*Title Page w/ ALL Author Contact Info.

Functional Multiple-set Canonical Correlation Analysis

Heungsun Hwang[1], Kwanghee Jung[1], Yoshio Takane[1], and Todd S. Woodward[2]

1. McGill University

2. University of British Columbia and British Columbia Mental Health and Addiction Research Institute

May 19, 2011

We thank the Associate Editor and three anonymous reviewers for their constructive comments that helped improve the quality of this article. Requests for reprints should be sent to: Heungsun Hwang, Department of Psychology, McGill University, 1205 Dr. Penfield Avenue, Montreal, QC, H3A 1B1, Canada. Email: heungsun.hwang@mcgill.ca.

Functional Multiple-set Canonical Correlation Analysis

Abstract

We propose functional multiple-set canonical correlation analysis for exploring associations among multiple sets of functions. The proposed method includes functional canonical correlation analysis as a special case when only two sets of functions are considered. As in classical multiple-set canonical correlation analysis, computationally, the method solves a matrix eigen-analysis problem through the adoption of a basis expansion approach to approximating data and weight functions. We apply the proposed method to functional magnetic resonance imaging (fMRI) data to identify networks of neural activity that are commonly activated across subjects while carrying out a working memory task.

Keywords: Functional data, multiple-set canonical correlation analysis, functional canonical correlation analysis, functional magnetic resonance imaging data.

1. Introduction

In psychology and many other fields of inquiry, data are increasingly being collected in the form of smooth curves or functions over time, space, and other continua. A few examples include motor control data (e.g., Mattar & Ostry, 2010; Olshen, Biden, Wyatt, & Sutherland, 1989), emotion speech production data (Lee, Bresch, & Narayanan, 2006), musical cognition/perception data (e.g., Almansa & Delicado, 2009; Vines, Krumhansl, Wanderley, & Levitin, 2006; Vines, Nuzzo, & Levitin, 2005), eye-tracking data (e.g., Jackson & Sirois, 2009), and brain imaging data (Tian, 2010).

Functional canonical correlation analysis is a tool for exploring the associations between a pair of such functional data (Leurgans, Moyeed, & Silverman, 1993; Ramsay & Silverman, 2005, Chapter 11). For example, Leurgans et al. (1993) shows how variation in the knee angle functions of children is related to their hip angle functions through a gait cycle. In functional canonical correlation analysis, a series of weight functions are obtained for each set of functional data such that the resultant components or weighted composites of the functional data are mutually orthogonal to each other within the same set of components, while maximally correlated with the different set of components.

Despite its usefulness, functional canonical correlation analysis is limited to the analysis of two functional datasets. In practice, it is not uncommon to collect more than two sets of functional data concurrently and investigate their interdependencies. For example, in functional neuroimaging studies, nearly all investigations collect data on multiple subjects, and it is increasingly common to collect data in multiple modalities

such as functional magnetic resonance imaging (fMRI), electroencephalography (EEG), and magnetoencephalography (MEG) (e.g., Correa, Li, Adali, & Calhoun, 2009).

In this paper, we propose an extension of functional canonical correlation analysis to the analysis of more than two sets of functional data. We call this proposed method functional multiple-set canonical correlation analysis. As the name explicitly suggests, the proposed method represents a functional version of classical multiple-set canonical correlation analysis (Carroll, 1968; Horst, 1961; Meredith, 1964). Classical multiple-set canonical correlation analysis is a multivariate technique for describing interrelationships among multiple sets of variables. Multiple-set canonical correlation analysis subsumes many extant statistical methods as special cases including canonical correlation analysis, discriminant analysis, principal components analysis, and simple/ multiple correspondence analysis (e.g., Gifi, 1990). In addition, multiple-set canonical correlation analysis has been recognized as a way of integrating the data observed from different sources such as subjects, stimuli, areas, etc. (e.g., Correa, Eichele, Adali, Li, & Calhoun, 2010; Takane & Oshima-Takane, 2002).

Technically, the proposed method extends multiple-set canonical correlation analysis to analyze associations among multiple sets of functional data. Accordingly, it consists of functional canonical correlation analysis as a special case when only two sets of functional data are considered. This method also subsumes functional principal components analysis (Rice & Silverman, 1991; Ramsay & Silverman, 2005, chapter 8) as a special case when each set of functional data reduces to a discrete value sampled from a single function. As will be shown in the application section, the proposed method can also be a useful tool for the fusion of functional data acquired from numerous sources.

The paper is organized as follows. Section 2 reviews classical multiple-set canonical correlation analysis, focusing on its problem formulation and computation. Section 3 briefly describes functional canonical correlation analysis for two sets of functions. Section 4 provides the technical underpinnings of the proposed functional multiple-set canonical correlation analysis. Section 5 presents the empirical application of the proposed method to brain imaging data. In particular, the method is employed to discover brain regions that are activated jointly across subjects during an fMRI experiment. The final section summarizes the implications of the proposed method and discusses directions for future research.

2. Multiple-set Canonical Correlation Analysis

In this section, we describe the problem formulation and computation of classical multiple-set canonical correlation analysis (MCCA) for multivariate variables. Note that we define MCCA in scalar notation to make clearer the connection between classical MCCA and its functional version proposed in Section 4.

The central problem of MCCA is to construct a series of components, called canonical variates, for each set of multivariate variables in such a way that the components maximize the association or homogeneity among them, or equivalently they minimize the departure from homogeneity among them (e.g., Gifi, 1990, p. 81). By evaluating the resultant (typically, low dimensional) canonical variates, MCCA can be of use in summarizing how multiple sets of variables are related to one another in a succinct manner.

For the moment, we focus on the leading canonical variates for K sets of variables (i.e., the number of dimensions = 1). Let x_{ipk} denote the pth variable value of the ith case in the kth dataset ($i = 1, \cdots, I$; $p = 1, \cdots, P_k$; $k = 1, \cdots, K$). Let w_{pk} denote the (canonical) weight for the pth variable in the kth set. Let $\eta_{ik} = \sum_{p=1}^{P_k} x_{ipk} w_{pk}$ denote the ith score of the kth leading canonical variate. We assume that x_{ipk} is variable-wise centered. The problem of MCCA can be formulated as maximization of the following optimization criterion.

$$\phi_1(w_{pk}) = \sum_{i=1}^{I} \left(\sum_{k=1}^{K} \eta_{ik} \right)^2 = \sum_{i=1}^{I} \left(\sum_{k=1}^{K} \sum_{p}^{P_k} x_{ipk} w_{pk} \right)^2, \tag{1}$$

with respect to w_{pk}, subject to the constraint $\sum_{i=1}^{N} \eta_{ik}^2 = 1$ (Carroll, 1968).

Let f_i denote the ith object score, characterizing the homogeneity among K leading canonical variates. Then, alternatively, the MCCA problem can be formulated to minimize the departure from the homogeneity among K canonical variates as follows.

$$\begin{aligned} \phi_2(f_i, w_{pk}) &= \sum_{k=1}^{K} \sum_{i=1}^{I} (f_i - \eta_{ik})^2 \\ &= \sum_{k=1}^{K} \sum_{i=1}^{I} (f_i - \sum_{p=1}^{P_k} x_{ipk} w_{pk})^2, \end{aligned} \tag{2}$$

with respect to f_i and w_{pk}, subject to the constraint $\sum_{i=1}^{I} f_i^2 = 1$. This criterion is also called the criterion for homogeneity analysis or K-set canonical correlation analysis (Gifi, 1990, Chapter 3; Yanai, 1998). As shown in (2), the object scores f_i's are indicative of the most highly correlated or homogenous one-dimensional representation of K datasets. Both formulations result in essentially the same solutions (Gifi, 1990; Takane, Hwang, & Abdi,

2008). In this paper, thus, we focus on (1) for the development of the proposed functional version of MCCA.

We used scalar notation in (1). From a computational perspective, however, it is more convenient to express this criterion in matrix notation because its optimization reduces to a matrix eigen-analysis problem. Moreover, as will be discussed in Section 4, optimization of the proposed functional version of MCCA also reduces to a matrix eigen-analysis problem similar to that in MCCA.

