HHS Public Access

Author manuscript

Neuroimage. Author manuscript; available in PMC 2018 April 01.

Published in final edited form as:

Neuroimage. 2017 April 01; 149: 63–84. doi:10.1016/j.neuroimage.2016.12.081.

A Family of Locally Constrained CCA Models for Detecting Activation Patterns in fMRI

Xiaowei Zhuang[a], **Zhengshi Yang**[a], **Tim Curran**[b], **Richard Byrd**[c], **Rajesh Nandy**[d], and **Dietmar Cordes**[a,b,*]

[a]Cleveland Clinic Lou Ruvo Center for Brain Health, Las Vegas, NV, 89106, USA

[b]Department of Psychology and Neuroscience, University of Colorado, Boulder, CO, 80309, USA

[c]Department of Computer Science, University of Colorado, Boulder, CO, 80309, USA

[d]School of Public Health, University of North Texas, Fort Worth, TX, 76107, USA

Abstract

Canonical correlation analysis (CCA) has been used in functional Magnetic Resonance Imaging (fMRI) for improved detection of activation by incorporating time series from multiple voxels in a local neighborhood. To improve the specificity of local CCA methods, spatial constraints were previously proposed. In this study, constraints are generalized by introducing a family model of spatial constraints for CCA to further increase both sensitivity and specificity in fMRI activation detection.

The proposed locally-constrained CCA (cCCA) model is formulated in terms of a multivariate constrained optimization problem and solved efficiently with numerical optimization techniques. To evaluate the performance of this cCCA model, simulated data are generated with a Signal-To-Noise Ratio of 0.25, which is realistic to the noise level contained in episodic memory fMRI data. Receiver operating characteristic (ROC) methods are used to compare the performance of different models. The cCCA model with optimum parameters (called optimum-cCCA) obtains the largest area under the ROC curve. Furthermore, a novel validation method is proposed to validate the selected optimum-cCCA parameters based on ROC from simulated data and real fMRI data. Results for optimum-cCCA are then compared with conventional fMRI analysis methods using data from an episodic memory task. Wavelet-resampled resting-state data are used to obtain the null distribution of activation.

For simulated data, accuracy in detecting activation increases for the optimum-cCCA model by about 43% as compared to the single voxel analysis with comparable Gaussian smoothing. Results from the real fMRI data set indicate a significant increase in activation detection, particularly in hippocampus, para-hippocampal area and nearby medial temporal lobe regions with the proposed method.

*Correspondence to: Dietmar Cordes, Ph.D. Cleveland Clinic Lou Ruvo Center for Brain Health, 888 W. Bonneville Ave, Las Vegas, NV, 89106. Phone: 1-702-483-6022; Fax: 1-866-372-2720. cordesd@ccf.org.

Keywords

Constrained Canonical Correlation Analysis (cCCA); functional Magnetic Resonance Imaging (fMRI); multivariate analysis; numerical optimization; episodic memory task

1. INTRODUCTION

Multivariate analysis methods have a broader range of uses than univariate analysis methods in functional Magnetic Resonance Imaging (fMRI) studies because these methods can simultaneously use a multitude of aspects in the data to determine activation patterns more reliably. The most common methods are based on independent component analysis (ICA) [Calhoun et al., 2009] and covariance and multiple-regression-derived analysis such as sparse partial least square (PLS) regression [Monteiro et al., 2015] and canonical correlation analysis (CCA). CCA is a statistical method that finds linear associations of two random variables so that the correlation between the variables is maximized [Hotelling, 1936], and can be defined by a multivariate extension of the general linear model [Thompson, 2005]. One of the first applications of CCA in neurological data analysis was carried out by Friman et al [2001] who applied CCA in local neighborhoods such as 3 3 (or similar) in-plane voxel regions to detect activation status in fMRI data during a mental calculation task. Other successful applications of CCA include fusion of multimodal neurological data (for a review, see [Calhoun and Sui, 2016]). Complimentary information from different modalities can be captured through joint multivariate data-driven approaches and therefore brain disease can be better investigated. For example, in studying schizophrenia, Sui et al applied a CCA+ ICA based model for multi-task data fusion [Sui et al., 2010] and a multi-set CCA analysis combining fMRI, electroencephalogram and structural MRI data [Sui et al ., 2014]. More recently, Levin-Schwartz et al [2016] proposed a principal component analysis (PCA) + CCA approach to jointly estimate the degree of similarities between multiple datasets with limited samples.

In this paper, we focused on improving the CCA models to detect activation maps more accurately from fMRI data. When CCA is applied locally in fMRI data, sensitivity in detecting activations is increased substantially if more than one voxel in the neighborhood was indeed activated. However, assigning activation status to a specific voxel in the neighborhood is less feasible because of the multivariate nature of the approach. Usually, for a 3 3 neighborhood CCA analysis, activation status is given to the center voxel to obtain maps with increased specificity whether or not the center voxel is active from a univariate standpoint. Several publications have shown the usefulness of this center-voxel assignment scheme, for example Friman et al [2001] and Nandy and Cordes [2004]. The most useful application of CCA is in obtaining activation maps for locations in the brain that are difficult to image, for example prefrontal and inferior temporal regions of the brain where susceptibility artifacts are usually strong and the signal-to-noise ratio (SNR) is smaller than in other cortical regions. Activation status in such regions are usually weak during memory-related task fMRI but critical for patients with amnestic mild cognitive impairment (aMCI) [Dickerson et al., 2004]. Local multivariate analysis may improve detection of activation in

areas where activations are weak [Nandy and Cordes, 2003] and therefore is more suitable in detecting activation during episodic memory tasks.

Ordinary CCA has been shown to yield a significant smoothing artifact, as activations of strongly active voxels tend to bleed into the neighboring voxels [Cordes et al., 2012a]. Therefore, ordinary CCA suffers from a low specificity and incorrectly classifies a voxel as active when there is a strong active voxel anywhere in the neighborhood irrespective of the activation state of the center voxel [Friman et al., 2001; Nandy and Cordes, 2004]. To address this problem of low specificity, a non-negative spatial weight constraint has been imposed to the CCA model [Friman et al., 2003; Ragnehed et al., 2009]. Recently, CCA has been improved further by considering stronger spatial constraints, called dominance constraints of the center voxel in the local neighborhood [Cordes et al., 2012b; Jin et al., 2012a]. Other improvements have been suggested by generalizing CCA in a non-linear context [Akaho, 2006; Hardoon et al., 2007] and by combining linear or non-linear relationships between different data sets [Dong et al., 2015a].

Spatial constraints are used to control the relationship between the weights of the center and all other neighboring voxels in each local neighborhood. Three suggested models use constraints $a_j > 0, \forall j$ [Friman et al., 2003], $\alpha_1 \geq \sum_{k=2}^{9} \alpha_k$ and $a_1 \geq \max_{k>1} a_k$ [Cordes et al., 2012b] where a_1 specifies the spatial weight of the center voxel and the $a_k\,(k > 1)$ are the weights of all other neighborhood voxels. Cordes et al [2012b] have shown that when a large spatial dominance constraint is applied to the center voxel $\left(\alpha_1 \geq \sum_{k=2}^{9} \alpha_k \quad \text{or} \quad \alpha_1 \geq \max_{k>1} \quad \alpha_k\right)$, weak localized activations could be detected with a significant reduction in the false positive rate (FPR). This method leads to an improved sensitivity for a given specificity.

However, with the spatial constraints suggested above, CCA methods still suffer from bleeding artifacts, especially in regions where SNR is large. In this study, we would like to further explore the spatial constraint space in CCA methods to see whether the local constrained CCA model can be improved and how much improvement is possible. An important theoretical interest is to investigate whether there is a more flexible and generalized model that covers the entire spatial constraint space. Therefore, the constraint in the CCA model could be determined based on the fMRI data set itself and thus our improved CCA model is more flexible in activation detection.

We address the above problem by proposing a family of spatial constrains for CCA (abbreviated as cCCA model), parameterized by two indices p and ψ in terms of $\alpha_1^p \geq \psi \quad \sum_{k=2}^{m} \alpha_k^p$ as the constraint. The two parameters p and ψ control the spatial weights of the center voxels (a_1) in relation to the local neighboring voxels (a_k, $k = 2, \ldots, m$). The new family-cCCA model incorporates previously proposed models when (p, ψ) = (1, 0) (non-negative spatial weight constraint [Friman et al., 2003, Cordes et al., 2012b], which we call non-negative-cCCA in this study), (p, ψ) = (1, 1) (sum constraint [Cordes et al., 2012b], which we call sum-cCCA in this study) and (p, ψ) = (∞, 1) (max constraint [Cordes et al., 2012b], which we call max-cCCA in this study), respectively. However, the solution of the

cCCA model is non-trivial. We use numerical optimization techniques to solve the cCCA model. Using simulations, we show the improvements of cCCA with optimum parameters p and ψ over standard mass-univariate analysis and previously proposed CCA-related models. We validate the selected optimum parameters of the family-cCCA model by using a novel validation technique based on area under the Receiver Operating Characteristic (ROC) curves. We apply the validated optimum-cCCA model to real fMRI data and demonstrate superior performance of cCCA over other analysis methods in activation detection. The optimum performance of cCCA is shown further by a classification and prediction for aMCI subjects and normal controls (NCs) using activation patterns in the medial temporal lobe (MTL) obtained from different analysis methods.

2. THEORY

2.1 Model Formulation

Let $\boldsymbol{Y} = (\boldsymbol{y_1}, \ldots, \boldsymbol{y_m}) \in \mathbb{R}^{t \times m}$ be a matrix of continuous variables representing the time courses (t time points) of m voxels (e.g., $m = 9$ for 3 × 3 regions in a 2D slice), and let $\boldsymbol{X} = (\boldsymbol{x_1}, \ldots, \boldsymbol{x_n}) \in \mathbb{R}^{t \times n}$ represent n functions used to model the blood oxygenation level-dependent (BOLD) response. Using cCCA, we find spatial weight vectors $\boldsymbol{\alpha} \in \mathbb{R}^{m \times 1}$ and $\boldsymbol{\beta} \in \mathbb{R}^{n \times 1}$, under certain spatial constraints, that maximize the correlation $\rho(\boldsymbol{\alpha}, \boldsymbol{\beta})$ between $\boldsymbol{Y\alpha}$ and $\boldsymbol{X\beta}$, which is also known as canonical correlation between $\boldsymbol{Y}$ and $\boldsymbol{X}$. Concretely, we maximize the multivariate function

$$\max_{\alpha,\beta} \rho(\alpha, \beta) = \frac{\operatorname{cov}(\boldsymbol{Y}\alpha, \boldsymbol{X}\beta)}{\sqrt{\operatorname{var}(\boldsymbol{Y}\alpha)\operatorname{var}(\boldsymbol{X}\beta)}} \quad (1)$$

defined for non-zero $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, under the following constraints on spatial coefficient vector $\boldsymbol{\alpha}$:

$$\begin{cases} \alpha_1 \geq 0, \ldots, \alpha_m \geq 0 \\ \alpha_1^p \geq \psi \sum_{k=2}^{m} \alpha_k^p \end{cases} \quad (2)$$

where $p > 0$ and $\psi > 0$ are two parameters of the model.

