*Author / title page

A Unified Approach to Multiple-set Canonical Correlation Analysis and Principal Components Analysis

Heungsun Hwang[1], Kwanghee Jung[2], 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

March 14, 2012

The authors thank the Associate Editor and three anonymous reviewers for their constructive comments. The MATLAB codes used for the simulation and empirical studies reported in this article may be obtained by contacting the first author.
Correspondence concerning the article should be addressed to Heungsun Hwang, Department of Psychology, McGill University, 1205 Dr. Penfield Avenue, Montreal, QC, H3A 1B1, Canada. Email: heungsun.hwang@mcgill.ca.

A Unified Approach to Multiple-set Canonical Correlation Analysis and Principal Components Analysis

Abstract

Multiple-set canonical correlation analysis and principal components analysis are popular data reduction techniques in various fields including psychology. Both techniques aim to extract a series of weighted composites or components of observed variables for the purpose of data reduction. However, their objectives of performing data reduction are different. Multiple-set canonical correlation analysis focuses on describing the association among several sets of variables through data reduction, whereas principal components analysis concentrates on explaining the maximum variance of a single set of variables. In this paper, we provide a unified framework that combines these seemingly incompatible techniques. The proposed approach embraces the two techniques as special cases. More importantly, it permits a compromise between the techniques in yielding solutions. For instance, we may obtain components in such a way that they maximize the association among multiple datasets, while also accounting for the variance of each dataset. We develop a single optimization function for parameter estimation, which is a weighted sum of two criteria for multiple-set canonical correlation analysis and principal components analysis. We minimize this function analytically. We conduct simulation studies to investigate the performance of the proposed approach based on synthetic data. We also apply the approach for the analysis of functional neuroimaging data to illustrate its empirical usefulness.

Keywords: Multiple-set canonical correlation analysis, principal components analysis, functional neuroimaging data.

1. Introduction

Multiple-set canonical correlation analysis and principal components analysis represent two data reduction techniques prevalent in many fields including psychology. Multiple-set canonical correlation analysis (Carroll, 1968; Horst, 1961; Meredith, 1964) is used to investigate how low dimensional representations (called canonical variates or variables) of several sets of variables are related to one another. It subsumes canonical correlation analysis as a special case when only two sets of variables are considered. Although this technique has been considered useful in studying interrelationships among multiple sets of variables, it is also being paid much attention as a tool for integrating data obtained from different sources such as subjects, stimuli, locations, or data acquisition techniques (e.g., Correa, Li, Adali, & Calhoun, 2009; Takane & Oshima-Takane, 2002). Principal components analysis, on the other hand, is used to examine how a single set of variables is explained by its low dimensional representations called principal components.

These two techniques are technically comparable in the sense that both aim to extract a series of weighted composites or components from each dataset for the purpose of data reduction. Nonetheless, their central objectives of extracting such weighted composites are different: In multiple-set canonical correlation analysis, the weighted composites of each dataset are obtained in such a way that they are maximally related to the corresponding weighted composites from the other datasets. Conversely, in principal components analysis, the weighted composites of a single dataset are obtained to account for the maximum variance of the dataset. Due to these distinctive objectives in data reduction, they remain separate techniques at large. Multiple-set canonical correlation analysis focuses on a simpler description of the association among several sets of variables, whereas principal components analysis concentrates on a summary of the variability of a single dataset. Note that multiple-set canonical correlation analysis becomes equivalent to principal components analysis when each dataset consists of a single variable only (e.g., Gifi, 1990). However, this is not a typical situation to which we apply multiple-set canonical correlation analysis.

In practice, multiple-set canonical correlation analysis and principal components analysis have frequently been used in a complementary manner. For example, Correa and her colleagues (Correa, Eichele, Adali, Li, & Calhoun, 2010) applied multiple-set canonical correlation analysis to fuse brain imaging data concurrently acquired from two different imaging modalities such as functional magnetic resonance imaging (fMRI) and electroencephalography (EEG). Subsequently, they used principal components analysis to obtain component loadings relating the canonical variates obtained from multiple-set canonical correlation analysis to each individual dataset. This is a two-step approach where the two data reduction techniques are carried out sequentially. This approach provides no assurance that the low dimensional data representations obtained from one technique are also optimal for the objective of the subsequent technique, because the former is obtained without reference to the latter (e.g., Arabie & Hubert, 1994; Chang, 1983). In addition, the two-step approach has no means to control for the degree of differential influence that each technique may have on the final solution.

In this paper, we propose a unified approach to multiple-set canonical correlation analysis and principal components analysis. This proposed approach accommodates the two data reduction techniques as special cases. More importantly, it can provide a compromise solution by controlling for the contribution of each technique to the final solution. The degree of compromise can be determined by weighing the contribution of each technique differentially. This capability can be of particular use in addressing an enduring issue in multiple-set canonical correlation analysis, which canonical variates of each dataset may not be well representative of the dataset because this technique concentrates only on how highly the canonical variates of each dataset are correlated to those from other datasets, not on how well they explain the variance of their own dataset (e.g., Lambert, Wildt, & Durand, 1988; van den Wollenberg, 1977). By applying the proposed approach, we may obtain low dimensional representations of several datasets, which are highly related to each other across the datasets and also account for the variance of each dataset well.

Mishra (2009) recently developed a hybrid approach to canonical correlation analysis and principal components analysis. The scope of his method is somewhat narrow in that principal components analysis was adopted for exclusively overcoming the same issue in canonical correlation analysis. Thus, this method adds a non-negative weight only to principal components analysis to control for its impact on the solution, while assigning no such weight to canonical correlation analysis. Consequently, it is difficult to regard the method as a unified approach to canonical correlation analysis and principal components analysis because it cannot subsume both the techniques as special cases. The method can have only canonical correlation analysis as a special case by setting the weight for principal components analysis equal to zero. Conversely, the method appears to be more flexible in that it can formulate the integration of principal components analysis in three different ways. Nonetheless, in most cases of weighing principal components analysis, this method may be considered a special case of the proposed approach, which involves two sets of variables only.