We can express criterion (1) in matrix notation, as follows.

$$\phi_1(\mathbf{w}) = \mathrm{tr}(\mathbf{w}'\mathbf{X}'\mathbf{X}\mathbf{w}), \tag{3}$$

subject to the constraint $\mathbf{w}'\mathbf{\Phi}\mathbf{w} = 1$, where $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_K]$ is an I by P row block matrix ($P = \sum_{k=1}^{K} P_k$), and $\mathbf{w} = [\mathbf{w}_1; \mathbf{w}_2; \cdots; \mathbf{w}_K]$ is a P by 1 column vector stacking $\mathbf{w}_k = [w_{1k}, w_{2k}, \cdots, w_{P_k k}]'$ one below another, and $\mathbf{\Phi}$ is a block diagonal matrix consisting of $\mathbf{X}_k'\mathbf{X}_k$ as the kth diagonal block. This criterion can be re-expressed as

$$\mathrm{tr}(\mathbf{m}'\mathbf{\Phi}^{-1/2}\mathbf{X}'\mathbf{X}\mathbf{\Phi}^{-1/2}\mathbf{m}), \tag{4}$$

subject to $\mathbf{m}'\mathbf{m} = 1$, where $\mathbf{m} = \mathbf{\Phi}^{1/2}\mathbf{w}$. Thus, maximizing (4) with respect to $\mathbf{m}$ is equivalent to obtaining the following eigenvalue decomposition (EVD):

$$\mathbf{\Phi}^{-1/2}\mathbf{X}'\mathbf{X}\mathbf{\Phi}^{-1/2} = \mathbf{\Sigma}\mathbf{\Delta}\mathbf{\Sigma}', \tag{5}$$

where $\mathbf{\Sigma}$ is a matrix of eigenvectors such that $\mathbf{\Sigma}'\mathbf{\Sigma} = \mathbf{I}$, and $\mathbf{\Delta}$ is a diagonal matrix consisting of eigenvalues in descending order. Then, $\mathbf{m} = \mathbf{\sigma}$, where $\mathbf{\sigma}$ is the first eigenvector in $\mathbf{\Sigma}$. In turn, $\mathbf{w}$ is obtained by

$$\hat{\mathbf{w}} = \mathbf{\Phi}^{-1/2}\mathbf{\sigma}. \tag{6}$$

In addition, an I by 1 vector of object scores, denoted by $\mathbf{f} = [f_1, f_2, \cdots, f_I]'$, can be calculated by

$$\mathbf{f} = \mathbf{X}\mathbf{\Phi}^{-1/2}\boldsymbol{\sigma} / \sqrt{\delta}\,, \tag{7}$$

where δ is the first eigenvalue of $\mathbf{\Delta}$ (e.g., Takane, Hwang, & Abdi, 2008).

When we consider multi-dimensional canonical variates (i.e., the number of dimensions > 1), orthonormalization constraints are imposed on the MCCA criterion, such as $\mathbf{W}'\mathbf{\Phi}\mathbf{W} = \mathbf{I}$, where $\mathbf{W} = [\mathbf{W}_1; \cdots; \mathbf{W}_K]$ is a P by L column block matrix of weights stacking $\mathbf{W}_k = [\mathbf{w}_{1k}, \cdots, \mathbf{w}_{Lk}]$ one below another, and L indicates the number of dimensions. Computationally, it is straightforward to obtain the weights $\mathbf{W}$ for multi-dimensional canonical variates: $\mathbf{W}$ are obtained based on the corresponding eigenvectors in $\mathbf{\Sigma}$ in (6).

3. Functional Canonical Correlation Analysis

In functional data analysis, the data observed are I curves, each of which has discrete observations or records of y_{ij} ($i = 1, \cdots, I$; $j = 1, \cdots, J$). We assume that a smooth function, denoted by $z_i(t)$, underlies each observed curve, and that the smooth function is available for argument t in some finite interval T ($t \in T$), over which all integrals are to be taken. The argument t can be time, space, wavelength or other continua. We generally assume a relationship between the observations of a curve and the underlying smooth function as follows.

$$y_{ij} = z_i(t_j) + e_{ij}\,, \tag{8}$$

where e_{ij} is an error (Ramsay & Silverman, 2005, p. 40).

Assume that we have two sets of such I curves, and the ith curve of the first and second sets has observed values of y_{ij1} and y_{ij2}, respectively. Let z_{i1} and z_{i2} denote smooth functions underlying an individual data curve in each set, where the observations y_{ij1} and y_{ij2} are associated with functional values $z_{i1}(t_{j1})$ and $z_{i2}(t_{j2})$, respectively.

In functional canonical correlation analysis for two sets of smooth functions (Leurgans, Moyeed, & Silverman, 1993; Ramsay & Silverman, 2005, Chapter 11), we aim to estimate a pair of weight functions successively in such a way that the same pair of canonical variates is maximally correlated with each other, while uncorrelated with different pairs. Again, let us focus on the leading canonical variates for now. Let $\beta_1(t_1)$ and $\beta_2(t_2)$ denote the (canonical) weight functions for the leading canonical variates of z_{i1} and z_{i2}, respectively. Let $\eta_{i1} = \int z_{i1}(t_1)\beta_1(t_1)dt_1$ and $\eta_{i2} = \int z_{i2}(t_2)\beta_2(t_2)dt_2$ denote the ith score of the leading canonical variate for each set. The problem of obtaining a pair of weights functions for the leading canonical variates is equivalent to maximizing the following criterion.

$$\phi_3(\beta_1, \beta_2) = \frac{\mathrm{cov}(\eta_{i1}, \eta_{i2})^2}{\left[\sum_{i=1}^{I}\eta_{i1}^2 + \lambda\int\left[D^2\beta_1(t_1)\right]^2 dt_1\right]\left[\sum_{i=1}^{I}\eta_{i2}^2 + \lambda\int\left[D^2\beta_2(t_2)\right]^2 dt_2\right]}, \tag{9}$$

where $\mathrm{cov}(\eta_{i1}, \eta_{i2})$ indicates the covariance between two canonical variates, λ indicates a non-negative smoothing parameter that controls for the degree of roughness in each weight function, and D^{m} denotes the derivative of order m.

This criterion is equivalent to maximizing the squared covariance between the two canonical variates subject to the constraints

$$\sum_{i=1}^{I}\eta_{i1}^2 + \lambda\int\left[D^2\beta_1(t_1)\right]^2 dt_1 = \sum_{i=1}^{I}\eta_{i2}^2 + \lambda\int\left[D^2\beta_2(t_2)\right]^2 dt_2 = 1, \tag{10}$$

(Ramsay & Silverman, 2005, p. 205). This way of combining regularization into each constraint is comparable to that proposed in regularized canonical correlation analysis (Vinod, 1976). In regularized canonical correlation analysis, the ridge type of regularization (Hoerl & Kennard, 1970) is used for two sets of multivariate data, whereas in functional canonical correlation analysis, a roughness penalty (e.g., $\int \left[D^2 \beta_1(t_1)\right]^2 dt_1$) is adopted as a regularization term that is the sum of the integrated squared second derivative of each weight function. In functional data analysis, a function's roughness is usually assessed by its curvature or second derivative (Ramsay & Silverman, 2005, p. 84). When a function is highly variable, its roughness penalty value will be large. Conversely, when a function is nearly linear, the penalty value will be close to zero.

In functional canonical correlation analysis, such regularization is necessary to control for smoothness of the estimated weight functions and to deal with the sample covariance matrices of functional data that can be singular. For more details, refer to Leurgans, Moyeed, and Silverman (1993) and Ramsay and Silverman (2005, p. 209).

4. Functional Multiple-set Canonical Correlation Analysis

4.1. Defining MCCA for functional data

We now consider K sets of data curves, and the ith curve of each set has observed values of y_{ijk}. Let z_{ik} denote a smooth function underlying an individual data curve in each set, which is available for argument t_k ($t_k \in T_k$). We also concentrate on the leading canonical variates for the moment. Let β_k denote the weight function for z_{ik} with values $\beta_k(t_k)$. Let $\eta_{ik} = \int \beta_k(t_k) z_{ik}(t_k) dt_k$ denote the ith score of the kth leading canonical variate. Then,

functional multiple-set canonical correlation analysis (FMCCA) aims to maximize the following criterion.

$$\begin{aligned}\phi_4(\beta_k) &= \sum_{i=1}^{I}\left(\sum_{k=1}^{K}\eta_{ik}\right)^2 + \lambda\sum_{k=1}^{K}\int[D^2\beta_k(t_k)]^2\,dt_k \\ &= \sum_{i=1}^{I}\left(\sum_{k=1}^{K}\int z_{ik}(t_k)\beta_k(t_k)dt_k\right)^2 + \lambda\sum_{k=1}^{K}\int[D^2\beta_k(t_k)]^2\,dt_k,\end{aligned} \tag{11}$$

with respect to β_k, subject to the constraint $\sum_{i=1}^{I}\eta_{ik}^2 + \lambda\int\left[D^2\beta_k(t_k)\right]^2 dt_k = 1$.

The proposed method can be viewed as a regularized version of MCCA, which uses a roughness penalty as a regularization term for each set of functional data. This method can subsume Takane et al.'s (2008) regularized multiple-set canonical correlation analysis as a special case where ridge-type regularization is adopted for each set of multivariate variables. Specifically, in the case of classical MCCA where K sets of multivariate variables are considered (i.e., x_{ipk} in Section 2), the weight function β_k is replaced by a vector of weights w_{pk} and the canonical variates score is by $\eta_{ik} = \sum_{p=1}^{P_k} x_{ipk} w_{pk}$.