For large ψ, the spatial constraint converges to a delta function, which is equivalent to the single voxel analysis and for small ψ, the spatial constraint converges to the simple non-negative constraint. The other parameter p is well-defined for all positive values. Under the condition $(p, \psi) = (1, 0)$, $(p, \psi) = (1, 1)$ and $(p, \psi) = (\infty, 1)$, the family-cCCA model reduces to the non-negative constrained CCA (non-negative-cCCA), sum constrained CCA (sum-cCCA), and max constrained CCA (max-cCCA), respectively. Since for every $\boldsymbol{z} \in \mathbb{R}^{m \times 1}, \lim_{p \to \infty} \|\boldsymbol{z}\|_p = \max_i |z_i|$, the max-cCCA can be approximated with a finite large p ($p \gg 1$) and $\psi = 1$, i.e. $(p, \psi) = (32, 1)$ cCCA in this study.

2.1.1 Estimating the canonical correlation coefficient giving a set of spatial weights—

Let $\boldsymbol{C_{YY}}$ and $\boldsymbol{C_{XX}}$ be the within-set covariance matrix and $\boldsymbol{C_{YX}}$ be the between-set covariance matrix. The canonical correlation coefficient is then given by

$$\rho(\alpha,\beta) = \frac{\alpha' \boldsymbol{C_{YX}} \beta}{\sqrt{\alpha' \boldsymbol{C_{YY}} \alpha \beta' \boldsymbol{C_{XX}} \beta}}. \quad (3)$$

In practice, we estimate the covariance matrices $\boldsymbol{C_{YY}}$, $\boldsymbol{C_{XX}}$, $\boldsymbol{C_{YX}}$ by the average of N samples in variable $\boldsymbol{Y}$ and $\boldsymbol{X}$, denoted by $\boldsymbol{S_{YX}}$, $\boldsymbol{S_{YY}}$, $\boldsymbol{S_{XX}}$ respectively. If $\boldsymbol{Y}$ and $\boldsymbol{X}$ are the samples after the mean of each column has been subtracted from the corresponding column, we obtain:

$$\begin{cases} \boldsymbol{S_{YX}} = \frac{1}{N-1} \boldsymbol{Y}' \boldsymbol{X} \\ \boldsymbol{S_{YY}} = \frac{1}{N-1} \boldsymbol{Y}' \boldsymbol{Y} \\ \boldsymbol{S_{XX}} = \frac{1}{N-1} \boldsymbol{X}' \boldsymbol{X} \end{cases}. \quad (4)$$

The multivariate function that we use is:

$$\rho(\alpha,\beta) = \frac{\alpha' \boldsymbol{S_{YX}} \beta}{\sqrt{\alpha' \boldsymbol{S_{YY}} \alpha \beta' \boldsymbol{S_{XX}} \beta}}. \quad (5)$$

This sample expression of the canonical correlation coefficient is defined everywhere except for $\boldsymbol{\alpha} = \mathbf{0}$ and $\boldsymbol{\beta} = \mathbf{0}$. Because $\boldsymbol{Y}$ and $\boldsymbol{X}$ are random matrices with coordinates being continuous and linearly independent, their sample covariance matrices are positive-definite with probability equal to 1 [Siotani et al., 1985], and therefore we obtain $\boldsymbol{\alpha}'\boldsymbol{S_{YY}}\boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{S_{XX}}\boldsymbol{\beta} > 0$ with probability equal to 1. Like the regular Pearson correlation coefficient, the canonical correlation coefficient is invariant under scalar multiplication of its arguments. Appendix A. 1 proves the existence of the maximum of the sample version of $\rho(\boldsymbol{\alpha}, \boldsymbol{\beta})$. Therefore, given N samples of $\boldsymbol{Y}$ and $\boldsymbol{X}$, the optimization problem combining Eq. (5) and Eq. (2) becomes:

$$\text{cCCA:} \quad \max_{\alpha,\beta} = \frac{\alpha' \boldsymbol{S_{YX}} \beta}{\sqrt{\alpha' \boldsymbol{S_{YY}} \alpha \beta' \boldsymbol{S_{XX}} \beta}} \quad \text{w. r. t.} \quad \begin{cases} \alpha_1 \geq 0, \ldots, \alpha_m \geq 0 \\ \alpha_1^p \geq \psi \sum_{k=2}^{m} \alpha_k^p \end{cases}. \quad (6)$$

CCA is a multivariate form of the general linear model [Thompson, 2005]. Therefore, we present a constrained multivariate multiple regression model which is mathematically equivalent to Eq. (6) in the following section.

2.1.2 Constrained multivariate multiple regression model—Using the same notation, CCA can be converted into an equivalent constrained multivariate multiple regression model:

$$\boldsymbol{Y}\alpha \;=\; \boldsymbol{X}\boldsymbol{B}\alpha \;+\; \boldsymbol{E}\alpha, \quad (7)$$

where $\boldsymbol{Y}$ and $\boldsymbol{X}$ are the data and design matrix, respectively, $\boldsymbol{a}$ is the spatial weight vector of $\boldsymbol{Y}$, $\boldsymbol{B}$ is the matrix of regression weights and $\boldsymbol{E}$ is the residual error matrix of the linear regression model $\boldsymbol{Y} = \boldsymbol{XB} + \boldsymbol{E}$. The least square solution of $\boldsymbol{B}$ equals $(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$. Furthermore, linear weights of $\boldsymbol{X}$ in cCCA ($\boldsymbol{\beta}$) can be represented as $\boldsymbol{B}\boldsymbol{a}$ which equals $(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}\boldsymbol{a}$. Therefore, cCCA can be rewritten in a constrained least square (cLS) format by

$$\mathrm{cLS}{:}\min_{\alpha}\| \boldsymbol{Y}\alpha \;-\; \boldsymbol{X}\beta \|^2 \text{w. r. t.} \begin{cases} \alpha_1 \;\geq\; 0, \ldots, \alpha_m \;\geq\; 0 \\ \alpha' \boldsymbol{S}_{\boldsymbol{YY}} \alpha \;=\; 1 \\ \beta \;=\; \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{Y}\alpha \\ \alpha_1^p \;\geq\; \psi \sum_{k=2}^{m} \alpha_k^p \end{cases} . \quad (8)$$

Both cCCA and cLS are equivalent under conditions $\boldsymbol{a}'\boldsymbol{S}_{\boldsymbol{YY}}\boldsymbol{a} = 1$ and $\boldsymbol{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}\boldsymbol{a}$, i.e., they have the same optimal solution(s). A formal proof of this proposition is presented in the Appendix A.2.

2.1.3 Size of local neighborhood—In analyzing small pixel regions such as 3×3 local neighborhoods, family-cCCA can reduce the number of false positives because of its improved spatial filter kernel, thus high specificity can be achieved particularly in small activated areas such as the hippocampus. Furthermore, we use a 3×3 2D in-plane neighborhood rather than a 3×3×3 3D voxel region in the family-cCCA model due to Echo Planar Imaging (EPI) limitations of our data acquisition. We are interested in obtaining functional detail in cross sections of the hippocampus rather than in the longitudinal hippocampal direction. Since our chosen slice thickness is much larger than the in-plane voxel size, a 3×3 2D neighborhood is optimal for the analysis. An analysis in only two dimensions is also optimal for gray matter in general if cortical unfolding is being used since the cortex can then be represented by a 2D sheet.

2.2 Solving the model

The above section 2.1.2 has shown that the cCCA model is equivalent to a constrained multivariate multiple regression model, and the linear weights of $\boldsymbol{X}$ in cCCA ($\boldsymbol{\beta}$) can be represented as $\boldsymbol{B}\boldsymbol{a}$ which equals $(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}\boldsymbol{a}$. Therefore, the cCCA model with a 3×3 2D in-plane neighborhood can be rewritten as:

$$\max_{\alpha} \frac{\alpha' \boldsymbol{S_{XY}} \beta}{\sqrt{\alpha' \boldsymbol{S_{YY}} \alpha \beta' \boldsymbol{S_{XX}} \beta}} \text{w. r. t.} \begin{cases} \alpha_1 \geq 0, \ldots, \alpha_9 \geq 0; \\ \alpha_1^p \geq \psi \sum_{k=2}^{9} \alpha_k^p \\ \beta = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{X}' \boldsymbol{Y} \alpha \end{cases}, \tag{9-1}$$

which is equivalent to:

$$\min_{\alpha} \| \boldsymbol{Y}\alpha - \boldsymbol{X}\beta \|^2 \text{w. r. t.} \begin{cases} \alpha_1 \geq 0, \ldots, \alpha_9 \geq 0 \\ \alpha_1^p \geq \psi \sum_{k=2}^{9} \alpha_k^p \\ \beta = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{X}' \boldsymbol{Y} \alpha \\ \alpha' \boldsymbol{S_{YY}} \alpha = 1 \end{cases}. \tag{9-2}$$

We solve the cCCA model (Eq. (9)) from two perspectives:

(i). Das and Sen [1994] showed that solution to the non-negative-cCCA equals an unconstrained CCA solution where the weights ($\boldsymbol{a}$) to one or several original voxels are chosen to be zero. Furthermore, a search needs to be carried out over all possible $\boldsymbol{a}$ configurations where one or more coordinates of $\boldsymbol{a}$ are equal to zero. The same idea is true for the cCCA problem. In section 2.2.1, we describe the sub-spaces of the original cCCA problem and the numerical optimization technique used to solve these sub-problems.