Dahl and Naes (2006) proposed so-called ridge generalized canonical analysis that incorporated a ridge parameter into the matrix eigen-analysis problem for multiple-set canonical correlation analysis. This method can be of use in stabilizing solutions by shrinking the influence of low-variance components. Computationally, it can also embrace the eigen-analysis problems for multiple-set canonical correlation analysis and principal components analysis by setting the ridge parameter at 0 and 1, respectively. However, it is unknown which optimization criterion is to be maximized by solving the eigen-analysis problem with the ridge parameter varying between 0 and 1. In other words, the method does not have an optimization criterion formulated under a clear objective of analysis. Conversely, the proposed approach provides a single optimization function, which is developed under the aim of combining multiple-set canonical correlation analysis and principal components analysis into a unified framework.

In Section 2, we explain the technical underpinnings of the proposed unified approach. We present a single optimization function for parameter estimation in the proposed approach. The optimization function is equivalent to a weighted sum of two criteria for multiple-set canonical correlation analysis and principal components analysis. We also show that this function can be optimized analytically. In Section 3, we evaluate the parameter recovery capability of the proposed approach and investigate its relative performance to multiple-set canonical correlation analysis and principal components analysis through the analysis of synthetic data. In Section 4, we illustrate the empirical usefulness of the proposed approach by analyzing functional neuroimaging data obtained from several subjects during a working memory experiment. In the final section, we summarize the theoretical and empirical implications of the proposed approach and discuss potential topics for future research.

2. The Proposed Unified Approach to Multiple-set Canonical Correlation Analysis and Principal Components Analysis

Let $\mathbf{Z}_k$ denote an N by p_k matrix of variables in the kth dataset ($k = 1, \cdots, K$), where N is the number of cases. Let $\mathbf{W}_k$ denote a p_k by D matrix of weights assigned to each variable in $\mathbf{Z}_k$, where D is the number of dimensions. Let $\mathbf{F}$ denote an N by D matrix of low dimensional data representations, often called object scores, which characterize the association or homogeneity among all $\mathbf{Z}_k$'s. Let $\mathbf{A}_k$ denote a D by p_k matrix of loadings relating $\mathbf{F}$ to $\mathbf{Z}_k$. Let α and β denote non-negative scalar values.

Our aim is to combine multiple-set canonical correlation analysis (MCCA) and principal components analysis (PCA) into a single framework. This means that we seek to obtain low dimensional representations (i.e., $\mathbf{F}$) of K sets of variables such that they maximize the association among them as in MCCA, while accounting for the variance of $\mathbf{Z}_k$ as much as possible like PCA. This aim can be achieved by minimizing the following optimization function,

$$f = \alpha \sum_{k=1}^{K} \mathrm{SS}(\mathbf{F} - \mathbf{Z}_k \mathbf{W}_k) + \beta \sum_{k=1}^{K} \mathrm{SS}(\mathbf{Z}_k - \mathbf{F}\mathbf{A}_k), \tag{1}$$

with respect to $\mathbf{F}$, $\mathbf{W}_k$, and $\mathbf{A}_k$, subject to the normalization constraint $\mathbf{F}'\mathbf{F} = \mathbf{I}$ for identification, and $\alpha + \beta = 1$, where SS($\mathbf{H}$) = tr($\mathbf{H}'\mathbf{H}$). When $\alpha = 1$, the first term in (1) is equivalent to the homogeneity criterion for MCCA (Gifi, 1990). When $\beta = 1$, the second term is equivalent to the criterion for PCA for $\mathbf{Z} = [\mathbf{Z}_1, \cdots, \mathbf{Z}_K]$. Accordingly, this optimization function is a weighted sum of two criteria for multiple-set canonical correlation analysis and principal components analysis. However, note that the same matrix of object scores (**F**) appears in both criteria. As a result, by minimizing (1), we can obtain **F** considering the objectives of multiple-set canonical correlation analysis and principal components analysis simultaneously.

As shown in (1), the proposed approach can deal with multiple-set canonical correlation analysis and principal components analysis as special cases: It becomes equivalent to multiple-set canonical correlation analysis when $\alpha = 1$ and $\beta = 0$, and reduces to principal components analysis for the entire data when $\alpha = 0$ and $\beta = 1$. In addition, this approach permits a compromise between the two techniques under $\alpha \neq 0$ and $\beta \neq 0$. We should a priori specify the values of α and β based on our research objectives or interests. By specifying $\alpha = \beta = .5$, we assume that both techniques contribute equally to the final solution. We may also adjust for the influence of the two techniques on the final solution by differently weighing the two criteria of (1). For example, we wish to weigh the first criterion more heavily than the second (i.e., $\alpha > \beta$), so that multiple-set canonical correlation analysis has a greater impact on the final solution than principal components analysis. This weighting scheme may be of use when we believe that it is more important to maximize the association among multiple datasets than to account for the variance of each dataset.

We can solve the minimization problem (1) in closed form. As $\mathbf{W}_k$ is involved only in the first criterion of (1) and $\mathbf{A}_k$ is only in the second, under $\mathbf{F}'\mathbf{F} = \mathbf{I}$, their least squares estimates are given by

$$\mathbf{W}_k = (\mathbf{Z}_k{'}\mathbf{Z}_k)^{-1}\mathbf{Z}_k{'}\mathbf{F}, \tag{2}$$

and

$$\mathbf{A}_k = (\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'\mathbf{Z}_k = \mathbf{F}'\mathbf{Z}_k. \tag{3}$$

Given the estimates of $\mathbf{W}_k$ and $\mathbf{A}_k$, $\mathbf{F}$ can be estimated as follows. Let