If we replace the roughness penalty term by $\sum_{k=1}^{K}\sum_{p=1}^{P_k} w_{pk}^2$, (11) becomes equivalent to the homogeneity criterion for regularized multiple-set canonical correlation analysis. Further, (11) reduces to (1) if no regularization is involved or equivalently when $\lambda = 0$.

4.2. Computational Considerations

In functional data analysis, it is typical to convert a functional optimization problem such as (11) to an approximately equivalent matrix eigen-analysis for multivariate variable values. We shall adopt the same strategy for optimization of (11). In particular, we utilize

a basis function expansion approach in which each function is approximated as a linear combination of known basis functions (Hastie, Tibshirani, & Friedman, 2009; Ramsay & Silverman, 2005). The data and weight functions in (11) can be generally expressed as basis function systems as follows.

$$z_{ik}(t_k) = \sum_{s=1}^{S_k} c_{isk}\psi_{sk}(t_k) = \mathbf{c}_{ik}'\boldsymbol{\psi}_k(t_k), \text{ and } \beta_k(t_k) = \sum_{s=1}^{S_k} \theta_{sk}\xi_{sk}(t_k) = \boldsymbol{\xi}_k(t_k)'\boldsymbol{\theta}_k, \qquad (12)$$

where $\boldsymbol{\psi}_k(t_k)$ and $\boldsymbol{\xi}_k(t_k)$ are S_k by 1 vectors of basis functions ($\psi_{sk}(t_k)$ and $\xi_{sk}(t_k)$) for the data and weight functions, respectively; and $\mathbf{c}_{ik}$ and $\boldsymbol{\theta}_k$ are S_k by 1 vectors of coefficients for these basis functions, consisting of c_{isk}, and θ_{sk} as elements. We may further express the simultaneous expansion of each set of all I data functions as

$$\mathbf{z}_k(t_k) = \mathbf{C}_k\boldsymbol{\psi}_k(t_k), \qquad (13)$$

where $\mathbf{C}_k$ is an I by S_k matrix of coefficients for an I by 1 vector-valued data function $\mathbf{z}_k(t_k)$.

Various basis functions are available to represent functions. For example, a Fourier series is widely used for approximating very stable, periodic functions. B-splines are considered useful for non-periodic functions involving local fluctuations. Refer to Ramsay and Silverman (2005, Chapter 3) for descriptions of different basis function systems.

Based on the basis function approximations, (11) can be re-expressed as

$$\phi_4(\beta_k) = \sum_{i=1}^{I}\left(\sum_{k=1}^{K}\int z_{ik}(t_k)\beta_k(t_k)dt_k\right)^2 + \lambda\sum_{k=1}^{K}\int[D^2\beta_k(t_k)]^2 dt_k$$

$$= \sum_{k=1}^{K} \mathrm{SS}\left(\mathbf{C}_k\left[\int \boldsymbol{\psi}_k(t_k)\boldsymbol{\xi}_k(t_k)'\,dt_k\right]\boldsymbol{\theta}_k\right) + \lambda\sum_{k=1}^{K}\boldsymbol{\theta}_k'\left[\int D^2\boldsymbol{\xi}_k(t_k)D^2\boldsymbol{\xi}_k(t_k)'\,dt_k\right]\boldsymbol{\theta}_k$$

$$= \sum_{k=1}^{K} \mathrm{SS}\left(\mathbf{C}_k\mathbf{Q}_k\boldsymbol{\theta}_k\right) + \lambda\sum_{k=1}^{K}\boldsymbol{\theta}_k'\mathbf{K}_k\boldsymbol{\theta}_k$$

$$= \sum_{k=1}^{K} \mathrm{SS}\left(\mathbf{A}_k\boldsymbol{\theta}_k\right) + \lambda\sum_{k=1}^{K}\boldsymbol{\theta}_k'\mathbf{K}_k\boldsymbol{\theta}_k, \qquad (14)$$

subject to $\mathrm{SS}\left(\mathbf{A}_k\boldsymbol{\theta}_k\right) + \lambda\boldsymbol{\theta}_k'\mathbf{K}_k\boldsymbol{\theta}_k = 1$, where SS(**H**) = $\mathrm{tr}\left(\mathbf{H}'\mathbf{H}\right)$, $\mathbf{Q}_k = \int \boldsymbol{\psi}_k(t_k)\boldsymbol{\xi}_k(t_k)'dt_k$, $\mathbf{A}_k = \mathbf{C}_k\mathbf{Q}_k$, and $\mathbf{K}_k = \int D^2\boldsymbol{\xi}_k(t_k)D^2\boldsymbol{\xi}_k(t_k)'dt_k$. In (14), $\mathbf{C}_k$ is typically estimated a priori and considered fixed. Based on (8), we obtain the least squares estimates of $\mathbf{C}_k$ as follows.

$$\mathbf{C}_k = \mathbf{y}_{jk}\boldsymbol{\psi}_k(t_k)'\left(\boldsymbol{\psi}_k(t_k)\boldsymbol{\psi}_k(t_k)'\right)^{-1}, \qquad (15)$$

where $\mathbf{y}_{jk}$ is an I by 1 vector of discrete records taken at the jth measurement point in the kth set (e.g., Ramsay & Silverman, 2005, Chapter 4). This least squares estimation implies that S_k should not exceed I. The computation of $\mathbf{Q}_k$ and $\mathbf{K}_k$ involves the integrals of products of a derivative D^m of basis functions. These integrals may be computed by using a numerical integration method such as the trapezoidal rule or Romberg integration (e.g., Press, Teukolsky, Vetterling, & Flannery, 1999). We used Ramsay's computer programs for this computation, which are publicly available in R and MATLAB (http://www.psych.mcgill.ca/misc/fda/) (also refer to Ramsay, Hooker, & Graves, 2009).

Let $\mathbf{A} = [\mathbf{A}_1, \mathbf{A}_2, \cdots, \mathbf{A}_K]$ denote an N by S row block matrix ($S = \sum_{k=1}^{K} S_k$),

$\boldsymbol{\theta} = [\boldsymbol{\theta}_1; \boldsymbol{\theta}_2; \cdots; \boldsymbol{\theta}_K]$ denote an *S* by 1 column vector, stacking $\boldsymbol{\theta}_k$ one below another, and $\boldsymbol{\Xi}$ denote a block diagonal matrix consisting of $\mathbf{K}_k$ as the *k*th diagonal block. Then, maximizing (14) is equivalent to maximizing

$$\mathrm{tr}(\boldsymbol{\theta}'(\mathbf{A}'\mathbf{A} + \lambda\boldsymbol{\Xi})\boldsymbol{\theta}), \tag{16}$$

subject to $\boldsymbol{\theta}'(\mathbf{D}_A + \lambda\boldsymbol{\Xi})\boldsymbol{\theta} = 1$, where $\mathbf{D}_A$ is a block diagonal matrix consisting of $\mathbf{A}_k'\mathbf{A}_k$ as the *k*th block (Takane, Hwang, & Abdi, 2008). This maximization problem with respect to $\boldsymbol{\theta}_k$ can be solved by EVD as shown in (5) with $\mathbf{X}'\mathbf{X} = \mathbf{A}'\mathbf{A} + \lambda\boldsymbol{\Xi}$ and $\boldsymbol{\Phi} = \mathbf{D}_A + \lambda\boldsymbol{\Xi}$. Once $\boldsymbol{\theta}_k$ is estimated, the corresponding weight function $\beta_k(t_k)$ can be obtained from (12). In addition, an *I* by 1 vector of object scores **f** can be calculated in a manner similar to (7). We recapitulate this estimation procedure in the Appendix. We can also estimate weight functions for multi-dimensional canonical variates in the proposed method, as described in Section 2.

We need to specify the value of λ prior to the above estimation procedure. As λ becomes large, a greater penalty is imposed on the roughness or variability of the estimated weight function, thereby leading the function to be more linear. Conversely, when λ becomes zero, no penalty term is imposed so that the estimated function tends to be close to the weight values obtained from classical MCCA.

We may determine the value of λ subjectively (Ramsay & Silverman, 2005, p. 206). At the same time, we may choose the value of λ in an automatic manner based on *G*-fold cross validation (Hastie et al., 2001, p. 214). In this method, the entire set of data is divided into *G* subsets. One subset is used as a test set, while the remaining subsets are as a training set for parameter estimation. The resultant parameter estimates are applied to the test set in order to calculate a prediction error. This procedure is repeated *G* times,

changing test and training sets systematically. Then, the cross-validation estimate of prediction error is calculated over all G test sets.