(ii). Eq. (9-1) and Eq. (9-2) can also be treated as an optimization problem with ten inequality constraints. Note that the formed optimization problem is not globally convex and has many local extrema. Various optimization algorithms have been introduced to solve these problems [Fletcher and Reeves, 1964; Soltanian-Zadeh et al., 1994; Shanno, 1985; Nocedal and Wright, 2006; Luenberger and Ye, 1984]. In section 2.2.2, we describe techniques used to find the global solution by either incorporating constraints into the objective function (Augmented Lagrangian method [Hestenes, 1969]) or partitioning the original variables into dependent and independent groups according to active constraints (Reduced Gradient method [Lasdon et al., 1978]).

2.2.1 Sub-space problems—Since the feasible solution of the model must include the center voxel ($a_1 > 0$), we consider each of the 2^{9-1} subspaces J where $J \subseteq \{2, \ldots 9\}$, and denote J^c as its complement. For instance, if $J = \{2, 4, 7\}$, then $J^c = \{3, 5, 6, 8, 9\}$. We replace by $a_1 \geq 0, \ldots, a_9 \geq 0$ by $\forall j \in J$, $a_j > 0$ and $\forall j \in J^c$, $a_j = 0$. In this way, the original cCCA problem Eq. (9) is replaced by 256 sub-problems:

$$\mathrm{cCCA}\,(J):\quad \max_{\alpha}\frac{\alpha'\boldsymbol{S_{YX}}\beta}{\sqrt{\alpha'\boldsymbol{S_{YY}}\alpha\beta'\boldsymbol{S_{XX}}\beta}}\text{w. r. t.}\begin{cases}\forall j \in J,\alpha_j > 0\\ \forall j \in J^c,\alpha_j = 0\\ \alpha_1^p \geq \psi\sum_{k=2}^{9}\alpha_k^p\\ \beta = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{Y}\alpha\end{cases}. \tag{10}$$

cCCA can hence be solved by considering 256 configurations in each neighborhood containing the center voxel. If $\alpha_j = 0$, we can treat this case by deleting the j^{th} row of $\boldsymbol{S_{YX}}$ and similarly the j^{th} row and column in $\boldsymbol{S_{YY}}$. In this way, the problem to solve is:

$$\mathrm{cCCA}\,(J):\quad \max_{\alpha}\frac{\alpha'\boldsymbol{S_{YX}}\beta}{\sqrt{\alpha'\boldsymbol{S_{YY}}\alpha\beta'\boldsymbol{S_{XX}}\beta}}\text{w. r. t}\begin{cases}\forall j \in J,\alpha_j > 0\\ \alpha_1^p \geq \psi\sum_{k=2}^{9}\alpha_k^p\\ \beta = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{Y}\alpha\end{cases}. \tag{11}$$

This problem can be further reduced to two cases whether the constraint $\alpha_1^p \geq \psi\sum_{k=2}^{9}\alpha_k^p$ holds strictly or not:

$$cCCA\quad strict\quad case\text{:}\quad \max_{\alpha}\frac{\alpha'\boldsymbol{S_{YX}}\beta}{\sqrt{\alpha'\boldsymbol{S_{YY}}\alpha\beta'\boldsymbol{S_{XX}}\beta}}\text{w. r. t}\begin{cases}\forall j \in J,\alpha_j > 0\\ \alpha_1^p > \psi\sum_{k=2}^{9}\alpha_k^p\\ \beta = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{Y}\alpha\end{cases}$$

(12-1)

and

$$cCCA\quad equal\quad case\text{:}\quad \max_{\alpha}\frac{\alpha'\boldsymbol{S_{YX}}\beta}{\sqrt{\alpha'\boldsymbol{S_{YY}}\alpha\beta'\boldsymbol{S_{XX}}\beta}}\text{w. r. t}\begin{cases}\forall j \in J,\alpha_j > 0\\ \alpha_1^p = \psi\sum_{k=2}^{9}\alpha_k^p\\ \beta = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{Y}\alpha\end{cases}.$$

(12-2)

In practice we solve Eq. (12-1) and Eq. (12-2) by the following method:

2.2.1.1 cCCA, strict case: The $\boldsymbol{a}$ that maximizes ρ from the solutions of $\nabla\rho = 0$ (see Appendix A. 3) and satisfies $\forall j \in J$, $a_j > 0$ and $\alpha_1^p > \psi\sum_{k=2}^{9}\alpha_k^p$ forms the strict case solution for ρ;

2.2.1.2 cCCA, equal case: *(i). A solution when* $p = 1$. Cordes et al [2012b] gives an efficient solution when $p = 1$, $\psi = 1$, which can be generalized to every $\psi > 0$ with $p = 1$. In this case, let matrix be $\boldsymbol{M} \in \mathbb{R}^{m\times m}$ be defined with diagonal elements $M_{ii} = 1$ for every i, and off-diagonal elements $M_{1j} = \psi$ for every $j > 1$, and 0 everywhere else. In Appendix A.4 we show that Eq. (12-2) is equivalent to the following optimization problem:

$$\text{cCCA:}\quad \max_{\alpha}\frac{\alpha'\boldsymbol{S_{YX}}\beta}{\sqrt{\alpha'\boldsymbol{S_{YY}}\alpha\beta'\boldsymbol{S_{XX}}\beta}}\text{w. r. t}\begin{cases} p &= 1\\ \phi &\geq 0\\ \alpha &= \boldsymbol{M}\phi\end{cases}. \tag{13}$$

Since the covariance matrix satisfies the property $T\boldsymbol{S_{UV}}R = \boldsymbol{S_{TU\,RV}}$ for random vectors $\boldsymbol{U}, \boldsymbol{V}$ and linear transformations T, R (see Appendix A. 5), it henceforth follows that the solution of Eq. (13) can be obtained by first replacing the data $\boldsymbol{Y}$ by $\widetilde{\boldsymbol{Y}} = \boldsymbol{YM}$, and then solving the optimization problem:

$$\text{cCCA:}\quad \max_{\phi}\frac{\phi'\boldsymbol{S_{\tilde{Y}X}}\beta}{\sqrt{\phi'\boldsymbol{S_{\tilde{Y}\tilde{Y}}}\phi\beta'\boldsymbol{S_{XX}}\beta}}\text{w. r. t}\begin{cases} p &= 1\\ \phi &\geq 0\\ \alpha &= \boldsymbol{M}\phi\end{cases}. \tag{14}$$

In other words, for $p = 1$, cCCA can be reduced to solving cCCA with non-negative constraints as in Eq. (14), which in turn can be reduced to solving 256 unconstrained ordinary CCA problems [Das and Sen., 1994]. Since CCA is equivalent to an eigenvalue problem, the case $p = 1$ and arbitrary ψ reduces to solving 256 eigenvalue problems that can be handled efficiently.

(ii) Solutions when $p \neq 1$. We use the Broyden–Fletcher–Goldfarb–Shanno (BFGS) with self-scaling [Shanno, 1985, Nocedal and Wright, 2006] method, which is an iterative gradient descent method, to solve Eq. (12-2) given a suitable starting point of $\boldsymbol{a}$. The search direction in each step is facilitated with the backtracking line search algorithm [Nocedal and Yuan, 1998]. To avoid the trivial solution of $\boldsymbol{a} = \boldsymbol{0}$ in our objective function, we arbitrarily set the weight of the first neighboring voxel (a_2) to be 1. In each optimization step, a_1 is replaced by $\sqrt[p]{\psi\sum_{k=2}^{9}\alpha_k^p}$ due to the cCCA constraint. The non-negative constraint is satisfied by substituting $\{\alpha_3, \alpha_4, \ldots, \alpha_9\} = \left\{\theta_3^2, \theta_4^2, \ldots, \theta_9^2\right\}$, i.e. spatial weights of all other neighboring voxels are represented by θ_i^2.

A self-scaling factor γ_k is applied to the Hessian matrix in each optimization step for faster convergence. With this self-scaling factor, the BFGS algorithm possesses favorable eigenvalue structure at each step, i.e. all the eigenvalues of the Hessian matrix at iteration k

($\boldsymbol{H}_k$) are spread below or above unity [Nocedal and Wright, 2006]. Therefore, rapid progress can be made at every stage. For example, in optimizing a 2D Rosenbrock function [Rosenbrock, 1960] starting at (1000, 1000), regular BFGS takes approximately 3000 iterations to converge whereas BFGS with γ_k reduces the number of steps towards optimum solution to around 900.

Finally, the optimal solution (maximum ρ) is obtained from the possible solutions of 2.2.1.1 and 2.2.1.2. To illustrate all steps in solving the cCCA problem by considering 256 separate configurations, a comprehensive flow chart is presented in Fig. 1.

Figure 1: flow chart in solving the cCCA problem with 256 separate configurations.

2.2.2. Other algorithms to solve the cCCA problem

2.2.2.1. Augmented Lagrangian: The Augmented Lagrangian algorithm incorporates constraints into the objective function in terms of additional quadratic penalty terms and explicit Lagrangian multiplier estimates [Nocedal and Wright, 2006]. Based on a proximal point approach [Rockafellar, 1976], we obtain the unconstrained form of the augmented Lagrangian sub-problems for constrained problems. The formed augmented Lagrangian function is then minimized (or maximized) with BFGS techniques. A detailed formulation of the augmented Lagrangian for cCCA is shown in Appendix A.6.

2.2.2.2. Reduced Gradient: From a computational viewpoint, the reduced gradient method is related to the simplex method in that the optimized variables are partitioned into dependent and independent sets [Luenberger and Ye, 1984]. In the reduced gradient method, inequality constraints are converted to equality constraints by adding non-negative slack variables [Lasdon et al, 1978]. We modified Eq. (9) for the reduced gradient as follows:

$$\begin{cases} \text{Objective function:} \quad \mathrm{g}(\alpha) = \max_{\alpha} \dfrac{\alpha' \boldsymbol{S}_{\boldsymbol{YX}} \left(\left(\boldsymbol{X}'\boldsymbol{X} \right)^{-1} \boldsymbol{X}'\boldsymbol{Y} \right) \alpha}{\sqrt{\alpha' \boldsymbol{S}_{\boldsymbol{YY}} \alpha \left(\left(\boldsymbol{X}'\boldsymbol{X} \right)^{-1} \boldsymbol{X}'\boldsymbol{Y}\alpha \right)' \boldsymbol{S}_{\boldsymbol{XX}} \left(\left(\boldsymbol{X}'\boldsymbol{X} \right)^{-1} \boldsymbol{X}'\boldsymbol{Y}\alpha \right)}} \\ \text{Subject to:} \quad \mathrm{h}(\alpha, \upsilon) = \psi \sum_{k=2}^{9} \alpha_k^p - \alpha_1^p + \upsilon = 0 \\ \text{Border constraints:} \quad \alpha_1 \geq 0, \ldots, \alpha_9 \geq 0, \upsilon \geq 0 \end{cases}$$

(15)

where ν is the slack variable added so that the inequality constraint is transformed into equality constraint. The optimized variable set becomes [a_1, …, a_9, ν]. We keep [a_2, …, a_9, ν] to be the independent variables and a_1 to be the dependent one. The reduced gradient (rGD) at each step is calculated as in Abadie and Carpentier [1969]. Iteration updates are taken by changing independent variables in the direction of negative rGD with those independent variables on their boundary held fixed if the movement would violate bounds. To continuously satisfy the equality constraints, dependent variables first move straightly

along the tangent plane of the surface and return to the constraint surface later through a correction step.