$\mathbf{\Omega}_k = \mathbf{Z}_k(\mathbf{Z}_k{'}\mathbf{Z}_k)^{-1}\mathbf{Z}_k{'}$, and $\mathbf{\Psi} = \mathbf{F}\mathbf{F}'$. Note that both $\mathbf{\Omega}_k$ and $\mathbf{\Psi}$ are idempotent and symmetric. Putting (2) and (3) into (1), minimizing (1) with respect to $\mathbf{F}$, subject to $\mathbf{F}'\mathbf{F} = \mathbf{I}$, is equivalent to minimizing

$$\begin{aligned}
f &= \alpha\sum_{k=1}^{K}\mathrm{SS}(\mathbf{F}-\mathbf{\Omega}_k\mathbf{F}) + \beta\sum_{k=1}^{K}\mathrm{SS}(\mathbf{Z}_k - \mathbf{\Psi}\mathbf{Z}_k) \\
&= \alpha\sum_{k=1}^{K}\mathrm{tr}(\mathbf{F}'\mathbf{F}-\mathbf{F}'\mathbf{\Omega}_k\mathbf{F}) + \beta\sum_{k=1}^{K}\mathrm{tr}(\mathbf{Z}'_k\,\mathbf{Z}_k - \mathbf{Z}'_k\,\mathbf{\Psi}\mathbf{Z}_k) \\
&= \alpha\sum_{k=1}^{K}\mathrm{tr}(\mathbf{F}'\mathbf{F}) + \beta\sum_{k=1}^{K}\mathrm{tr}(\mathbf{Z}'_k\,\mathbf{Z}_k) - \alpha\sum_{k=1}^{K}\mathrm{tr}(\mathbf{F}'\mathbf{\Omega}_k\mathbf{F}) - \beta\sum_{k=1}^{K}\mathrm{tr}(\mathbf{Z}'_k\,\mathbf{\Psi}\mathbf{Z}_k) \\
&= \alpha KD + \beta\sum_{k=1}^{K}\mathrm{tr}(\mathbf{Z}'_k\,\mathbf{Z}_k) - \sum_{k=1}^{K}\mathrm{tr}(\alpha\mathbf{F}'\mathbf{\Omega}_k\mathbf{F} + \beta\mathbf{F}'\mathbf{Z}_k\mathbf{Z}'_k\,\mathbf{F}) \\
&= \alpha KD + \beta\sum_{k=1}^{K}\mathrm{tr}(\mathbf{Z}'_k\,\mathbf{Z}_k) - \mathrm{tr}\left(\mathbf{F}'\left[\sum_{k=1}^{K}\alpha\mathbf{\Omega}_k + \beta\mathbf{Z}_k\mathbf{Z}'_k\right]\mathbf{F}\right).
\end{aligned} \tag{4}$$

Minimizing (4) thus reduces to maximizing

$$\mathrm{tr}\left(\mathbf{F}'\left[\sum_{k=1}^{K}\alpha\mathbf{\Omega}_k + \beta\mathbf{Z}_k\mathbf{Z}'_k\right]\mathbf{F}\right), \tag{5}$$

with respect to $\mathbf{F}$. This maximization is equivalent to calculating the following eigenvalue decomposition,

$$\sum_{k=1}^{K}\alpha\mathbf{\Omega}_k + \beta\mathbf{Z}_k\mathbf{Z}'_k = \mathbf{\Gamma}\mathbf{\Delta}\mathbf{\Gamma}', \tag{6}$$

where $\mathbf{\Gamma}'\mathbf{\Gamma} = \mathbf{I}$, and $\mathbf{\Delta}$ is a diagonal matrix consisting of eigenvalues as elements. Then, $\mathbf{F}$ is obtained as the first D columns of $\mathbf{\Gamma}$ (e.g., Yanai, 1998).

3. Simulation Studies

We carry out two simulation studies to investigate the performance of the proposed approach based on synthetic data.

3.1. Simulation Study 1

In the first study, we focused on how well the proposed approach recovered the parameter values of $\mathbf{W}_k$ and $\mathbf{A}_k$ under different sample sizes. We did not evaluate recovery of $\mathbf{F}$ because the number of object scores changed with sample size. The process of generating synthetic data can be summarized as follows: We chose the parameter values of $\mathbf{W}_k$ and $\mathbf{A}_k$. For each sample size, we drew an N by D matrix of object scores ($\mathbf{F}$) from the standard normal distribution and subsequently, normalized it such that $\mathbf{F}'\mathbf{F} = \mathbf{I}$. We also drew an N by D+ p_k matrix, denoted by $\mathbf{E}_k$, from a normal distribution with mean 0 and standard deviation σ. We then generated an N by p_k matrix of variables $\mathbf{Z}_k$ by $\mathbf{Z}_k = [\mathbf{Y}_k + \mathbf{E}_k]\mathbf{Q}_k'(\mathbf{Q}_k\mathbf{Q}_k')^{-1}$, where $\mathbf{Y}_k = [-\alpha\mathbf{F}, \beta\mathbf{F}\mathbf{A}_k]$, and $\mathbf{Q}_k = [-\alpha\mathbf{W}_k, \beta\mathbf{I}]$. This way of generating $\mathbf{Z}_k$ was derived from (1) or minimizing the sum of squares of $\mathbf{E}_k = [\alpha\mathbf{F}, \beta\mathbf{Z}_k] - [\alpha\mathbf{Z}_k\mathbf{W}_k, \beta\mathbf{F}\mathbf{A}_k] = \mathbf{Z}_k[-\alpha\mathbf{W}_k, \beta\mathbf{I}] - [-\alpha\mathbf{F}, \beta\mathbf{F}\mathbf{A}_k] = \mathbf{Z}_k\mathbf{Q}_k - \mathbf{Y}_k$ for each set of variables.

For this study, we considered six sample sizes (N = 20, 50, 100, 200, 500, and 1000). In addition, we specified that $K = 4$, $p_1 = p_2 = p_3 = p_4 = 10$, $D = 3$, $\alpha = \beta = .5$, and $\sigma = .03$, which was equal to the average standard deviation of $\mathbf{Y}_k$ at N = 1000. At each sample size, we generated 500 samples, each of which was analyzed by the proposed approach. To assess the properties of parameter estimates obtained under the proposed approach, we computed the relative biases, standard deviations, and mean square errors (MSE) of the estimates of weights and loadings across different sample sizes. In particular, the MSE is proportional to the sum of the bias and standard deviation of an estimate, indicating how close the estimate is to its parameter value on average (Mood, Graybill, & Boes, 1974).