Specifically, let $\mathbf{f}^{*(g)}$ denote a vector of object scores of the gth test set, calculated based on the weight functions estimated from the remaining training set ($g = 1, \cdots, G$). Let $\mathbf{f}^{(g)}$ denote the gth subset of object scores which are calculated based on the entire functional data. We then calculate a cross-validation estimate of prediction error as follows.

$$\varepsilon(\lambda) = \frac{1}{G}\sum_{g=1}^{G} \mathrm{SS}\left(\mathbf{f}^{(g)} - \mathbf{f}^{*(g)}\right). \tag{17}$$

We repeat these procedures over a range of the values of λ. The value of λ associated with the minimum $\varepsilon(\lambda)$ is be chosen as the final one.

5. An Empirical Application to Functional Neuroimaging Data

In this section, we apply the proposed method to functional neuroimaging data to demonstrate its empirical usefulness. The present example uses a subset of the functional magnetic resonance imaging (fMRI) data originally collected for an investigation of working memory impairment in schizophrenia (Cairo, Woodward, & Ngan, 2006). fMRI records signal variation in blood-oxygen level dependent (BOLD) signal, which is correlated with signal variation in blood flow. The basic element of spatial measurement in fMRI is referred to as a voxel, which is, for the data analyzed in the current study, a 4 × 4 × 4 mm cube of imaged neural matter. BOLD signal changes are recorded continuously over scans (or time points) in every voxel in the brain. These measurements

of signal changes, called time courses, can be considered smooth functions of time (Tian, 2010).

5.1. Example Data

We provide a summary of the data analyzed here. Refer to Cairo et al. (2006) for full details regarding experimental conditions, the nature of sample, and data acquisition. In this example, we used four healthy subjects who performed a verbal working memory task under four different memory load conditions while undergoing fMRI. During a single trial of this task, the subjects viewed a string of 2, 4, 6 or 8 different uppercase letters for 4 seconds (encoding), which they were instructed to remember over a 6 second delay (delay). Following the delay, a single lowercase letter was shown for 1 second. Subjects were asked to decide whether this letter had been included in the preceding letter string (probe). The probe stage was followed by an inter-trial interval of 3, 4, or 5 seconds in duration. Each subject was presented with two stimulus runs, 9 occurrences of each memory load (2, 4, 6 or 8 letters) per run, producing a total of 72 working memory trials ($2 \times 9 \times 4$). Each stimulus run consisted of 214 scans of the entire brain. The duration of each brain scan, called time resolution (TR), was three seconds. There was no time interval between scans. Thus, a single stimulus run took 642 seconds (3×214). The timing of stimulus presentation was identical for all subjects. The BOLD signals in 23,621 voxels of the whole brain were extracted from each of the 214 scans collected from each subject.

Accordingly, we had four sets of functional data, each of which contained 23,621 cases, representing voxels, and 214 columns, representing scans measured for each

subject. The BOLD signal was realigned, spatially normalized and smoothed prior to analysis using Statistical Parametric Mapping (SPM2) (Wellcome Institute of Cognitive Neurology, London, UK, http://www.fil.ion.ucl.ac.uk/spm).

To relate the resultant canonical variates back to experimental conditions, we constructed a design matrix that contained models for expected BOLD increases in response to the timing of stimulus presentations. The design matrix was built on the finite impulse response (FIR) model (Goutte, Nielsen, & Hansen, 2000) for estimating the average change in BOLD signal at specific time points. The time points coded by the FIR model corresponded to the 1^{st} to 8^{th} repetition times following stimulus presentation, referred to as peristimulus time. The columns of the design matrix model BOLD signal change in peristimulus time for the four load conditions (i.e., 8 repetition times × 4 load conditions). Thus, the rows of the design matrix must match the number of scans (214). The value 1 was placed in rows in the design matrix for which BOLD signal amplitude was to be modeled, and 0 was in the other rows. The TR for these data was 3 seconds; therefore, the BOLD signal was modeled for each condition separately over a 24-second window in peristimulus time. The encoding, delay and probe stages were covered by this 24-second window for each of four load conditions (2, 4, 6, and 8 letters). Refer to Metzak et al. (2010, 2011) for more details with respect to the construction of the design matrix with extension to multiple subjects.

5.2. Analyses and Results

The aim of our analysis was to integrate signal variation in four subjects' brain voxels into highly-correlated, low-dimensional representations and to identify regions of the

brain which might form networks that were positively or negatively activated among the subjects who performed the same working memory tasks.

We used the B-spline basis system to approximate both data and weight functions for all datasets, considering their non-periodic signal changes over scans. The B-spline basis consists of piecewise polynomials of prescribed order for subintervals of T_k which are separated by so-called breakpoints or knots. Polynomials in adjacent subintervals are constrained to join smoothly at the breakpoint of the subintervals. Thus, in using the B-spline basis, it is necessary to specify the number of knots and the order of polynomials. For this example, we set the number of knots at 72 so as to place at least two subintervals for each load condition. Moreover, the polynomial order selected was 4, i.e., a cubic polynomial, which is one of the most widely used orders in practice (Hastie, Tibshirani, & Friedman, 2009, p. 120). We used Ramsay's computer programs (available at http://www.psych.mcgill.ca/misc/fda/) to construct the B-spline basis under the prescribed numbers of knots and polynomial order.

We applied five-fold cross validation for selecting the value of λ. Figure 1 displays the cross-validated prediction error estimates $\varepsilon(\lambda)$ against the common logarithms of different values of λ, varying from 10^1 to 10^{10} by a factor of 10 (i.e., $\log_{10}(\lambda) = \{1, 2, \cdots, 10\}$). As shown in the figure, the minimum value of $\varepsilon(\lambda)$ was achieved at $\lambda = 10^2$. Thus, we chose $\lambda = 10^2$ for the proposed method.

Insert Figure 1 about here

As with other fMRI analyses, it is crucial to examine whether or not solutions obtained from the proposed method are directly relevant to experimental manipulations. To interpret the canonical variates obtained from the proposed method in the context of the present experimental conditions, we calculated so-called predictor weights (Hunter & Takane, 2002) for each dataset. The predictor weights are equivalent to the regression coefficients obtained by regressing each weight function $\beta_k(t_k)$ on the design matrix aforementioned. These predictor weights indicate the contribution of a specific time point on the temporal variation in the functional networks represented by particular canonical variates over scans, thereby showing whether the networks are associated with the hemodynamic response function (HRF) shapes expected from the experimental design. For example, in the present experiment, an HRF peak was expected in response to the 4-second encoding stage (between 8-15 seconds after trial onset) for each memory-load condition (Metzak et al., 2010, 2011).

Figure 2-A presents the estimated weight functions for the leading canonical variates of the four subjects. The weight function shows an overall pattern of fluctuation over 214 scans of the neural network represented by the leading canonical variate for each subject. Despite differences in their time-course fluctuations, these weight functions appear to vary in a relatively similar manner. The mean correlation among the four leading canonical variates is .61. Figure 2-B displays the mean predictor weight values (averaged over subjects) that represent the response of a functional network to the memory task at different memory-load conditions. Inspection of these predictor weights suggests that the network represented by the leading canonical variates shows activation changes similar to the HRF shapes expected from the experimental design (e.g., peak

activation between 8-15 seconds after trial onset in response to an encoding stage). This suggests that the network associated with the leading canonical variates involves voxel activations that are related to the experimental stimulus presentations.

Figure 3-A exhibits the estimated weight functions for the second-dimensional canonical variates. These weights functions appear to change more distinctively across the four subjects, compared to those for the leading canonical variates. The mean correlation among these canonical variates is .49. Figure 3-B displays the mean predictor weight values (averaged over subjects) for each memory-load condition. As shown in this figure, the predictor weights do not match expected HRF shapes. Thus, the network represented by the second-dimensional canonical variates is less likely to be activating in response to the stimulus timing.

Although they are not presented here to preserve space, we have also examined other subsequent canonical variates (i.e., $L > 2$). Similarly to the second-dimensional case, no functional networks associated with the higher-dimensional canonical variates involved voxel activation that was sensitive to the experimental design.