2.3 Selection of a suitable algorithm

In the above section (section 2.2), we have described three different techniques in solving the cCCA problem. We use the following algorithms to solve Eq. (9-1) and Eq. (9-2).

(a) BFGS with 1 starting point for entire 3×3 region;

(b) BFGS with 1 starting point for each 256 sub-problems;

(c) Augmented Lagrangian with 1 starting point for entire 3×3 region;

(d) Reduced Gradient with 1 starting point for entire 3×3 region;

(e) educed Gradient with 9 starting points for entire 3×3 region.

To select a suitable algorithm out of the listed algorithms, we evaluate both accuracy and efficiency of each algorithm.

2.3.1 Accuracy—For each 3×3 neighborhood, we define the largest correlation output across all the algorithms to be the benchmark correlation $\rho_{\text{benchmark}}$. Then the accuracy for that neighborhood is defined to be the probability of the output correlation from one specific algorithm, $\rho_{\text{algorithm}}$, that is within the tolerance of $\rho_{\text{benchmark}}$. In other words, given a small number $\epsilon>0$ as tolerance, we define accuracy of a specific algorithm for one 3×3 neighborhood as:

$$\text{Accuracy} = Prob\left[\left|\rho_{\text{algorithm}} - \rho_{\text{benchmark}}\right| \leq \epsilon\right]. \quad (16)$$

The average accuracy of all neighborhoods is then defined as the accuracy of that specific algorithm.

2.3.2 Efficiency—We time each algorithm in solving one 3×3 neighborhood, that is, determine the activation status for one center voxel. The computation time for each algorithm averaging over all 3×3 neighborhoods is then defined as the efficiency for that specific algorithm. We list the efficiency and accuracy of all five algorithms for a fixed (p, ψ) = (2,1) averaging over 700 3×3 neighborhoods in Table.1.

Table. 1: Efficiency and accuracy for each method solving cCCA problem with (p, ψ) = (2,1).

2.3.3 Trade-off between accuracy and efficiency—As shown in Table.1, given ϵ = 0.001, algorithm (b) BFGS with 1 starting point for each 256 sub-problems achieves an accuracy of 100%. However, it needs several seconds to determine the activation status of one voxel. If we relax our tolerance by 10% (ϵ =0.01), method (e) Reduced Gradient with 9 starting points for entire 3×3 region reaches a 100% accuracy while bringing the calculation time down to less than one second. Therefore, method (b) is the optimum algorithm if highest accuracy is of concern whereas method (e) is optimal with the highest efficiency.

2.4 Statistical analysis

After the optimum $\boldsymbol{a}$ is obtained, we compute Wilk's Λ statistic [Mardia et al., 1979; Friman et al., 2001; Jin et al., 2012b] to classify voxels to be active for a specific contrast of interest $\boldsymbol{c}$ using:

$$\begin{aligned} \beta &= \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{Y}\alpha \\ E &= (\boldsymbol{Y}\alpha - \boldsymbol{X}\beta)'(\boldsymbol{Y}\alpha - \boldsymbol{X}\beta) \\ H &= (\boldsymbol{c}\beta)'\left(\boldsymbol{c}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{c}'\right)^{-1}(\boldsymbol{c}\beta) \\ \Lambda &= \frac{E}{E+H}. \end{aligned} \tag{17}$$

Here, the error is labeled E, the contrast vector $\boldsymbol{c}$, and the hypothesis H. We further compute a monotonically increasing statistic $F = \left(\frac{1-\Lambda}{\Lambda}\right)\frac{\upsilon_E}{\upsilon_H}$, where ν_E and ν_H denote the degree of freedom of E and H, respectively [Cordes et al., 2012b], so that the probability of activation increases with the magnitude of F.

2.4.1 Calculation of the null distribution—To obtain the null distribution, all of the above analysis is repeated on wavelet-resampled resting-state time courses [Bullmore et al., 2001, Breakspear et al., 2004] until a stable maximum statistic is obtained. Breakspear et al. [2004] have demonstrated that constrained resampling of the resting-state data in the wavelet domain allows construction of bootstrapped data with the following essential properties: (1) spatial and temporal correlations are preserved; (2) the irregular geometry of the intracranial images is maintained; (3) there is adequate type I error control; and (4) expected experiment-induced correlations are included. Therefore, the spatiotemporal resampled data in wavelet domain can be further used in testing the null hypothesis. We use the distribution of the maximum statistic to correct for multiple comparisons and obtain the family-wise-error-rate (FWE) [Nandy and Cordes, 2007].

2.5 Validating the family-cCCA model.

To verify that the performance of the family-cCCA model on real fMRI data is consistent with the performance on simulated data when the same parameter set is applied, we propose a novel validation method. This method inspects the relationship and consistency between the areas under the ROC curves (AUC) from simulated data and from real fMRI data when using the family-cCCA model to detect activation. Here we consider 42 representative family-cCCA models with parameters $p \in \{0.5,1,2,4,8,16,32\}$ in combination with $\psi \in \{1,2,4,8,16,32\}$.

2.5.1 AUCs for simulated data—ROC curves are straight-forward to compute when applying the family-cCCA model to simulated data since the activation truth is well defined. The 42 cCCA models yield 42 ROC curves. AUCs for a FPR in the interval [0, 0.1] are then computed, which we denote as $\mu_{i=1,\ldots,42}$. Then, the parameter combination (p, ψ) that gives the largest μ_i is selected as the optimum parameter combination of the family-cCCA model.

2.5.2 AUCs for real fMRI data: Generate apparent ROC curves—Computing ROC curves when applying the family-cCCA model to detect activation in real fMRI data is difficult due to the unknown activation status. Here, we propose the following method to generate "apparent ROC curves". Each of the cCCA methods are applied to real fMRI data. We set a certain threshold to define the 'apparent ground truth' and label *active (**A**)* and *inactive (**I**)* neighborhoods using the following schema: A 3×3 neighborhood is labeled ***A*** if the center voxel is labeled as active with current cCCA method for threshold T. Similarly, a 3×3 neighborhood is labeled ***I*** if the center voxel is labeled as inactive with current cCCA method for threshold T. Next, we use a sequence of thresholds $t_1 \ldots t_T$ in an ascending order within the range of the image statistics. We choose n pairs of 3×3 neighborhoods (***R, R′***) randomly from set A, and for each pair of (***R, R′***) we create a synthetic 3×3 neighborhood ***R″*** with the center voxel from ***R*** and the neighborhood from ***R′*** (without center voxel). The assumption here is that if center voxels in ***R*** and ***R′*** are truly active or inactive, then ***R″*** should be declared active or inactive, respectively. We then apply the same cCCA method to ***R″***. For each of the n pairs of neighborhoods (***R, R′***) from A, true positive (TP) or false negative (FN) declaration can occur if the current cCCA method labels the center voxel of ***R″*** as active or inactive, respectively. Similarly, the same cCCA method is applied to ***R″*** for each of the n pairs of neighborhoods (***R, R′***) from I. A false positive (FP) or true negative (TN) label is assigned if current cCCA method labels the center voxel of ***R″*** as active or inactive, respectively. A ROC curve is then constructed using these active or inactive declarations with respect to each threshold t_j. The area underneath the ROC curve, integrating over the same false positive range as in simulated data, is calculated and denoted as $\nu_{i=1,\ldots,42}$. The process of generating apparent ROC curves is illustrated in Fig. 2.

Figure 2. Generate apparent ROC curves for real fMRI data.

2.5.3 Model validation—The family-cCCA model is finally validated using the Pearson correlation and Bland Altman plot [Bland JM and Altman DG., 1986] between the sequences $\mu_{i=1,\ldots,42}$ from simulated data and $\nu_{i=1,\ldots,42}$ from real fMRI data. A strictly monotonic relationship between $\mu_{i=1,\ldots,42}$ and $\nu_{i=1,\ldots,42}$ is expected.

The correlation between two set of variables shows the extent to which changes in the value of one variable are correlated to the changes in the value of the other [Udovičić et al., 2007]. A larger positive correlation demonstrates a higher linear dependency between the cCCA model performance on real fMRI data and our simulation. Therefore, the optimum parameters selected from the simulation can be considered also optimal in analyzing real fMRI data.

If there is a strictly monotonic relationship between $\mu_{i=1,\ldots,42}$ and $\nu_{i=1,\ldots,42}$, a perfect agreement should be observed after each variable is normalized by its mean value. The Bland Altman plot quantifies the agreement between these two set of variables by studying the mean difference and constructing limits of agreement [Giavarina D., 2015]. We plot the differences between the normalized two set of variables against the average between them. 95% limits of agreement for each comparison (mean ± 1.96 standard deviation of the difference) are also plotted, which tells us how far apart measurement by two methods are

acceptable. We consider the two set of variables consistent if the mean difference is 0 and all the differences are within this 95% limit.

3. MATERIAL AND METHODS

3.1 Imaging

FMRI data of 14 subjects (7 subjects diagnosed with amnestic mild cognitive impairment (aMCI) and 7 normal controls (NC)), were collected with Institutional Review Board approval using a 3.0 T GE scanner. Diagnosis of aMCI was made by trained professionals based on Petersen Criteria [Peterson et al., 2001]. Subjects in two groups (aMCI and NC) were matched in age, education and right-handedness.

The EPI protocol had the following specifications: TR 2000 ms, TE 30 ms, parallel imaging factor of 2, 25 slices (coronal oblique, perpendicular to the long axis of hippocampus), slice thickness/gap=4.0 mm/1.0 mm, 288 time frames, in-plane resolution 96×96 interpolated to 128×128, yielding a voxel size of $1.72\times1.72\times5$ mm^3. High resolution structural images were also acquired including a T1-weighted image ($0.43\times0.43\times1$ mm^3) and a coplanar T2-weighted image ($0.43\times0.43\times2.5$ mm^3).