Figure 1 displays the average relative bias, standard deviation, and mean square error of the estimates of weights and loadings across different sample sizes. We considered absolute values of relative bias greater than ten percent indicative of an unacceptable degree of bias (e.g., Bollen, Kirby, Curran, Paxton, & Chen, 2007; Lei, 2009). As shown in this figure, the proposed approach on average yielded positively biased estimates of weights and loadings. Nonetheless, the degrees of relative bias were acceptable for all sets of estimates because they were smaller than ten percent in absolute value. Moreover, the parameter estimates of the proposed approach were generally associated with quite small standard deviations across all the sample sizes. Lastly, the proposed approach involved very small average MSE values of all parameter estimates across the sample sizes. The average MSE values of the estimates tended to decrease with sample size increased.

Thus, the proposed approach seemed to recover parameters sufficiently well. The estimates of the proposed approach had a positive yet tolerant level of bias, while they involved a very small level of variability. Consequently, the mean square errors of the estimates were nearly zeros, indicating that the estimates were quite close to the parameters on average.

Insert Figure 1 about here

3.2. Simulation Study 2

In the second study, we investigated the relative performance of the proposed approach, as compared to MCCA and PCA. In particular, we compared the capability of these three methods to recover parameters under different values of α altering from .05 to .95. The three methods entail different sets of parameters: $\mathbf{F}$ and $\mathbf{W}_k$ in MCCA, $\mathbf{F}$ and $\mathbf{A}_k$ in PCA, and $\mathbf{F}$, $\mathbf{W}_k$ and $\mathbf{A}_k$ in the proposed approach. Thus, we concentrated on how well the methods recovered $\mathbf{F}$, which was the only common set of parameters among them. As stated earlier, however, recovery of $\mathbf{F}$ can be affected by sample size. To avoid this issue, we fixed $N = 100$ for the second study. For each value of α, we generated 500 random samples based on the same data generation procedure used for the first simulation study.

To examine the relative accuracy of object score estimates obtained from the three methods, we calculated the average mean absolute deviations (MAD) between true and estimated object scores for each value of α. Figure 2 exhibits the average MAD values for the three methods across different values of α. As expected, when α became close to zero, the proposed approach resulted in MAD values similar to those from PCA. In such cases, MCCA yielded larger MAD values than the other methods. Conversely, when α approached one, the proposed approach resulted in MAD values similar to those from

MCCA, whereas PCA produced the largest MAD values. Moreover, the proposed approach tended to provide smaller MAD values than MCCA and PCA across all the values of α. This indicates that the proposed approach recovered the common set of parameters (object scores) better than MCCA and PCA.

Insert Figure 2 about here

4. An Application to Functional Neuroimaging Data

In this section, we apply the proposed approach to functional neuroimaging data in order to demonstrate its empirical usefulness. The present example is part of fMRI data obtained from a verbal working memory study (Cairo, Woodward, & Ngan, 2006; Metzak, Riley, Wang, Whitman, Ngan, & Woodward, 2011). 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 over scans in every voxel in the brain.

4.1. The Data

In this application, four women subjects, who were right-handed, healthy and native English speakers (age range 18-35), performed a variable load delayed recognition working memory task while undergoing fMRI. The variable load delayed recognition memory task consisted of encoding, maintenance, and retrieval phases and four different memory load conditions. During a single trial of this task, the subjects viewed a string of 2, 4, 6 or 8 different uppercase consonants for 4 seconds (the encoding phase), which they were instructed to remember over a short 6-second delay (the maintenance phase). Following the delay, a single lowcase consonant was shown for 1 second. Subjects were asked to decide whether this letter had been included in the preceding letter string (the retrieval phase). The probe stage was followed by an inter-trial interval of 6 seconds in duration. Each stimulus run consisted of 214 scans of the entire brain, and 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.

Thus, we had four sets of variables, each of which was composed of 23,621 cases, representing voxels, and 214 variables, indicating scans measured for each subject. The BOLD signal was realigned, spatially normalized and smoothed prior to analysis using Statistical Parametric Mapping (SPM2).

Furthermore, a design matrix was developed that explicitly reflected a finite impulse response modeling BOLD signal changes to the stimulus presentation scheme of the experiment for each scan. Each set of the original data was then decomposed into a portion explained by this design matrix and the residual portion unexplained by the design matrix (Metzak, Feredoes, Takane, Wang, Weinstein, Cairo, Ngan, & Woodward, 2011; Woodward, Cairo, Ruff, Takane, Hunter, & Ngan, 2006). Technically, this decomposition was carried out by regressing each set of the data on the design matrix (e.g., Takane & Shibayama, 1991). We used only the portion of the data explained by the design matrix for actual analyses. This allowed analyzing the data that were directly relevant to the experimental conditions.

4.2. Analysis Objectives

The main objective of our analysis was to integrate signal variation in four subjects' BOLD signal into highly-correlated low dimensional representations so as to identify regions of the brain, which were commonly activated among the subjects who were performed the same working memory task. This technically required low dimensional data integration of brain voxels over multiple subjects. Thus, multiple-set canonical correlation analysis might be a sensible choice to achieve the main objective (e.g., Correa et al., 2010). However, at the same time, we sought to obtain these low dimensional representations (or object scores) in such a way that they also explained the data sufficiently well. This would be our secondary objective of analysis.

A wide range of effortful cognitive tasks consistently lead to not only increases in activity in fronto-parietal brain regions but also concomitant decreases in activity in ventro-medial brain regions (Fox, Snyder, Vincent, Corbetta, Van Essen, & Raichle, 2005). These anticorrelated networks have been coined the task-positive and task-negative network, respectively (Fox et al., 2005). Due to the emergence of these networks in past working memory research (Metzak, Feredoes, Takane, Wang, Weinstein, Cairo, Ngan, & Woodward, 2011a; Metzak et al., 2011b), we expected variations in these networks to emerge in the current analysis.

4.3. Analysis Results

At first, we had applied the proposed approach to the data under $\alpha = \beta = 0.5$. This means that both criteria for multiple-set canonical correlation analysis and principal components analysis in (1) had the same degree of influence on yielding the final solution. In this application, we found that the PCA criterion was responsible for 99% of the optimization function value, while the MCCA criterion accounted for only 1%. Thus, if we adopt such an equal weighting scheme, the PCA criterion would have a dominant effect on the final solution, forcing the solution to be almost identical to the PCA solution. This analysis would not be well suited to fulfilling our main objective.

