal* **23,** 819–826.

Hadjipantelis, P. Z., Aston, J. A. D., Müller, H. G., and Evans, J. P. (2015). Unifying amplitude and phase analysis: A compositional data approach to functional multivariate mixed-effects modeling of Mandarin Chinese. *Journal of the American Statistical Association* **110,** 545–559.

Hallin, M., Paindaveine, D., and Verdebout, T. (2010). Optimal rank-based testing for principal components. *The Annals of Statistics* **38,** 3245–3299.

Hoff, P. D. (2009). A hierarchical eigenmodel for pooled covariance estimation. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **71,** 971–992.

Krzanowski, W. J. (1984). Principal component analysis in the presence of group structure. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **33,** 164–168.

Lock, E. F., Hoadley, K. A., Marron, J. S., and Nobel, A. B. (2013). Joint and individual variation explained (JIVE) for integrated analysis of multiple data types. *The Annals of Applied Statistics* **7,** 523–542.

Pepler, P. T., Uys, D. W., and Nel, D. G. (2016). A comparison of some methods for the selection of a common eigenvector model for the covariance matrices of two groups. *Communications in Statistics - Simulation and Computation* **45,** 2917–2936.

Power, J. D., Cohen, A. L., Nelson, S. M., Wig, G. S., Barnes, K. A., Church, J. A., Vogel, A. C., Laumann, T. O., Miezin, F. M., and Schlaggar, B. L. (2011). Functional network organization of the human brain. *Neuron* **72,** 665–678.

Schott, J. (1999). Partial common principal component subspaces. *Biometrika* **86,** 899–908.

Thomas Yeo, B. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., Roffman, J. L., Smoller, J. W., Zöllei, L., Polimeni, J. R., Fischl, B., Liu, H., and Buckner, R. L. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology* **106,** 1125–1165.

Wang, W., Zhang, X., and Li, L. (2019). Common reducing subspace model and network alternation analysis. *Biometrics* **0,**.

Xu, X., Chen, C. Y.-H., and Härdle, W. K. (2019). Dynamic credit default swap curves in a network topology. *Quantitative Finance* **0,** 1–22.

Ye, J., Dobson, S., and McKeever, S. (2012). Situation identification techniques in pervasive computing: A review. *Pervasive and Mobile Computing* **8,** 36 – 66.

Zhou, G., Cichocki, A., Zhang, Y., and Mandic, D. P. (2016). Group component analysis for multiblock data: Common and individual feature extraction. *IEEE Transactions on Neural Networks and Learning Systems* **27,** 2426–2439.