For each subject, task activation episodic memory data and resting-state data (same EPI parameters, subject had eyes closed) were used in the analysis. Specifically, the memory task involved encoding and recognition activity using stimuli of human faces paired with occupations. The task consisted of six periods of encoding, distraction (control), recognition activity, and short instructions where words on the screen reminded subjects of the task component ahead. Four regressors (instruction, encoding, distraction, recognition) were specified in our design matrix by convolving the timing of each condition with the canonical hemodynamic response function (HRF). The encoding task consisted of seven novel visual stimuli, and the recognition task consisted of fourteen stimuli, half novel (“new stimuli”) and half identical (in random order) to the items seen in the previous encoding task (“old stimuli”). An active control task (button press responding to the letter “Y” or “N”) was used as a distraction task between each pair of encoding and recognition activity because simple fixation or rest is known to induce activations in the default mode, which could potentially induce activations in the medial temporal lobe (MTL) [Raichle and Snyder, 2007].

Figure 3: Experimental design.

3.2 Preprocessing

First 10 seconds of the EPI data acquisition (5 volumes) were removed to avoid incomplete steady-state magnetization. All other time frames were slice-timing corrected and realigned to the mean EPI image in SPM12 (http://www.fil.ion.ucl.ac.uk/spm/). The data were high-pass filtered with a cut-off frequency of 1/120 Hz to remove temporal drift [Holmes et al., 1997]. For the cCCA analysis, no spatial smoothing was applied. For comparison with mass-univariate analysis, Gaussian in-plane spatial smoothing with full-width-at-half-maximum (FWHM) of 2.24 voxels (3.8 mm) was used for both task-related and resting-state data. This FWHM was used to compare cCCA results applied to a 3×3 pixel area. A justification for this choice is provided in Appendix A.7.

3.3 Simulation

To compute ROC curves to evaluate and compare the different models indexed by (p, ψ), we simulated groups of 3×3 voxels with active or inactive center voxel to mimic real fMRI data sets. Three key points to carry out the simulation are: (1) how many of the 8 neighbors are assumed to be active, (2) how time courses of each voxel are simulated so that both spatial and temporal correlations are kept, and (3) how a suitable SNR is obtained.

3.3.1 Determine number of active neighbors—To establish how many of the 8 neighboring voxels should be simulated as active, we looked at real data. Single voxel analysis was applied to real fMRI data and voxels with top 5% significance were labeled as active. For every active or inactive voxel, we counted how many of the neighboring voxels were also active to obtain the empirical distribution of the number of active neighbors given an active or inactive center voxel separately. Simulated spatial activation patterns were then generated on a 39×39 grid following this empirical distribution. Scenarios of 3×3 neighborhoods without any activation voxels were also included in the simulation as in real fMRI data.

3.3.2. Simulate time courses of each voxel—We constructed the simulated time courses by adding pure noise $\boldsymbol{y}_{\text{null}}$ to pure activation time courses $\boldsymbol{y}_{\text{active}}$ with a noise fraction f that is close to real fMRI data [Cordes et al., 2012b]:

$$\boldsymbol{y}_{\text{simulated}} = \begin{cases} (1-f)\,\boldsymbol{y}_{\text{active}} + f\,\boldsymbol{y}_{\text{null}} & \text{active voxel} \\ \boldsymbol{y}_{\text{null}} & \text{inactive voxel} \end{cases}. \quad (18)$$

Note that both $\boldsymbol{y}_{\text{active}}$ and $\boldsymbol{y}_{\text{null}}$ were variance normalized.

Both $\boldsymbol{y}_{\text{active}}$ and $\boldsymbol{y}_{\text{null}}$ were acquired from the real fMRI time series to have realistic temporal autocorrelations in the simulated data. The vector $\boldsymbol{y}_{\text{active}}$ was the time course from the most significant voxel in single voxel analysis. The vectors $\boldsymbol{y}_{\text{null}}$ were obtained from wavelet-resampled resting-state data [Bullmore et al., 2001, Breakspear et al., 2004]. To ensure the same spatial correlation as in real fMRI data, a Gaussian Point Spread Function (PSF) was first estimated from the wavelet-resampled resting-state time series and then applied to the simulated time series.

3.3.3. Determine SNR—To compute the SNR, we use the same definition as in Cordes et al [2012b]:

$$\text{SNR} = \frac{1-f}{f} \times \left(\frac{\Sigma\lambda_{\text{i}}}{\Sigma\varsigma_i}\right)^{\frac{1}{2}}, \quad (19)$$

where λ_{i} and ς_i were the eigenvalues of the covariance matrix of $\boldsymbol{y}_{\text{active}}$ and $\boldsymbol{y}_{\text{null}}$. Since both $\boldsymbol{y}_{\text{active}}$ and $\boldsymbol{y}_{\text{null}}$ were variance normalized in our simulation, SNR only depended on the noise fraction f. To achieve the noise fraction f as in real fMRI data, we applied single voxel analysis to simulated data for f from 0 to 1 with a step size of 0.05. The average correlation

value with a significance level of $p<0.05$ (uncorrected), was compared with the average correlation value of the same significance level acquired from real fMRI time series. The noise fraction, *f*, at which the mean correlation value from simulation matches the real data was chosen as the noise fraction for our simulation, yielding a corresponding $\mathrm{SNR}=\frac{1-f}{f}$. To compare the performance of each method under low SNR case, another simulation with a higher noise fraction was also generated.

3.4 Smoothing artifact

Using simulated data, the smoothing artifact of the family-cCCA model was evaluated as a function of noise levels and analysis methods. We generated 7×7 neighborhoods with the true activation only occurring within the 3×3 pixel regions. The centermost voxel was always kept active and the number of active neighbors in the 3×3 regions ranged from 1 to 8. Time series were simulated as described in section 3.3.2 with noise fraction *f* ranging from 5% to 95% in steps of 10%. For each combination of (*s, f*), 100 7×7 neighborhoods were generated and then analyzed with the following methods: (1) unconstrained CCA, (2) non-negative-cCCA (covered by (p, ψ) = (1, 0) in the family-cCCA model), (3) sum-cCCA (covered by (p, ψ) = (1, 1) in the family-cCCA model), (4) family-cCCA with (p, ψ) = (1,2), (5) family-cCCA with (p, ψ) = (1,4), (6) family-cCCA with (p, ψ) = (1,6), (7) family-cCCA with (p, ψ) = (1,8), (8) single voxel analysis (covered by (p, ψ) = (1, ∞) in the family-cCCA model).

We limited our analysis to the center 5×5 region in each simulated 7×7 neighborhood so that every voxel when analyzed with CCA-related methods all had 8 neighbors. We defined the active ratio (AR) for each method in analyzing a specific neighborhood as:

$$\mathrm{AR} = \frac{\text{Number of active voxel: detected}}{\text{Number of active voxel: simulated}} \begin{cases} = 1 & \text{No artifact} \\ > 1 & \text{Smoothing artifact} \\ < 1 & \text{Less detected activation} \end{cases} \quad (20)$$

To determine the smoothing artifact for different analysis method under various noise levels, an average AR $\left(\overline{AR}\right)$ was calculated over 100 simulated neighborhoods for each combination of (*s, f*). Therefore, for each analysis method at a specific noise level, there are 9 $\overline{AR}$ values corresponding to 9 cases of number of active neighbors in the center 3×3 neighborhood ($s = \{0,1,\ldots,8\}$). $\overline{AR}$ was further averaged over these 9 cases based on the probability of each case showing up in all 2^{9-1} configurations in a 3×3 neighborhood as:

$$\overline{\overline{AR}} = \sum_{s=0}^{8} \left(\frac{\binom{8}{s}}{256} \times \overline{AR}_s \right). \quad (21)$$

The acquired $\overline{\overline{AR}}$ was used to represent the smoothing artifact for each analysis method at different noise levels.

3.5 Data analysis

The following benchmark models were defined and applied to both simulated data and real fMRI data: (1) Mass-univariate analysis, i.e. Single Voxel + Gaussian Smoothing (SV+GS), (2) unconstrained CCA (unc-CCA), (3) non-negative-cCCA ((p, ψ) = (1, 0)), (4) sum-cCCA ((p, ψ) = (1, 1)).

3.5.1 Simulated data—Forty-two cCCA models given by parameters: ψ = 1, 2, 4, 8, 16, 32 and p = 0.5, 1, 2, 4, 8, 16, 32 and the above 4 benchmark methods, were used in analyzing the simulated data. Note that sum-cCCA was covered by model (p, ψ) = (1, 1) resulting in 45 models in total.

Evaluation of each method was obtained both qualitatively and quantitatively. Activation maps from each method were obtained, thresholded and compared with the simulated patterns, forming the qualitative measurements of the model. True positive rate (TPR) (sensitivity), true negative rate (TNR, specificity), false positive rate (FPR) and false negative rate (FNR) were calculated at each threshold (from minimum to maximum of the activation map) and ROC curves were generated. The point at the lowest Total False Rate, i.e. (FPR + FNR), was used to threshold the activation map. Sensitivity, specificity, detection accuracy and F_1 score $\left(\frac{2\mathrm{TP}}{2\mathrm{TP}+\mathrm{FP}+\mathrm{FN}}\right)$ [Fawcett, 2006] for that specific threshold formed the quantitative measurements of model performance. The AUC was calculated within the range of FPR from 0 to 0.1, which is the range that is most important in neuroscience, to provide another quantitative measurement of the overall model performance. The (p, ψ) cCCA model which gives the largest AUC within the range of FPR $\in$ [0,0.1] was selected as the optimum-cCCA model and the corresponding p and ψ were defined as the optimum parameters.

The family-cCCA model was validated using the method described in section 2.5. A total of 42 models with different p and ψ were considered and 1521 (39×39) voxels from simulation and real fMRI data were used. For each (p, ψ) cCCA model, the initial threshold T used to classify real data as active set A and inactive set I was p<0.05 (FWE), acquired from wavelet-resample resting-state data; the process to create synthetic neighborhoods was repeated 500 times. The correlation between two measurements $\mu_{i=1,\ldots,42}$(simulation) and $\nu_{i=1,\ldots,42}$(real fMRI data) was obtained and the Bland–Altman plot was generated.