Insert Figures 2 and 3 about here

Thus, we concentrated only on the object scores of voxels (f_i's) obtained based on the leading canonical variates in order to identify neural regions that were commonly activated among the four subjects while they performed the working memory task. Figure 4 exhibits five slice images constructed from these leading (voxel) object scores. The far-right sagittal brain image indicates which slices of the brain the first four images

represent. All the images present dominant 5% of the object scores mapped onto a structural brain image template, where positive scores are in red and yellow and negative scores are in blue and green. The images of these object scores represent neural networks that were activated in all four subjects during the experiment, suggesting that these regions were likely to be functionally connected across the subjects. The neural networks associated with the leading canonical variates comprise the elements of both task-positive and task-negative networks (Fox, Snyder, Vincent, Corbetta, Van Essen, & Raichle, 2005). The task-positive network is dominated by increased activation in dorsal anterior cingulate cortex/supplementary motor area, left dorso-lateral prefrontal cortex (DLPFC), left insula, bilateral inferior frontal cortices (including Broca's area), and bilateral inferior parietal cortices. This network of regions is thought to activate in response to the memory task. The task-negative network is a system of functionally connected brain regions that are thought to be suppressed during the task. Here this is characterized by decreased activation in medial superior frontal cortex, bilateral middle/inferior temporal cortices, ventral anterior cingulate cortex, right fusiform gyrus, posterior cingulate cortex, and bilateral angular gyri.

Insert Figure 4 about here

For comparison purposes, we also applied classical MCCA to the same data, considering the data multivariate. Figures 5 and 6 display the estimated weight values for the first- and second-dimensional canonical variates, respectively. Due to their severe local fluctuations over scans, it is difficult to describe the time-dependent patterns of the

weights values in a clean manner. More importantly, the mean predictor weight values calculated from these canonical variates do not vary in a fashion similar to the expected HRF shapes, as also shown in the figures. This suggests that the networks represented by the first- and second-dimensional canonical variates obtained from classical MCCA involve voxel activation irrelevant to the experimental stimulus presentations. This was the case in subsequent higher-dimensional canonical variates, although again they are not provided here to conserve space.

Insert Figures 5 and 6 about here

---

6. Concluding Remarks

We proposed an extension of functional canonical correlation analysis to the analysis of more than two sets of functional data. This method represents a regularized version of multiple-set canonical correlation analysis, which adopts a roughness penalty as a regularization term for each set of functional data. The proposed method has proved useful in investigating which brain regions were activated during an fMRI experiment on verbal working memory. Specifically, the method was able to produce experimental condition-specific canonical variates of signal changes over scans. The object scores of voxels obtained by integrating these canonical variates reflected neural networks that were activated commonly in multiple subjects while performing the experiment. Conversely, classical MCCA resulted in canonical variates that were not directly attributed to the working memory experimental conditions. This suggests that MCCA

showed the association among several sets of BOLD signal changes derived from any sources of brain activity other than the experimental conditions. Thus, it is unlikely to be worthwhile to interpret the results of MCCA from substantive perspectives.

We may further refine and extend the proposed method to improve its data-analytic capability and applicability. For example, we may incorporate linear constraints into the proposed method for more elaborate analyses. In the present application, we applied the proposed method to the original data, and then related the results of the method to a design matrix that took experimental conditions into account. Conversely, as an initial stage of analysis, we may regard such a design matrix as linear constraints and decompose the data into different parts based on the design matrix (e.g., Metzak et al., 2010). In the next stage, the proposed method can be applied only to the portion of the data explained by the design matrix. This may result in a solution that is more directly relevant to experimental conditions. Takane and Shibayama (1991) developed a comprehensive approach to imposing linear constraints on the row and column sides of a data matrix (also see Hwang & Takane, 2002; Takane & Hwang, 2002; Takane & Hunter, 2001). We may adopt a similar strategy for the proposed method.

In addition, we may integrate some cluster analysis into the proposed method to identify heterogeneous subgroups of canonical variates, which involve distinctive patterns of weight functions. The capturing of such group-level heterogeneity has been an issue of theoretical and empirical importance in various fields including psychology (e.g., Wedel & Kamakura, 1998). All of these possibilities warrant future theoretical and empirical work.

Appendix: A summary of the FMCCA estimation procedure

Given basis functions for data and weight functions, $\boldsymbol{\psi}_k(t_k)$ and $\boldsymbol{\xi}_k(t_k)$, respectively, and the value of λ, the estimation procedure of FMCCA, under $L = 1$, involves the following steps.

Step 1: We compute $\mathbf{Q}_k = \int \boldsymbol{\psi}_k(t_k)\boldsymbol{\xi}_k(t_k)'dt_k$ and $\mathbf{K}_k = \int D^2\boldsymbol{\xi}_k(t_k)D^2\boldsymbol{\xi}_k(t_k)'dt_k$.

Step 2: We estimate $\mathbf{C}_k$ by

$$\mathbf{C}_k = \mathbf{y}_{jk}\boldsymbol{\psi}_k(t_k)'\left(\boldsymbol{\psi}_k(t_k)\boldsymbol{\psi}_k(t_k)'\right)^{-1}, \tag{A.1}$$

where $\mathbf{y}_{jk}$ is an I by 1 vector of discrete observations taken at the jth measurement point in the kth set. We subsequently compute $\mathbf{A}_k = \mathbf{C}_k\mathbf{Q}_k$.

Step 3: We construct $\mathbf{A} = \left[\mathbf{A}_1, \mathbf{A}_2, \cdots, \mathbf{A}_K\right]$ and $\boldsymbol{\Phi} = \mathbf{D}_A + \lambda\boldsymbol{\Xi}$, where $\mathbf{D}_A$ and $\boldsymbol{\Xi}$ are block diagonal matrices consisting of $\mathbf{A}_k'\mathbf{A}_k$ and $\mathbf{K}_k$ as the kth diagonal block, respectively.

Step 4: We obtain the following eigenvalue decomposition

$$\boldsymbol{\Phi}^{-1/2}(\mathbf{A}'\mathbf{A} + \lambda\boldsymbol{\Xi})\boldsymbol{\Phi}^{-1/2} = \boldsymbol{\Sigma}\boldsymbol{\Delta}\boldsymbol{\Sigma}', \tag{A.2}$$

where $\boldsymbol{\Sigma}$ is a matrix of eigenvectors, and $\boldsymbol{\Delta}$ is a diagonal matrix of eigenvalues.

We obtain $\boldsymbol{\theta} = \left[\boldsymbol{\theta}_1; \boldsymbol{\theta}_2; \cdots; \boldsymbol{\theta}_K\right]$ by $\boldsymbol{\theta} = \boldsymbol{\Phi}^{-1/2}\boldsymbol{\sigma}$, where $\boldsymbol{\sigma}$ is the first eigenvector in $\boldsymbol{\Sigma}$. We then obtain each element of the weight function for the kth set, $\beta_k(t_k)$, by $\beta_k(t_k) = \boldsymbol{\xi}_k(t_k)'\boldsymbol{\theta}_k$. We calculate a vector of object scores, $\mathbf{f}$, by $\mathbf{f} = \mathbf{A}\boldsymbol{\Phi}^{-1/2}\boldsymbol{\sigma}/\sqrt{\delta}$, where δ is the first eigenvalue of $\boldsymbol{\Delta}$ (refer to Equations (16) and (17) in Takane at al., 2008).

References

Almansa, J., & Delicado, P. (2009). Analysing musical performance through functional data analysis: rhythmic structure in Schumann's Traumerei. *Connection Science*, 21, 207-225.

Cairo, T. A., Woodward, T. S., & Ngan, E. T. C. (2006). Decreased encoding efficiency in schizophrenia. *Biological Psychiatry*, 59(8), 740-746.

Carroll, J. D. (1968). A generalization of canonical correlation analysis to three or more sets of variables. *Proceedings of the 76th Annual Convention of the American Psychological Association*, 227-228.

Correa, N. M., Eichele, T., Adali, T., Li, Y., & Calhoun, V. D. (2010). Multi-set canonical correlation analysis for the fusion of concurrent single trial ERP and functional MRI. *Neuroimage*, 50, 1438-1445.

Correa, N. M., Li, Y., Adali, T., & Calhoun, V. D. (2009). Fusion of fMRI, sMRI, and EEG data using canonical correlation analysis. *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, 385-388.

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., & Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences*, 102, 9673-9678.

Goutte, C., Nielsen, F. A., & Hansen, L. K. (2000). Modeling the hemodynamic response in fMRI using smooth FIR filters. *IEEE Transactions on Medical Imaging*, 19, 1188-1201.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd Edition)*. New York: Springer.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12, 55-67

Horst, P. (1961). Generalized canonical correlations and their applications to experimental data. *Journal of Clinical Psychology*, 17, 331-347.

Hunter, M. A. & Takane, Y. (2002). Constrained principal component analysis: Various applications. *Journal of Educational and Behavioral Statistics*, 27, 105-145.

Hwang, H., & Takane, Y. (2002). Generalized constrained multiple correspondence analysis. *Psychometrika*, 67, 211-224.

Jackson, I., & Sirois, S. (2009). Infant cognition: going full factorial with pupil dilation. *Developmental Science*, 12, 670 – 679.

Lee, S., Bresch, E., & Narayanan, S. (2006). An exploratory study of emotional speech production using functional data analysis techniques. *Proceedings of the 7th International Seminar on Speech Production*.