Accordingly, we decided to set $\alpha = .99$ and $\beta = .01$; in other words, the MCCA criterion was to have a disproportionately favorable effect on the final solution over the PCA criterion. In this way, we might be able to balance out the influence of MCCA and PCA on the final solution. In the present application, we concentrate on the first two-dimensional solutions (i.e., $D = 2$) because it was difficult to interpret the subsequent dimensions. More importantly, as will be shown below, investigating the first two dimensions offers sufficient insights for the basic questions of analysis.

Figure 3 exhibits five slice images constructed from the (voxel) object scores that were calculated based on the leading dimension. The far-right sagittal brain image indicates which slices of the brain the first four images represent. All the images display the dominant 5% of the object scores mapped onto a structural brain image template. In this case, all the dominant object scores were positive, displayed in red and yellow. The images of these object scores represent functional networks that were positively activated in all four subjects during the experiment, suggesting that these regions were likely to be functionally connected across the subjects. These regions include the bilateral dorsolateral prefrontal cortices, dorsal anterior cingulate cortex, bilateral precental gyri, bilateral inferior frontal cortices, bilateral inferior and superior parietal cortices, and bilateral inferior occipital gyri.

Insert Figure 3 about here

Figure 4 displays five slice images of the object scores of the voxels corresponding to the second dimension. Again, the far-right sagittal brain image indicates which slices of the brain the first four images are. All the images present the dominant 5% of the object scores. Interestingly, all these dominant scores were negative scores, displayed in blue. This indicates that the functional network for the second dimensional solution comprised the elements of task-negative (i.e., default) networks. The task-negative network is a system of functionally connected brain regions that are thought to reduce activation during the memory task. Here this is characterized by decreased activation in the bilateral postcentral gyri, ventral anterior cingulate cortex, posterior cingulate gyrus, bilateral angular gyri, bilateral insular cortices, and bilateral primary auditory cortices.

Insert Figure 4 about here

The solutions in Figures 3 and 4 provide useful information with respect to the brain regions commonly activated among the four subjects and their characteristics. For comparison purposes, nonetheless, we also present the solutions obtained from multiple-set canonical correlation analysis. This analysis corresponds to applying the proposed approach under $\alpha = 1$ and $\beta = 0$. We do not provide the solutions obtained from PCA only, or equivalently the proposed approach under $\alpha = 0$ and $\beta = 1$, because our main objective (i.e., integrating signal variation in each subject's brain voxels into low-dimensional representations that are highly correlated across subjects) could not be achieved by using PCA alone, which focuses solely on a single dataset.

Figure 5 shows five slice images constructed from the voxel object scores that were calculated based on the leading dimension obtained from multiple-set canonical correlation analysis. The far-right sagittal brain image indicates which slices of the brain the first four images represent. All the images present the dominant 5% of the object scores. In this case, a majority of the dominant object scores were positive, indicating that the images of these object scores are likely to represent task-positive networks that were positively activated in all four subjects during the experiment. These regions include the left inferior frontal gyrus, dorsal anterior cingulate cortex, bilateral postcentral gyrus, bilateral inferior parietal cortex, bilateral thalamus and bilateral cerebellum. Thus, it seems that the first dimensional solution obtained from multiple-set canonical correlation analysis bears a strong resemblance to that in Figure 3.

Insert Figure 5 about here

Figure 6 displays five slice images of the object scores of the voxels corresponding to the second dimension obtained from multiple-set canonical correlation analysis. All the images present the dominant 5% of the object scores where positive scores are in red and yellow, and negative scores are in blue. The functional network for the second dimension contained the elements of both task-positive and negative networks. The task-positive network is dominated by increased activation in the right inferior frontal gyrus, dorsal anterior cingulate cortex, and bilateral superior parietal gyri. The task-negative network is characterized by decreased activation in the ventral anterior cingulate cortex, posterior cingulate cortex, bilateral angular gyri, bilateral primary auditory cortices, bilateral inferior temporal cortices, insular cortex, primary visual cortex, posterior inferior gyrus, inferior occipital gyrus, middle frontal gyrus, posterior cingulate gyrus, primary auditory cortex, primary motor cortex, lateral premotor area, and primary somasensory cortex. Thus, the second dimension from multiple-set canonical correlation analysis resulted in a different solution from that in Figure 4. In the previous analysis, the same dimension involved task-negative networks only.

It is difficult to compare these two analyses directly because they had different objectives of analysis under different weighting schemes. The previous analysis under α = .99 and β = .01 appears to provide a more interpretable solution than multiple-set canonical correlation analysis, because it shows that the first dimension was related exclusively to task-positive networks and the second was only to task-negative networks, suggesting that the task-positive networks explain the variance of each subject's brain activity well in the first dimension, whereas the task-negative networks account for the variance well in the second dimension. Conversely, the multiple-set canonical correlation analysis results produced a better formed task-negative network, and functional networks involving both the task-negative and task-positive aspects have been reported in previous analyses of similar working memory tasks (Metzak et al., 2011a,b).

Insert Figure 6 about here

6. Concluding Remarks

We proposed a unified approach to multiple-set canonical correlation analysis and principal components analysis. This approach permits handling these two widely used data reduction techniques in an integrated and interactive manner. As stated in the Introduction section, the proposed approach can be used to address a long-standing issue inherent to multiple-set canonical correlation analysis, i.e., no available mechanism for extraction of canonical variates such that they also explain the variance of their own dataset well.

We investigated the parameter recovery capability of the proposed approach through the analysis of synthetic data. In general, the approach resulted in parameter estimates that were quite close to their parameters. Moreover, it was found to recover object scores better than MCCA and PCA.

As illustrated in the empirical application section, we can adjust for weighting schemes for multiple-set canonical correlation and principal components analysis in the proposed approach. This may be of help in conducting an analysis that is better suited for our research objectives. In addition, the weighting feature of the proposed approach allows exploring alternative solutions under more diverse hypotheses. Nevertheless, choices of the scalar weights are contingent on how to set the objectives of analysis. This means that it is difficult to select the values of the scalar weights in an automated manner.