3.5.2 Real fMRI data—Four benchmark models and the family-cCCA model with optimum parameters p and ψ (optimum-cCCA), acquired and validated through simulation, were used in analyzing real fMRI episodic memory data. The non-smoothed fMRI data, preprocessed in section 3.2, was input for single voxel analysis, CCA and cCCA analysis with window size of 3×3 in a 2D slice. The smoothed data were input into SV+GS analysis. Activation maps for contrast Encoding v/s Control for different analyses were obtained and thresholded at both different significance levels (p<0.1, 0.01, 1e-3, 1e-4, 1e-5, (uncorrected) and 0.05 (FWE)). All thresholds for different methods were acquired non-parametrically as

described in section 2.4.1. Activation maps were co-registered to the co-planar high-resolution T2 image for display.

Apparent ROC curves for four benchmark cases and the optimum-cCCA model were generated as described in section 2.5.2. Original threshold used to acquire apparent ground truth is $p<0.05$ (FWE) and the process of exchanging neighbors was repeated 500 times for each method. Each analysis method was then applied to the exchanged neighborhoods and ROC curves were generated. Sensitivity, specificity, F1 score and accuracy for each method were also calculated at significance level of $p<0.05$ (FWE).

3.5.3 Classification and prediction between aMCI subjects and NCs—To further demonstrate the better performance of the optimum-cCCA model, a classification and prediction of activation patterns between aMCI subjects and NCs were carried out. Features generated from each analysis methods at different significance levels were input into classification separately. Prediction accuracy was then calculated and used to compare different model performance.

3.5.3.1 Feature set: Memory activation *F*-maps for contrast Encoding v/s Control, detected by single voxel analysis, four benchmark models and optimum-cCCA model, were obtained at different significance levels ($p < 0.1$, 1e-2, 1e-3, 1e-4, and 1e-5 (uncorrected)) separately. Activation maps were further co-registered to the subject T1 image. High-resolution T1 scan for each subject was input into Freesurfer [Fischl., 2012, Iglesias et al., 2015] to acquire subject-specific MTL subregion masks, including: CA1 (Cornu Ammonis area 1), CA234&DG (combining CA2, CA3, CA4 and Dentate Gyrus), SUB (Subiculum), ERC (Entorhinal Cortex), PHC (Parahippocampal Cortex) and FUS (Fusiform Gyrus). Activation percentages $\left(\frac{\text{Number of voxels in one subregion passing threshold}}{\text{Total number of voxels in one subregion}}\right)$ for each subject in the above MTL subregions at different significance levels were then extracted as feature set for classification and prediction. Note that the feature sets were generated independently from different analysis methods at various significance levels and prediction accuracy was calculated for each feature set. Statistical thresholds for all the significance levels were calculated from null data as described in section 2.4.1.

3.5.3.2 Classification method: Previous study has shown that Radial Basis Function Network (RBFN) classifier [Brommhead and Lowe, 1988, Haykin, 2008] is more powerful in prediction of aMCI using local activation patterns in MTL subregions when using limited data samples [Jin et al., 2015]. Therefore, RBFN with leave-2-out cross-validation method was applied to determine the predication accuracy. In each leave-2-out validation loop, one subject from each group was left out for testing purpose so that the training sample was balanced between the two groups. The cross-validation loop was repeated for every combination of one subject from each group being left out, i.e. $\binom{7}{1} \times \binom{7}{1} = 49$ times. To test the statistical significance of the obtained prediction accuracy, we used the permutation test to compute the p-value at the 0.05 level non-parametrically. The group indices (aMCI subjects or NCs) were randomly permuted.

The exact same analysis was ran on every possible random permutation to acquire the null-distribution of the prediction accuracy. Note that this process was also done for features generated from each analysis method separately.

4. RESULTS

4.1 Simulation

Fig. 4(A) (B) shows the empirical distribution of active neighbors with active (Fig. 4(A)) and inactive (Fig. 4(B)) center for real data (blue bars) and simulated pseudo-real data (orange bars). The active voxels are labeled as described in section 3.3.1. As shown in Fig.4 (A) and (B), the empirical distribution of active neighbors with an active or inactive center from real data and simulated data are approximately matched. The Pearson correlations between distributions are 0.9285 (p-value<0.001) for that with active center and 0.9738 (p-value<0.001) for that with inactive center. In Fig. 4(C), the average correlation value for significance level of p<0.05 (uncorrected) is shown by the solid blue line for real data and dashed red line for the pseudo-real data against different noise fractions f. The matched value for f in simulation is around 0.8, marked by the black arrow. Therefore, we use the simulation with noise fraction $f = 0.8$ (SNR = 0.25) to mimic real fMRI data and the simulation with noise fraction $f = 0.85$ to demonstrate the cCCA model performance for a higher noise case.

Figure 4. Simulation method.

The performance of all 45 models (described in section 3.4.1), measured by AUC with FPR $\in [0, 0.1]$, applied to simulated data is listed in Table 2. The optimum parameters for the family-cCCA model is achieved at $(p, \psi) = (1,8)$ for simulation with $f = 0.8$, and $(p, \psi) = (1,2)$ for simulation with $f = 0.85$ (highlighted in orange in Table (2). We also highlighted all (p, ψ) combinations that gives an AUC within the range of of the maximum value (highlighted in blue in Table (2)).

Table 2. Area under ROC for simulated data using different analysis methods.

The simulation results for the 4 benchmark models and the optimum cCCA model for $f = 0.8$ and $f = 0.85$ cases are shown in Figs. 5 and 6, respectively. As can be seen in Fig. 5(A), optimum-cCCA has the best performance in considering both sensitivity and specificity (Fig. 5(B) for F_1 score and accuracy and Fig. 5(C) for ROC curve). Even in the higher noise case, as can be seen in Fig. 6, optimum-cCCA still outperforms all the five benchmark models as outlined in Table 3.

Figure 5 and Figure 6. Simulation results for simulation with noise fraction $f = 0.8$ and $f = 0.85$, respectively.

Table 3. Improvement in detection accuracy of optimum-cCCA model relative to benchmark models.

Using our validation method as described in section 2.5, we obtain a Pearson correlation of 0.9874 (p-value<0.001) between sequence $\mu_{i=1,\ldots,42}$ from the simulated data and sequence

$\nu_{i=1,\ldots,42}$ from real fMRI data (Fig. 7(A)). This high positive correlation demonstrates the linear dependence between the performance of the proposed cCCA model in analyzing real fMRI data and simulated pseudo-real data. In addition, as shown in the Bland-Altman plot for mean normalized $\mu_{i=1,\ldots,42}$ and $\nu_{i=1,\ldots,42}$ in Fig. 7(B), almost all of the differences between $\mu_{i=1,\ldots,42}$ and $\nu_{i=1,\ldots,42}$ are within the 95% limits of agreement. Therefore, the optimum coefficient selected from the simulation ($(p, \psi) = (1,8)$) can also be considered as optimum parameters of the cCCA model for this fMRI data set.

Figure 7. Correlation plot and Bland-Altman plot for $\mu_{i=1,\ldots,42}$ and $\nu_{i=1,\ldots,42}$.

4.2 Smoothing artifact

Two representative simulated 7×7 neighborhoods with no active neighbor (s=0) and 8 active neighbors (s=8) are shown in Fig. 8(B) and Fig. 8(D) separately. The corresponding $\overline{\mathrm{AR}}$ map for different analysis methods with various noise levels are shown in Fig. 8(C) and Fig. 8(E), respectively. The final weighted average $\overline{\mathrm{AR}}$ map $\left(\overline{\overline{AR}}\right)$ is shown in Fig. 8(F). All activations are thresholded at p-value of 0.05 (FWE), using the method described in section 2.4.1. The horizontal axis of the $\overline{\mathrm{AR}}$ maps labels the noise fraction while vertical axis labels different analysis methods applied, ranking from the weakest to the strongest center dominant constraint (top to bottom). The yellow color ($\overline{\mathrm{AR}} = 9$) indicates that the smoothing artifact is significant whereas dark blue color indicates $\overline{\mathrm{AR}}<1$, i.e. detected activation is less than the real activation and therefore, there are false negatives. Light blue color indicates $\overline{\mathrm{AR}}=1$; that is, there is no smoothing artifact and all true activations are being detected. As shown in Fig. 8(C) (E) and (F), unc-CCA has the most significant smoothing artifact and as the constraint of the center voxel increases in the analysis methods, the smoothing artifact decreases. Constrained CCA models with strong center dominance constraint show no smoothing artifacts for all noise levels.

Figure 8. Smoothing artifact for different analysis methods as a function of noise fraction in simulated data.

4.3 Real fMRI data analysis

Fig. 9(A) shows activation maps for a representative subject, produced by single voxel analysis, four benchmark models and optimum-cCCA model with contrast Encoding v/s Control, thresholded at p<0.001 (uncorrected) and overlaid on the subject T2 image. *F*-statistic described in section 2.4 is used to display the activation maps. Table 4 lists the statistical thresholds, *F* statistics for contrast Encoding v/s Control) at p-value of 1e-5, 1e-4, 1e-3 and 1e-2 (uncorrected), obtained from null data as described in section 2.4.1.

Table 4. Statistical thresholds at p-value of 1e-5, 1e-4, 1e-3 and 1e-2, obtained from wavelet-resampled resting-state data, *F* statistics for contrast Encoding v/s Control.

Apparent ROC curves for each analysis method, calculated as described in section 2.5.2, are shown in Fig. 9(B). Sensitivity (blue bars), specificity (orange bars), accuracy (grey bars) and F1 score (yellow bars), calculated at p<0.05 (FWE) are displayed in Fig. 9(C). Optimum-cCCA outperforms other four analysis method with the largest area under the

apparent ROC curves (red curve in Fig. 9(C)) and the highest accuracy and F1 score (grey and yellow bars in Fig. 9(B)). Average activation percentages for MTL subregions (mean and standard deviation, averaging over 14 subjects) are listed in Table 5.

Figure 9. Analysis results on real fMRI data.

Table 5. Percentage of active voxels in MTL subregions for different analysis methods.

Classification and prediction accuracy between aMCI subjects and NCs are shown in Fig. 10, obtained with RBFN classifier and leave-2-out cross-validation methods. Horizontal axis labels the different significance level used to obtain features, i.e. activation percentages in the MTL subregions. For each significance level, feature sets generated from single voxel analysis (blue bars), SV+GS (orange bars), unc-CCA (grey bars), non-negative-cCCA (yellow bars) sum-cCCA (purple bars) and optimum-cCCA (green bars) are input into classification to obtain prediction accuracy. The dashed black curve is the 95^{th} percentile of the null distribution of prediction accuracy, acquired from permutation tests.