Leurgans, S. E., Moyeed, R. A., & Silverman, B. W. (1993). Canonical correlation analysis when data are curves. *Journal of the Royal Statistical Society* B, 55, 725-740.

Mattar A. A. G., & Ostry, D. J. (2010). Generalization of dynamics learning across changes in movement amplitude. *The Journal of Neurophysiology*. 104, 426-438.

Meredith, W. (1964). Rotation to achieve factorial invariance. *Psychometrika*, 29, 187-206.

Metzak, P. D., Feredoes, E., Takane Y., Wang, L., Weinstein, S., Cairo, T., Ngan, E.T.C, & Woodward, T. S. (2010). Constrained principal component analysis reveals functionally connected load-dependent networks involved in multiple stages of working memory. *Human Brain Mapping,* DOI: 10.1002/hbm.21072.

Metzak, P.D., Riley, J., Wang, L., Whitman, J.C., Ngan, E.T.C., & Woodward, T.S. (2011). Decreased efficiency of task-positive and task-negative networks during working memory in schizophrenia. *Schizophrenia Bulletin,* DOI: 10-1093/schbul/sbq154.

Olshen, R.A., Biden, E. N. Wyatt, M. P., & Sutherland, D. H. (1989). Gait analysis and the bootstrap. *Annals of Statistics*, 17, 1419-1440.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (1999). *Numerical Recipes in C. The Art of Scientific Computing (2nd Edition)*. Cambridge University Press.

Ramsay, J. O., Hooker, G., & Graves, S. (2009). *Functional Data Analysis with R and Matlab*. New York: Springer.

Ramsay, J. O., & Silverman, B. W. (2005). *Functional Data Analysis (2nd Edition*). New York: Springer.

Rice, J. A., & Silverman, B. W. (1991). Estimating the mean and covariance structure non-parametrically when the data are curves. *The Journal of the Royal Statistical Society B*, 53, 233-243.

Takane, Y., & Hunter, M. A. (2001). Constrained principal component analysis: A comprehensive theory. *Applicable Algebra in Engineering, Communication, and Computing*, 12, 391-419.

Takane, Y., & Hwang, H. (2002). Generalized constrained canonical correlation analysis. *Multivariate Behavioral Research*, 37, 163-195.

Takane, Y., Hwang, H., & Abdi, H. (2008). Regularized multiple-set canonical correlation analysis. *Psychometrika*, 73, 753-775.

Takane, Y., & Oshima-Takane, Y. (2002). Nonlinear generalized canonical correlation analysis by neural network models. In S. Nishisato, Y. Baba, H. Bozdogan, and K. Kanefuji (Eds.), *Measurement and Multivariate Analysis* (pp. 183-190). Tokyo: Springer Verlag.

Takane, Y., & Shibayama, T. (1991). Principal component analysis with external information on both subjects and variables. *Psychometrika*, 56, 97-120

Tian, T. S. (2010). Functional data analysis in brain imaging studies. *Frontiers in Quantitative Psychology and Measurement*.

Vines, B. W., Krumhansl, C. L., Wanderley, M. M., & Levitin, D. J. (2006). Cross-modal interactions in the perception of musical performance. *Cognition*, 101, 80-113.

Vines, B. W., Nuzzo, R. L., & Levitin, D. J. (2005). Quantifying and analyzing musical dynamics: Differential calculus, physics and functional data techniques. *Music Perception*, 23(2), 137-152.

Vinod, H. D. (1976). Canonical ridge and econometrics of joint production. *Journal of Econometrics*, 4, 47-166.

Wedel, M., & Kamakura, W. A. (1998). *Market Segmentation: Conceptual and Methodological Foundations*. Boston: Kluwer Academic Publishers.

Yanai, H. (1998). Generalized canonical correlation analysis with linear constraints. In C. Hayashi, N. Ohsumi, K. Yajima, Y. Tanaka, H.-H. Bock, & Y. Baba (Eds.). *Data Science, Classification, and Related Methods* (pp. 539-546). Tokyo: Springer-

Figure 1. The cross-validation prediction error values $\varepsilon(\lambda)$ against the common logarithms of different values of the smoothing parameter λ.

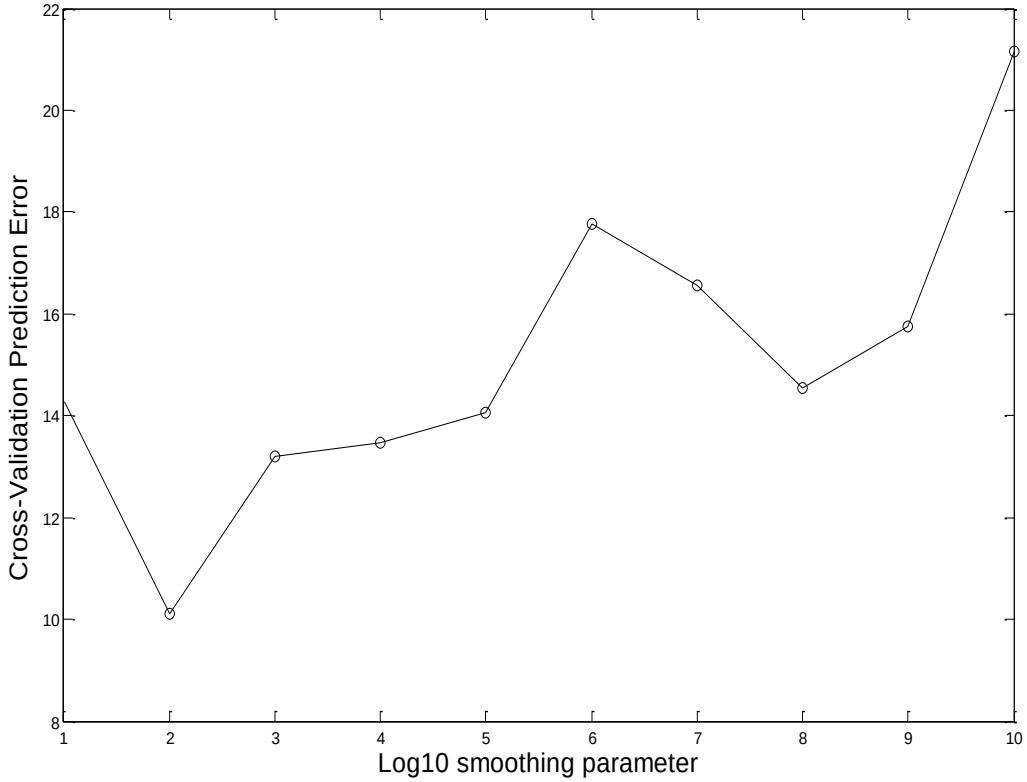
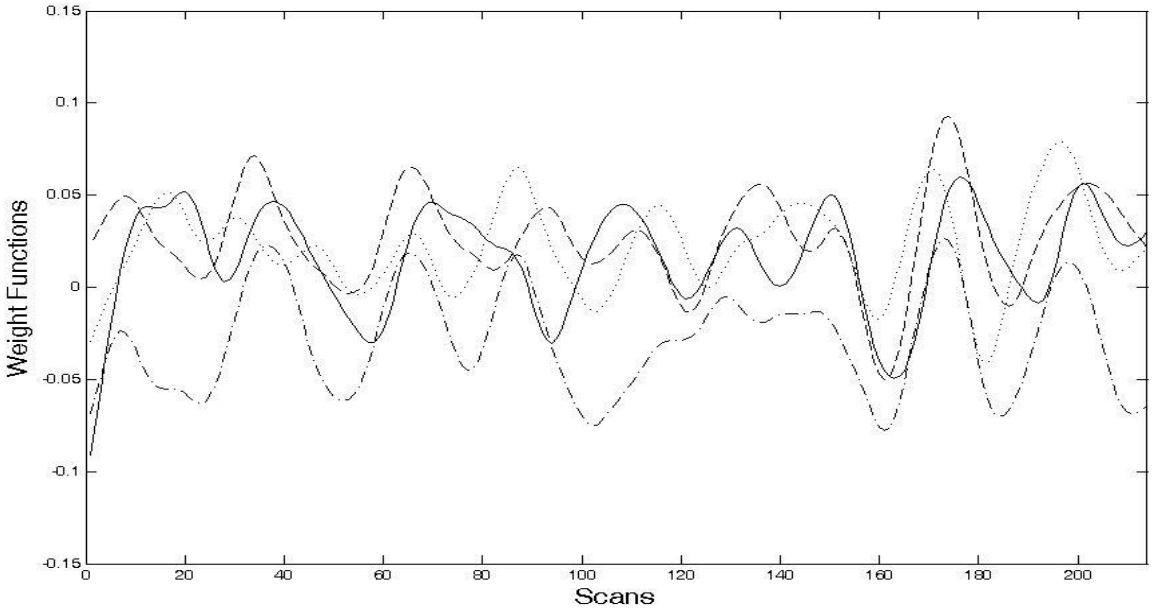


Figure 2. (A) The estimated weight functions for the first-dimensional canonical variates of four subjects obtained from the proposed method (… = subject 1, - - - = subject 2, ____ = subject 3, and __ . __ = subject 4), and (B) the mean predictor weights (averaged over subjects) for each load condition across the first-dimensional canonical variates (● = 2 letters, ■ = 4 letters, ▲= 6 letters, and * = 8 letters). The predictor weights at the first time point are adjusted to zero, and all other values scaled accordingly.