In practice, we suggest that the researcher begin by adopting equal weights (i.e., $\alpha = \beta =$ .5) because these values were to recover parameters sufficiently well, as provided in our simulation study, and then probe alternative solutions based upon differential weighting.

We may further extend and refine the proposed approach. Firstly, we may consider a nonlinear version of the proposed approach for the analysis of discrete data. In this nonlinear version, discrete data may be converted to be continuous through the adoption of a certain type of data transformation such as optimal scaling (Bock, 1960; Young, 1981). The nonlinear extension will include nonlinear multiple-set canonical correlation analysis and nonlinear principal components analysis as special cases (Gifi, 1990). Moreover, we may extend the proposed approach to deal with functional data. Due to advances in technology, in various areas of psychology, data are being collected in the form of curves, surfaces or images as a function of time, space, or other continua; for example, eye-tracking data (e.g., Jackson & Sirois, 2009), music cognition data (e.g., Vines, Nuzzo, & Levitin, 2005), facial temperature data (e.g., Jang & Lee, 2009), event-contingent social interaction data (e.g., Moskowitz, Russell, Zuroff, Bleau, Pinard, & Young, 2006), etc. A functional version of the proposed approach will be a promising tool that integrates functional multiple-set canonical correlation analysis (Hwang, Jung, Takane, & Woodward, in press) and functional principal components analysis (Rice & Silverman, 1991; Ramsay & Silverman, 2005, chapter 8) into a single framework. Lastly, we may develop a constrained version of the proposed approach, which imposes a variety of linear constraints (e.g., zero, non-additivity, or equality constraints) on both criteria of (1) in a manner similar to Takane and Shibayama (1991). This constrained version will embrace constrained principal components analysis (Takane & Hunter, 2001; Takane &

Shibayama, 1991) and constrained canonical correlation analysis (Takane & Hwang, 2002) as special cases. Future studies are warranted to investigate the technical and empirical feasibility of these extensions.

References

Arabie, P., & Hubert, L. (1994). Cluster analysis in marketing research. In R. P. Bagozzi (Ed.), *Advanced Methods of Marketing Research* (pp. 160-189). Oxford: Blackwell.

Bock, R. D. (1960). *Methods and applications of optimal scaling*. Psychometric Laboratory Report #25, University of North Carolina.

Bollen, K. A., Kirby, J. B., Curran, P. J., Paxton, P. M., & Chen, F. (2007). Latent variable models under misspecification: Two-stage least squares (2SLS) and maximum likelihood (ML) estimators. *Sociological Methods and Research*, *36*, 46-86.

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.

Chang, W. (1983). On using principal components before separating a mixture of two multivariate normal distributions. *Applied Statistics*, *32*, 267-275.

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.

Dahl, T., & Naes, T. (2006). A bridge between Tucker-1 and Carroll's generalized canonical analysis. *Computational Statistics and Data Analysis, 50*, 3086-3098.

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 USA*, 102, 9673-9678.

Gifi, A. (1990). *Nonlinear Multivariate Analysis*. Chichester: Wiley.

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

Hwang, H., Jung, K., Takane, Y., & Woodward, T. S. (in press). Functional multiple-set canonical correlation analysis. *Psychometrika*.

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

Jang, S. L., & Lee, J. H. (2009). A preliminary study to develop an emotion elicitation scenario: Using emotion appraisal components. Paper presented at the 2009 International Conference on Asia Pacific Psychology (ICAPP), Yonsei University, Seoul, Korea, August 24-25.

Lambert, Z. V., Wildt, A. R., & Durand, R. M. (1988). Redundancy analysis: An alternative to canonical correlation and multivariate multiple regression in exploring interset associations. *Psychological Bulletin*, *104*, 282-289.

Lei, P.-W. (2009). Evaluating estimation methods for ordinal data in structural equation modeling. *Quality and Quantity, 43,* 495-507.

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. (2011a). Constrained principal component analysis reveals functionally connected load-dependent networks involved in multiple stages of working memory. *Human Brain Mapping*, *32 (6)*, 856-871.

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

Mishra, S. K. (2009). Representation-constrained canonical correlation analysis: A hybridization of canonical correlation and principal component analysis. Working Papers NEHU-Economics-4. Available at http://works.bepress.com/sk_mishra/4.

Mood, A. M., Graybill, F. A., & Boes, D. C. (1974). *Introduction to the Theory of Statistics*. McGraw-Hill.

Moskowitz, D. S., Russell, J., Zuroff, D. C., Bleau, P., Pinard, G., & Young, S. N. (2006). Interpersonal behaviour, affect, and cognition in the everyday lives of individuals with social anxiety disorder: An investigation of baseline and change during paroxetine treatment. Report for GlaxoSmithKline.

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., & 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

van den Wollenberg, A. L. (1977). Redundancy analysis: An alternative for canonical analysis. *Psychometrika*, *42*, 207-219.

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.

Woodward, T. S., Cairo, T. A., Ruff, C. C., Takane, Y., Hunter, M.A. & Ngan, E.T.C. (2006). Functional connectivity reveals load dependent neural systems underlying encoding and maintenance in verbal working memory. *Neuroscience*, *139*, 317-325.

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-Verlag.

Young, F. W. (1981). Quantitative analysis of qualitative data. *Psychometrika, 46*, 357-388.

Figure 1. The average relative bias (RBIAS), standard deviation (SD), and mean square error (MSE) of weight and loading estimates obtained from the proposed approach.