Figure 10. Classification and prediction accuracy between aMCI subjects and normal controls based on activation percentages obtained from different analysis methods.

5. DISCUSSION

This study introduces a novel family-cCCA model as a more flexible technique to detect fMRI activation. We solve the family-cCCA model with numerical optimization techniques that includes a BFGS algorithm with backtracking line search method (Fig.1, section 2.2.1), an augmented Lagrangian algorithm (section 2.2.2.1) and a reduced gradient method (section 2.2.2.2). The family-cCCA model is validated on both simulation studies (Fig. 5 and 6) and real episodic memory fMRI data for 7 aMCI subjects and 7 NCs (Fig. 9 and 10). Results demonstrate superior performance of the family-cCCA model.

5.1 Performance of family-cCCA model

It is difficult to use real fMRI data to compare performance of different models since the location of activation is unknown. Instead, we use simulated data with known ground truth to assess the performance of each analysis method.

5.1.1 Simulation method—We simulate pseudo-real data based on Cordes et al [2012b] with several improvements. First, we include 3×3 neighborhoods without any activation voxel in our simulation and evaluate the activation status for every voxel in the 39×39 grid. Therefore, the Euclidean distances between 3×3 neighborhoods with active voxels are varying. Further, as shown in Fig. 4(A) and Fig. 4(B), the empirical distributions of active neighbors given an active or inactive center in our simulation are highly correlated with those in real fMRI data. Second, both temporal and spatial correlation within and between time series are kept in our simulation. Time courses for each voxel are combinations of $\boldsymbol{y}_{\text{active}}$ and $\boldsymbol{y}_{\text{null}}$ from real fMRI data so that the temporal correlation is preserved. A Gaussian PSF estimated from wavelet-resampled resting-state time series is applied to the 39×39 simulation grid so that the spatial correlation in simulated data is the same as in real

fMRI data. Due to above improvements, the selected noise fraction in our simulation that mimics real fMRI data is also different from the one in Cordes et al [2012b].

The first simulation is done with an SNR close to real episodic memory data (SNR = 0.25) and is used to determine the optimum parameters of the family-cCCA model in analyzing real fMRI data. The second simulation is done with a lower SNR (higher noise fraction) and is used to compare the family-cCCA model to the various other models in a more noisy scenario since the SNR of some regions in fMRI could be lower than the average SNR [Glover and Law, 2001]. In assessing our family-cCCA model performance, we calculate the AUC over the interval of FPR ∈ [0, 0.1] for 42 different combinations of p and ψ (Table 2). The family-cCCA model which gives the largest area (i.e. AUC integrating over FPR ∈ [0, 0.1]) is selected as the optimum-cCCA model. The optimum parameters are achieved at (p, ψ) = (1,8) for simulation with SNR of 0.25 and (p, ψ) = (1,2) for simulation with a lower SNR, respectively. These parameters are fMRI data–set-dependent and are discussed later.

5.1.2 Model performance: activation maps thresholded at min (FPR + FNR)— To compare the optimum-cCCA model to the various existing models [Friston et al., 1994, Friman et al., 2003, Cordes et al., 2012b], we threshold the activation map from each analysis with the statistic value corresponding to the lowest combined false rate, min (FPR + FNR) , which represents the best performance that a specific model can achieve (Fig. 5(A) and Fig. 6(A)). With this threshold, CCA with no or weak constraints (unc-CCA and non-negative-cCCA) produce block-type smoothing artifacts, leading to an increase of false positives (2nd and 3rd orange bars in Fig. 5(B)). Conversely, methods with strong dominance constraints of the center voxel (for example sum-cCCA) give fewer artifacts and lead to a higher specificity (4th orange bars in Fig. 5(B)). Low specificity of standard CCA also occurs because CCA allows more variables (8 in our case) to model the signal, which may lead to overfitting. Therefore, when evaluating activation of the center voxel in a neighborhood, stronger spatial constraints are required to explain the signal with more variables while limiting the number of false activations. However, a constraint that is too strong may reduce the set of feasible solutions thereby decreasing the sensitivity. As represented by the 1st blue bar in Fig. 5(B), mass-univariate analysis produces a lower sensitivity, especially as compared to the weak cCCA models (2nd and 3rd blue bars in 5(B)). The proposed family-cCCA model is more flexible than the above weak or strong constrained CCA models since it covers a larger space of possible center dominance constraints and therefore improves sensitivity for a given specificity in fMRI activation detection (5th blue and orange bars in Fig. 5(B)).

In considering the trade-off between sensitivity and specificity, we calculate F_1 scores $\left(\frac{2\mathrm{TP}}{2\mathrm{TP}+\mathrm{FP}+\mathrm{FN}}\right)$ as an important quantitative measurement to further evaluate the optimum-cCCA model in comparison to the four benchmark models. In both simulations, optimum-cCCA outperforms all the other benchmark models and has highest F_1 score (5th yellow bar in Fig. 5(B) and Fig. 6(B)).

5.1.3 Model performance: ROC curves—To assess the overall model performance, rather than comparing statistical numbers at a single threshold, ROC curves for the five

methods are shown in Fig. 5(C) and Fig. 6(C) and the AUCs over the interval of FPR from 0 to 0.1 for all 45 models are calculated for each method and listed in Table 2.

As shown in Fig. 5(C), the ROC curve of SV+GS (grey line) is close to sum-cCCA (green line) in the simulation with a higher SNR while in the simulation with a lower SNR, sum-cCCA has a much larger AUC as compared to SV+GS (Fig. 6(C) and Table. 2(B)). This result is consistent with previous conclusions that high correlated noise decreases the effectiveness of mass-univariate analysis while multivariate analysis methods take advantages of local neighboring information and thus improves activation detection performance [Friman et al., 2003, Cordes et al., 2012b]. Using local neighborhood information is a key element in multivariate analysis methods to obtain better activation maps. Further, CCA with stronger constraints (e.g. max-cCCA, sum-cCCA, and optimum-cCCA) perform better than unc-CCA or non-negative-cCCA, especially in higher noise case (see higher AUCs in Table 2). As previous findings indicate, if the constraint of cCCA does not guarantee dominance of the center voxel, artifacts are observed [Cordes et al., 2012b]. The proposed family-cCCA model covers all the analysis methods with different parameters and determines the optimum parameters based on the noise level in real fMRI data. Therefore, the proposed family-cCCA model is more flexible in analysis and less sensitive to noise, as represented by the red curves in Fig. 5(C) and Fig. 6(C). For any fMRI data, optimum parameters of cCCA model can be roughly selected based on the noise level in fMRI. Given high SNR fMRI data, cCCA with a stronger dominance constraint of the center voxel should be applied whereas for low SNR fMRI data more local information should be incorporated i.e. a weaker constraint should be used in family-cCCA.

We also listed the AUC of mass-univariate analysis without Gaussian Smoothing (SV) in both simulations in Table 2. As shown in Table 3, optimum-cCCA still outperforms SV analysis with an increase in AUC of 1.42% and 13.30% in both simulations. In general, in noiseless data, SV is the best analysis method since it has the highest specificity. As noise increases, SV is not as good as multivariate methods since it assumes independent neighboring voxels and has a lower sensitivity. This fact is one reason why SV analysis for data with lower noise level has a higher AUC than four benchmark models in our simulation (see Table 2(A)). Another reason for the superior performance of SV in this case is the way simulated data is generated. The active time series in simulation is the combination of the most significant voxel obtained from SV analysis in real fMRI data ($\boldsymbol{y}_{\text{active}}$) and the wavelet resampled resting state data ($\boldsymbol{y}_{\text{null}}$) with a noise fraction f. With a lower noise level, the active simulated time series are dominated by $\boldsymbol{y}_{\text{active}}$ obtained from SV analysis. Therefore, SV outperforms the other methods which use local neighborhoods information without a strong enough constraint. The smoothing artifact here is significant for SV+GS and those multivariate cCCA methods without a strong spatial constraint. For this reason a more comparable benchmark is SV+GS instead of SV alone. As noise increases, multivariate analysis with less strong spatial constraint (e.g. sum-cCCA) starts outperforming the SV analysis (see Table 2(B)) since incorporating local information increases the sensitivity. The family-cCCA model, in both noise level cases, determines the optimum parameters for the constraint based on the data itself. Therefore, more flexible constraint is obtained from fMRI data itself and optimum-cCCA still outperforms SV analysis.

In summary, the optimum-CCA model, incorporating the local spatial activation information while keeping the appropriate dominance of the center voxel, more accurately detects activations in noisy data.

5.2 Smoothing artifact of different analysis models

As discussed in section 5.1.2, CCA with weak or no constraint will produce significant smoothing artifacts. The proposed family-cCCA model covers a larger space of the constraints and therefore we conducted another simulation to evaluate the smoothing artifact in the family-cCCA model.

We select 8 analysis models listed in section 3.3, ranked from the weakest to the strongest center dominance constraint. Other than unc-CCA, the remaining 7 models are covered by the family-cCCA model. We calculate $\overline{\mathrm{AR}}$, the ratio of detected activated voxels in the 5×5 blue striped area to the true active voxels in the center 3×3 orange area, to evaluate the smoothing artifact. As shown in Fig. 8(C) and (E), when noise fraction is less than 75%, we obtain $\overline{\mathrm{AR}}_{\mathrm{s}}$ much greater than 1 for unc-CCA method and family-cCCA model with weak constraints (yellow color in Fig. 8(C), light green color in Fig. (E)). This behavior occurs because the activation of the center voxels are bleeding into all 8 neighbors, thus produce a significant smoothing artifact and create false positives. Conversely, the proposed family-cCCA model with stronger center dominance constraints ((p, ψ) = (1,4); (p, ψ) = (1,6); (p, ψ) = (1,8) and also single voxel analysis) show no smoothing artifact ($\overline{\mathrm{AR}}_{=1}$, light blue color in Fig. 8(C) and (E)). Further, as the noise fraction f increases, the smoothing artifacts decreases and for $f > 75\%$, all the analysis methods are conservative since they all produce false negatives. Similar performances are also observed in other simulated cases with different number of active voxels as well as in the average $\overline{\mathrm{AR}}$ map (Fig. 8(F)).