(A)



(B)

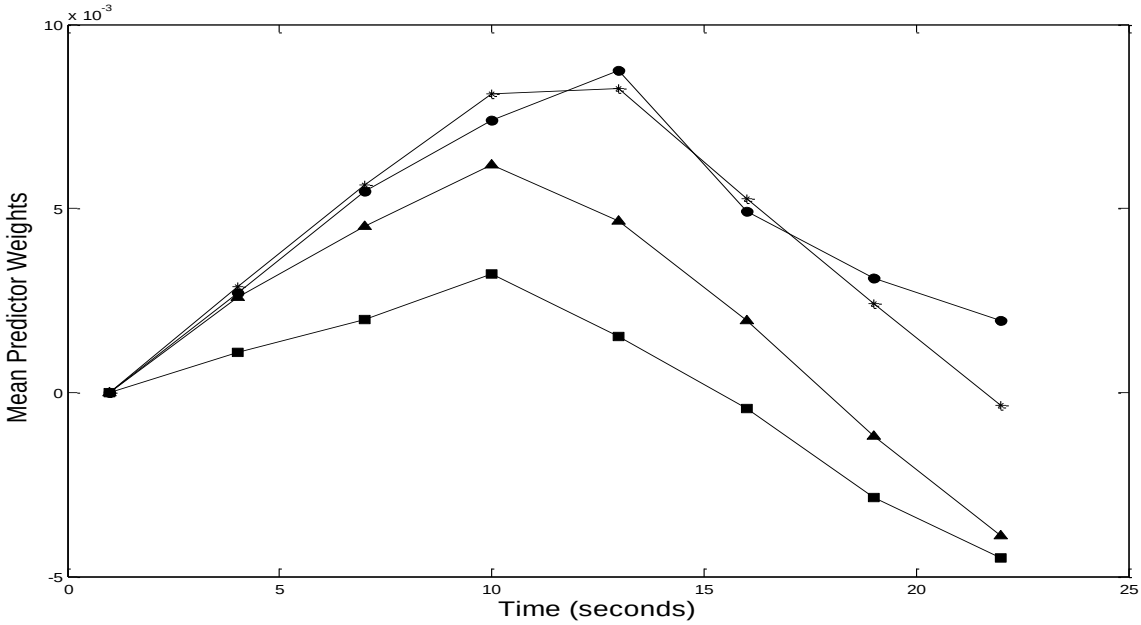
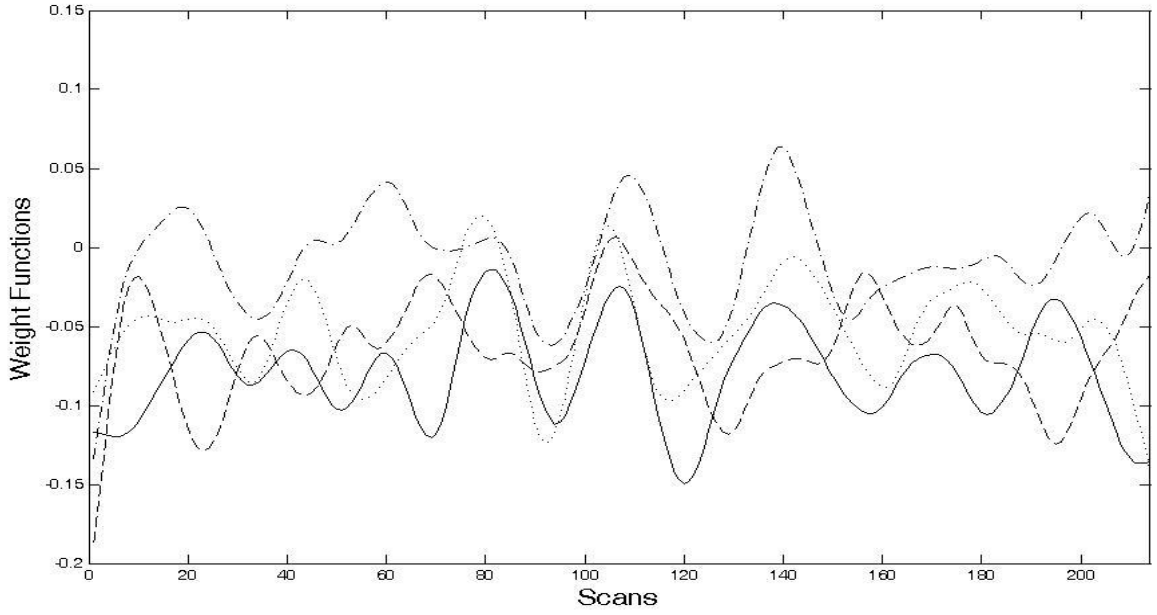


Figure 3. (A) The estimated weight functions for the second-dimensional canonical variates of four subjects obtained from the proposed method (… = subject 1, - - - = subject 2, ____ = subject 3, and __ . __ = subject 4), and (B) the mean predictor weights (averaged over subjects) for each load condition across the second-dimensional canonical variates (● = 2 letters, ■ = 4 letters, ▲= 6 letters, and * = 8 letters). The predictor weights at the first time point are adjusted to zero, and all other values scaled accordingly.

(A)



(B)

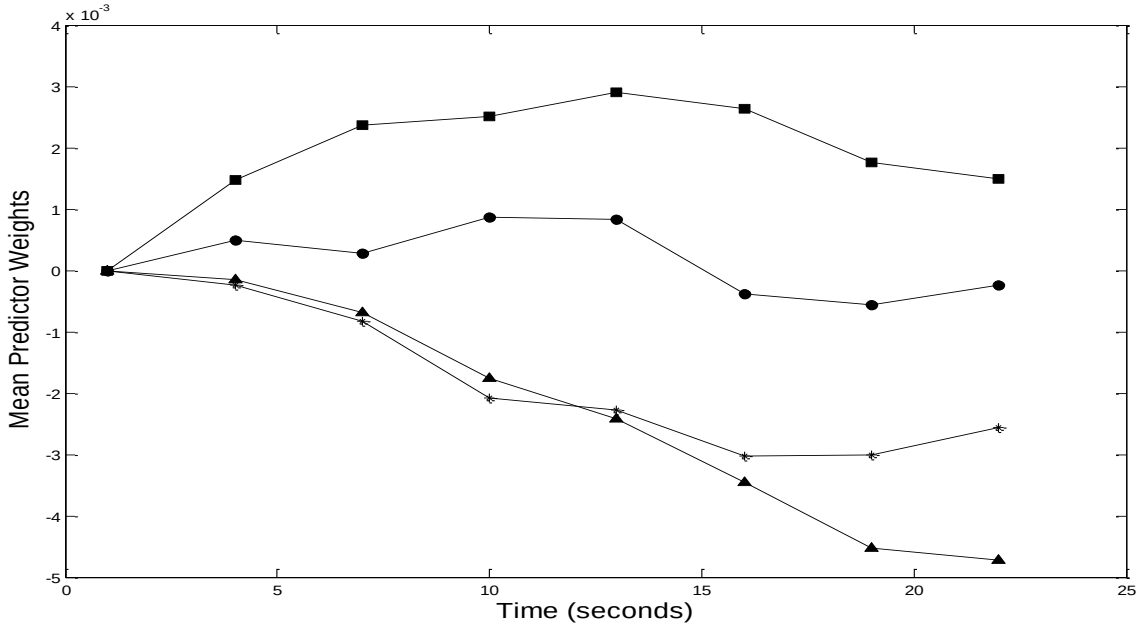


Figure 4. Slice images of the object scores of voxels for the leading canonical variates obtained from the proposed method. The dominant 5% of the object scores are displayed with positive scores in red and yellow and negative scores in blue and green.