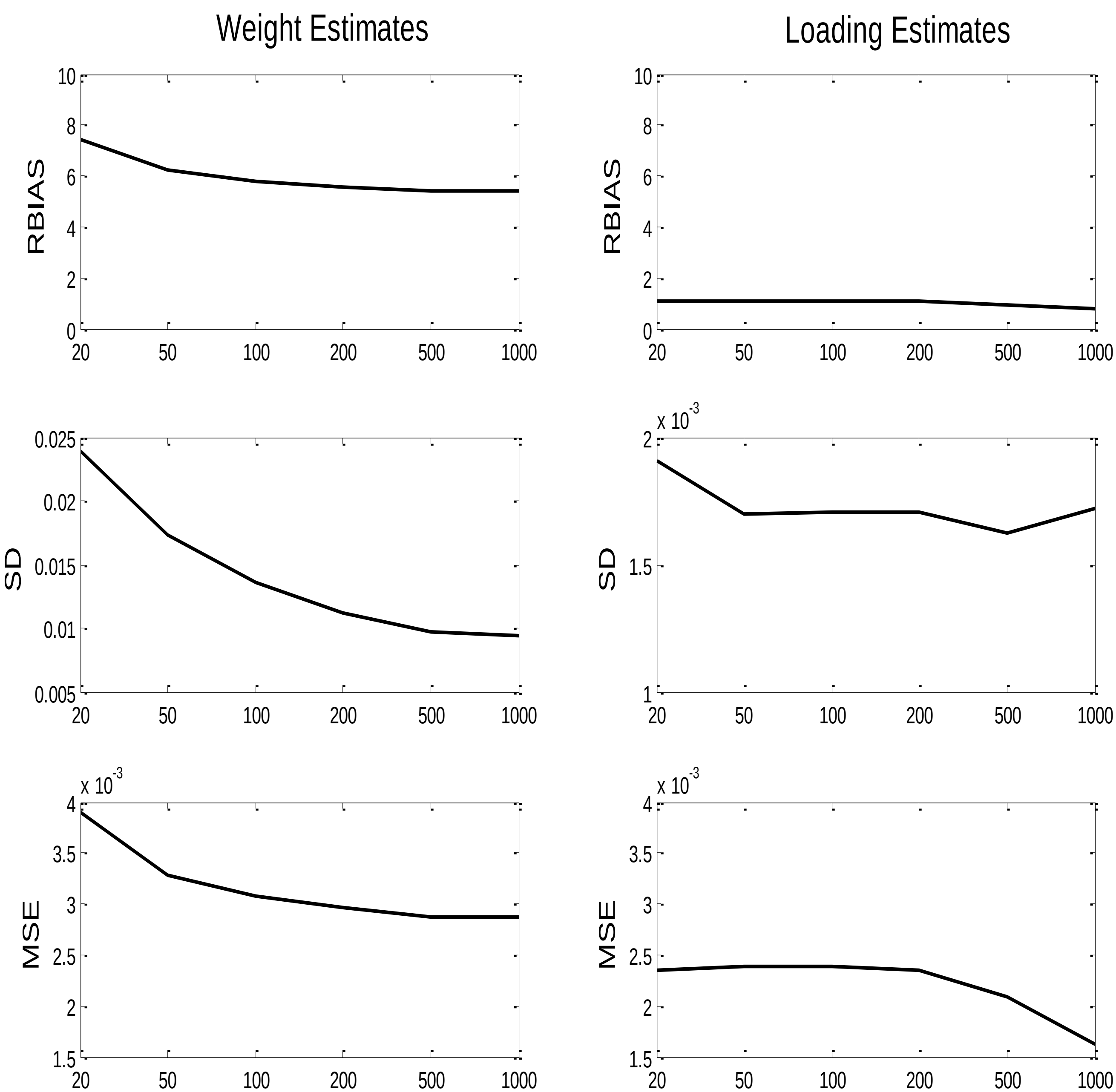

Figure 2. The average mean absolute deviations of object score estimates (MAD(F)) obtained from the proposed approach (____), MCCA (__ . __), and PCA (_ _ _), under different values of α.

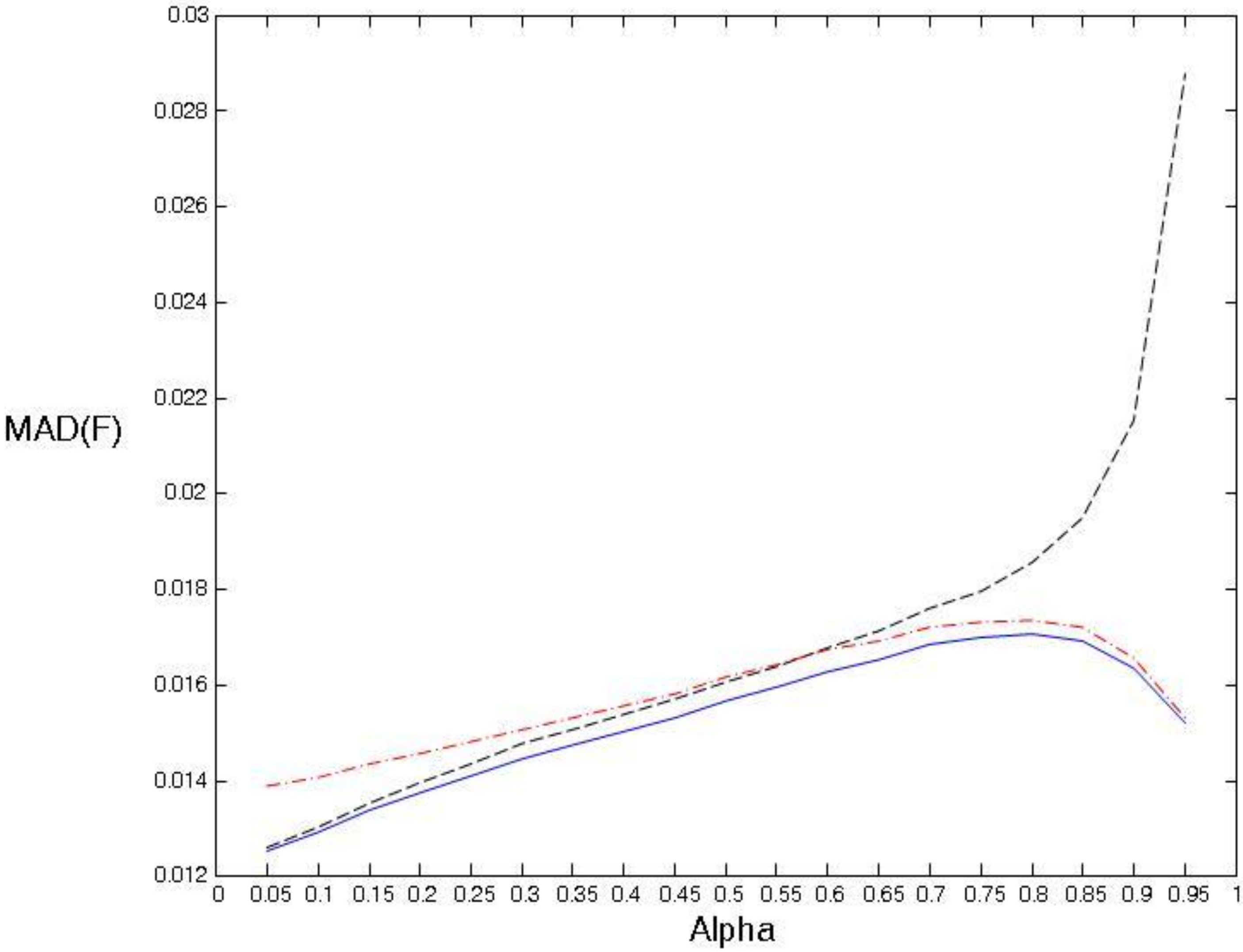

Figure 3. The first dimensional solution obtained from the proposed approach under α = .99 and β = .01.

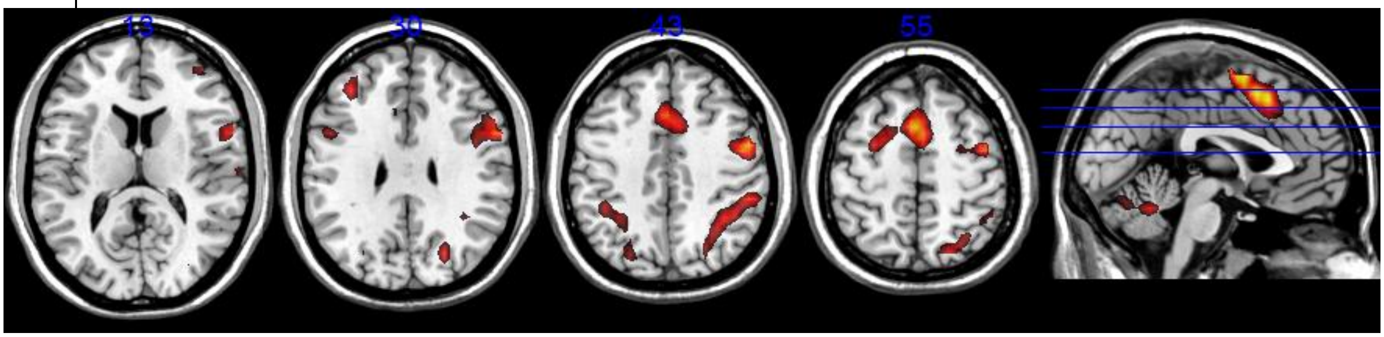

Figure 4. The second dimensional solution obtained from the proposed approach under α = .99 and β = .01.

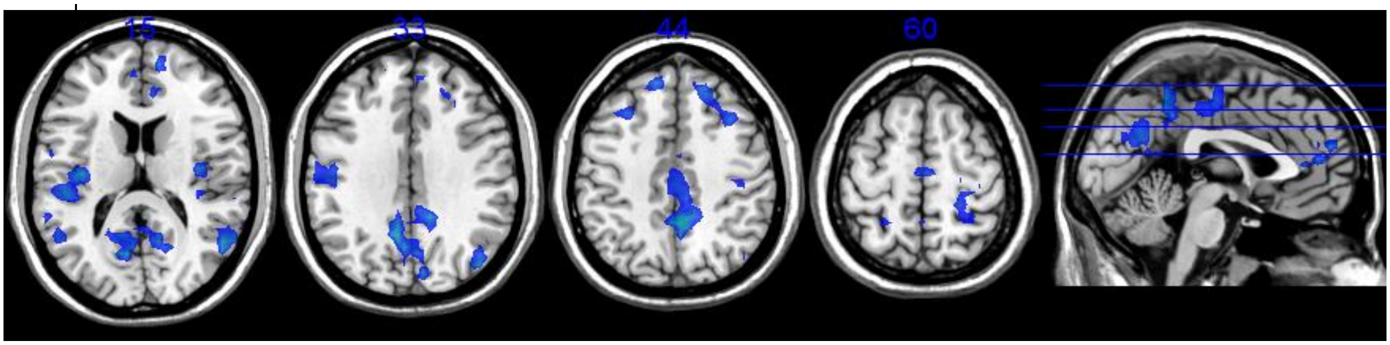

Figure 5. The first dimensional solution obtained from multiple-set canonical correlation analysis or equivalently the proposed approach under $\alpha = 1$ and $\beta = 0$.

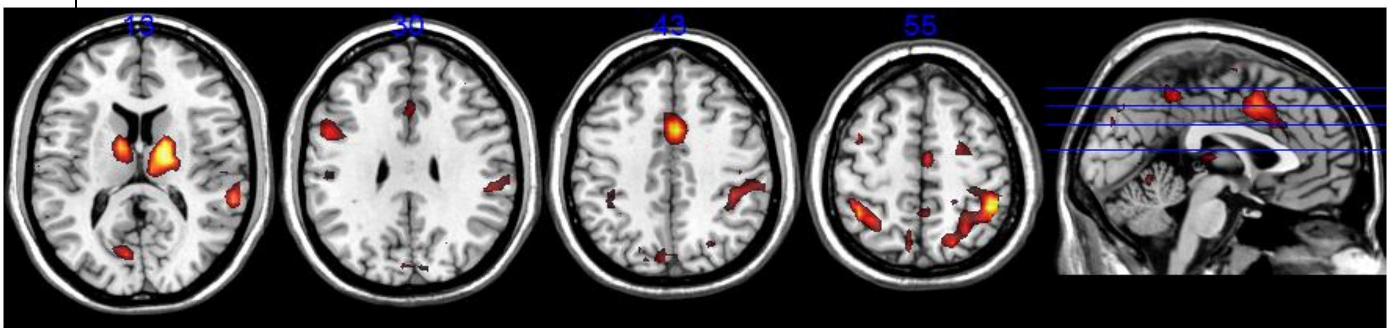

Figure 6. The second dimensional solution obtained from multiple-set canonical correlation analysis or equivalently the proposed approach under $\alpha = 1$ and $\beta = 0$.