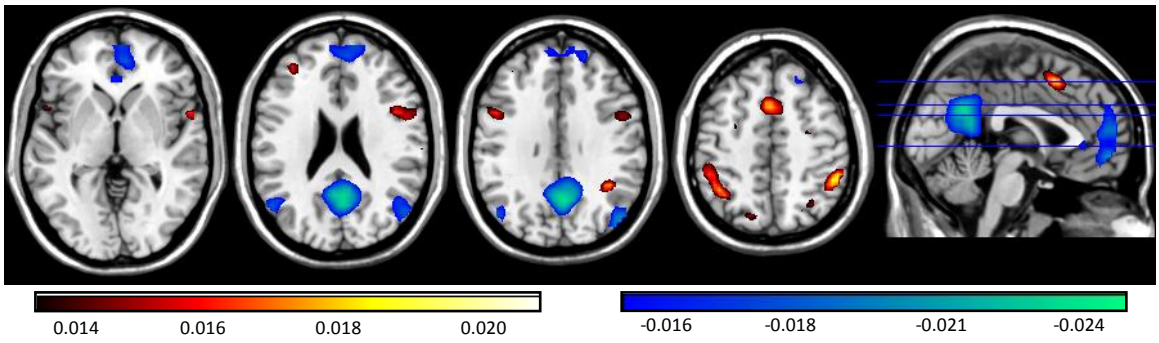
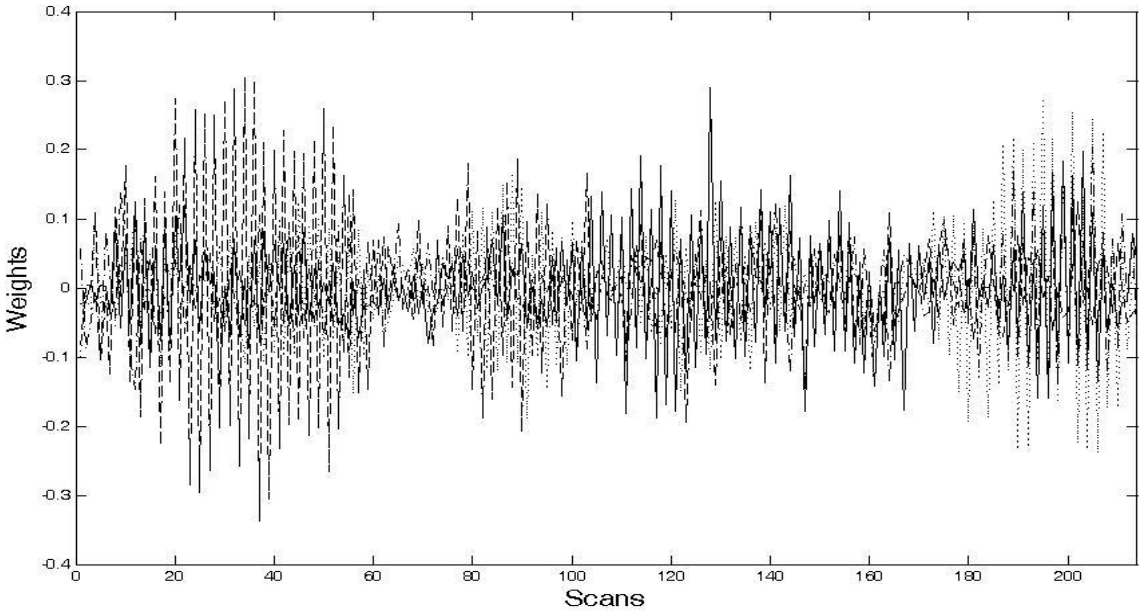


Figure 5. (A) The estimated weight values for the first-dimensional canonical variates of four subjects obtained from multiple-set canonical correlation analysis (… = subject 1, - - - = subject 2, ____ = subject 3, and __ . __ = subject 4), and (B) the mean predictor weights (averaged over subjects) for each load condition across the first-dimensional canonical variates (● = 2 letters, ■ = 4 letters, ▲= 6 letters, and * = 8 letters). The predictor weights at the first time point are adjusted to zero, and all other values scaled accordingly.

(A)



(B)

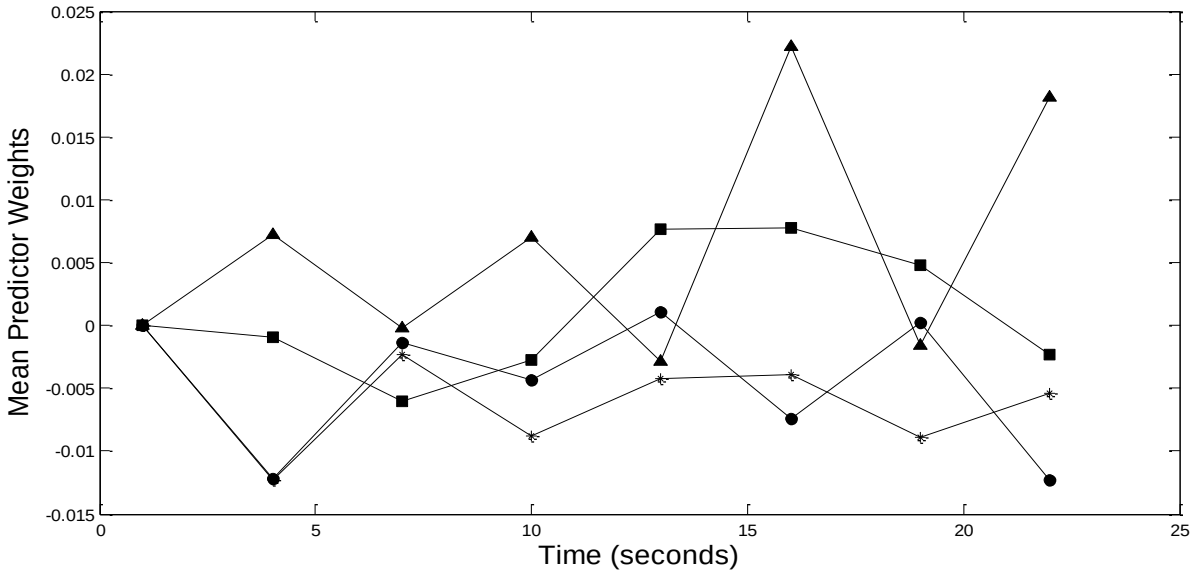
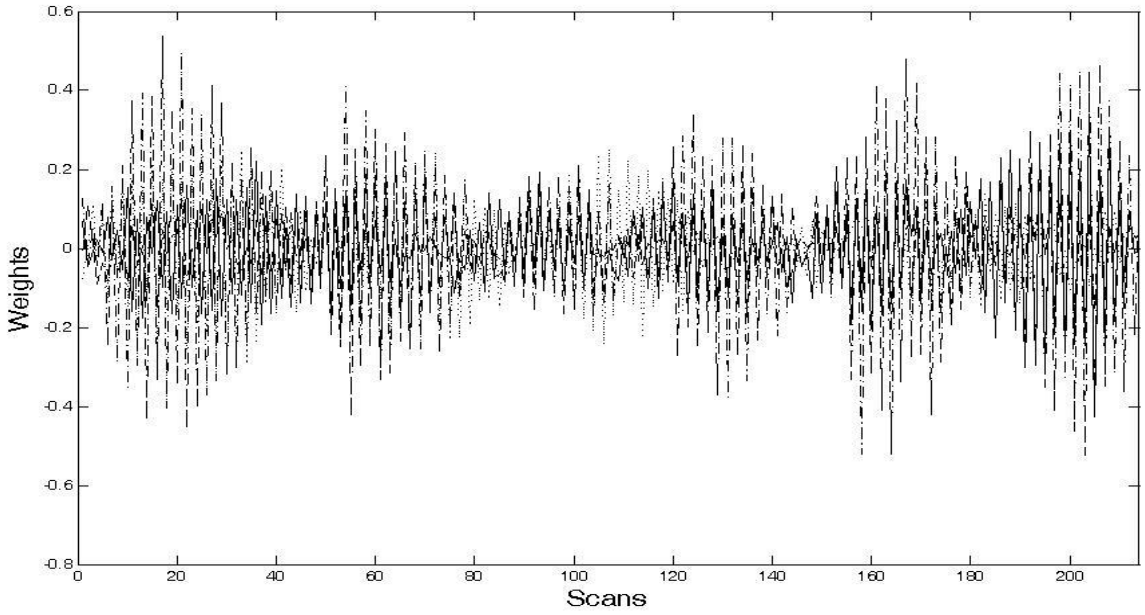


Figure 6. (A) The estimated weight values for the second-dimensional canonical variates of four subjects obtained from multiple-set canonical correlation analysis (… = subject 1, - - - = subject 2, ____ = subject 3, and __ . __ = subject 4), and (B) the mean predictor weights (averaged over subjects) for each load condition across the second-dimensional canonical variates (● = 2 letters, ■ = 4 letters, ▲= 6 letters, and * = 8 letters). The predictor weights at the first time point are adjusted to zero, and all other values scaled accordingly.

(A)



(B)