5.3 Episodic memory activation in fMRI data analysis

Episodic memory is the ability to encode and retrieve personal experiences. Previous studies have shown that episodic memory is facilitated by neural pathways in the medial temporal lobe (MTL), which includes hippocampus, para-hippocampal areas and other nearby regions [Squire 1992; Zeineh et al., 2003; Squire et al., 2004; Dickerson and Eichenbaum, 2010]. In particular, using a face-name encoding and retrieval task, Zeineh et al., [2003] have shown that more activation can be detected in Cornu Ammonis (CA) areas and Dentate Gyrus (DG) during encoding while fusiform is active regardless of encoding and retrieval. Therefore, in our face-occupation encoding and retrieval task paradigm, activations in hippocampus, para-hippocampal areas and other nearby MTL regions are expected.

Activation detection in these regions is often complicated due to the low activation and low SNR caused by susceptibility artifacts from the sphenoid sinus [Cordes, et al., 2012b]. Mass-univariate analysis could possibly miss the activation due to the low sensitivity for data with low SNR (because of signal dropout in or near the hippocampus). CCA models on the other hand incorporate multiple time series with optimum weight in a local neighborhood and are thus more sensitive to detect activations. However, more activation identified by the multivariate methods is not always better as artifacts can be introduced and larger fraction of

false positives in null data can be observed (See Table. 4). Therefore, multivariate analysis methods need to be applied together with proper constraints.

5.3.1 Activation detection—As can be seen from Fig. 9(A), mass univariate analysis, with or without Gaussian smoothing, misses most of the activations in hippocampus or para-hippocampal areas at a significance level of $p<0.001$ (uncorrected). Unc-CCA and non-negative-cCCA detect large activations in the entire brain but also produce smoothing artifacts in cerebrospinal fluid (CSF) area as well as a higher statistic (FP) on null data. Our optimum-cCCA model detects bilateral activations in MTL (Fig. 9(A)) with a much lower threshold on null data and no artifacts. The superior performance of the optimum-cCCA model is further demonstrated by the apparent ROC curves and related quantitative measurements. Mass univariate analysis is specific but conservative at $p<0.05$ (FWE), reflected in a high specificity (1st orange bar in Fig 9(B)) but a low sensitivity (1st blue bar in Fig 9(B)). CCA with no or weak constraint has a larger sensitivity than univariate analysis (2nd and 3rd blue bars in Fig. 9(B)) but very low specificity (2nd and 3rd orange bars Fig 9 (B)). Optimum-cCCA model produces the most balanced sensitivity and specificity (5th blue and orange bars in Fig 9(B)), and thus shows the highest accuracy (5th grey bar in Fig 9(B)) and F1 score (5th yellow bar in Fig 9(B)). Optimum-cCCA model outperforms all other benchmark cases with the largest area under apparent ROC curves, as shown by the red curve in Fig. 9(C). These observations are also consistent with the simulation results discussed in section 5.1.2 and previous findings [Friman et al., 2003, Cordes et al., 2012].

5.3.2 Classification and prediction between aMCI subjects and NCs—Pathological changes for aMCI subjects occur early in the MTL [Dickerson et al., 2004]. Given our episodic memory paradigm, which is targeted at MTL activation, aMCI subjects and NCs will reveal different functional changes. A superior fMRI analysis model will accurately detect MTL activations for both aMCI subjects and NCs, and thus makes a classification and prediction based on MTL activation patterns possible. We carried out this classification and prediction based on the activation percentages of MTL subregions generated from different analysis methods to further demonstrate the better performance of the optimum-cCCA model.

The feature sets input into classification are the activation percentages in the MTL subregions, obtained from single voxel analysis, four benchmark methods and optimum-cCCA model at various significance levels ($p<0.1$, 1e-2, 1e-3, 1e-4 and 1e-5). As shown in Fig 10, the prediction accuracy is a function of significance level used to threshold the activation maps. Given a low significance (e.g. $p<0.1$, uncorrected), every voxel will be active and there is no difference between activation patterns of aMCI subjects and NCs (last blue, orange, grey, yellow purple and green bars in Fig. 10). Conversely, if the threshold is too high, no active voxel is detected for both aMCI subjects and NCs, thus results in failed classification. As shown by the 2nd and 3rd green bars in Fig 10, feature sets generated from the optimum-cCCA model at the significance level of $p<1e-3$ and 1e-4 produce the highest prediction accuracy. These prediction accuracies are greater than the 95th percentile of null distribution (dashed black line in Fig. 10), thus is statistically reliable. Therefore, optimum-cCCA is superior as its detected activations best discriminate aMCI subjects from NCs.

5.3.3 Signed statistics—In this study, we are more interested in activation differences between conditions Encoding and Control. Therefore, the unsigned F statistics are used. However, in some analysis, negative activation is also important, thus a signed statistic is useful. The Λ statistics in Eq. (17) and F statistics introduced in section 2.4 can also be converted to a signed Λ and F statistics ($\Lambda_{\pm}$ and $F_{\pm}$) following Calhoun et al. [2004] and Cordes et al. [2012b]:

$$\begin{aligned} \Lambda_{\pm} &= \operatorname{sign}\left(\boldsymbol{c}'\beta\alpha\right) \times \Lambda \\ F_{\pm} &= \operatorname{sign}\left(\boldsymbol{c}'\beta\alpha\right) \times F \end{aligned} \tag{22}$$

Of note, non-parametric significance test should be applied since the parametric distributions of these signed statistics are unknown.

5.4 Computational consideration

All calculations are performed in MATLAB (The Mathworks, Inc., version R2015b) on a Dell-workstation with Intel Xeon E5-2687W architecture running at a clock speed of 3.4GHz and equipped with 96GB of memory. As listed in Table 1 and discussed in section 2.3.3, the proposed cCCA model with reduced gradient algorithm takes around 1.5 hours to analyze one fMRI data set (approximately 100,000 voxels). If the optimum parameter p equals 1, the family-cCCA could be solved within 15 minutes as outlined in section 2.2.1.2 and Appendix A.4. For the data set with optimum parameter $p \neq 1$ but needing a faster evaluation, a possible alternative is to approximate the optimum-cCCA model with another cCCA model with p equals 1 and a different $\hat{\psi}$. As listed in Table 2(A), in simulation with SNR of 0.25, (p, ψ) = (2,16) gives AUC of 0.0854, which is close to 0.0870 given by(p, ψ) = (1,8). The differences in AUC is 1.87%. Therefore, the proposed cCCA model could be solved faster by using an approximate $(p, \psi) = \left(1, \hat{\psi}\right)$ cCCA model for a quicker evaluation.

5.5 Limitations and Future studies

Data specific optimum model parameters—As shown in Table 3, optimum-cCCA outperforms all other models. However, these optimum parameters for the family-cCCA model is fMRI data-specific, i.e. specific to the task and contrast. As discussed in section 5.1, the optimum parameters can be roughly selected based on noise level in the given data set. To obtain exact data-and-contrast-specific optimum parameters of the cCCA model, the entire process of generating and analyzing a simulation close to the contrast in question needs to be repeated.

3D local neighborhood approach—A 3×3×3 3D neighborhood consists of 27 voxels and contains more local information than a 2D neighborhood. A family-cCCA model with a 3D local neighborhood can further increase the sensitivity while maintaining specificity. In this study, we use a 2D local neighborhood due to the high in-plane non-isotropic resolution of collected fMRI data. For future fMRI data with isotropic resolution a family-cCCA model with 3D local neighborhood could be implemented to further improve model performance.

Modeling hemodynamic response function—In the family-cCCA model proposed in this study, we consider constraints only on the spatial weights of each voxel. The hemodynamic function (HRF) is assumed to be identical for the entire brain. In a future study, another aspect is to simultaneously estimate a voxel-specific HRF and obtain optimal weights for the family-cCCA model. In univariate analysis, Pedregosa et al [2013] tried to solve this problem by jointly estimating the HRF and the activation patterns. For multivariate cases, Dong et al [2015b] proposed a CCA-involved local multimodal serial analysis method, where they replaced the design matrix with EEG data with different lags to count for the various onset times of HRFs. In future extension of this research project, we would like to include estimation of HRFs from fMRI data to obtain a more accurate model of cCCA where different voxels or regions may have a different HRF profile for the response.

Group Analysis—To the best of our knowledge, there is no analytically derived group analysis method for CCA-related models which have spatial constraints. In future studies, we would like to test a possible solution to this problem based on Fisher's data fusion [Fisher, 1925] method. The values of the individual test statistics obtained non-parametrically, either p-value or F-statistics, could be combined from multiple subjects to obtain single group statistic at each voxel level.

6. CONCLUSION

In this study, we proposed a new family-constrained CCA model as a more flexible tool for fMRI data analysis. Using numerical optimization techniques, we solved the cCCA problem accurately and efficiently. Simulations as close to real fMRI data were constructed to evaluate the family-cCCA model performance. Results demonstrate that the family-cCCA model with optimum parameters could increase the sensitivity for a given specificity in activation detection, especially for data sets with low SNR, as compared to the mass-univariate analysis and conventional CCA-related methods. The selected optimum-cCCA parameters were further validated with the proposed novel validation method based on area under the ROC curves and demonstrated to be also optimal in analyzing real fMRI data. The optimum parameters for the family-cCCA model are specific to our data set and chosen contrast.

The family-cCCA model with optimum parameters outperforms mass-univariate analysis in detecting small localized activations in noisy fMRI data, such as hippocampal activations in episodic memory fMRI data. Optimum-cCCA also has more statistical power as compared to the weak constrained CCA methods, since it produces less false positives on null data. Furthermore, the family-cCCA model is more flexible and reliable than the strong dominance CCA methods since it covers a larger space of possible center dominance constraints.

ACKNOWLEDGEMENTS

This research project was supported by the NIH (grant number 1R01EB014284 and COBRE grant 1P20GM109025). We would like to acknowledge Martin Merener's contribution to the initial stages of this research project.

APPENDIX A

A.1 Technical aspects regarding canonical correlation ρ

A.1.1 The domain

Note that ρ is undefined for those $\boldsymbol{a}$ and $\boldsymbol{\beta}$ for which $\boldsymbol{Y}'\boldsymbol{a}$ or $\boldsymbol{X}'\boldsymbol{\beta}$ has zero-variance. However, if the centered coordinates of $\boldsymbol{Y}$ and $\boldsymbol{X}$ are linearly independent, then $\boldsymbol{Y}'\boldsymbol{a}$ or $\boldsymbol{X}'\boldsymbol{\beta}$ will have zero-variance only if $\boldsymbol{a}$ or $\boldsymbol{\beta}$ are zero (respectively). Hence, the domain of ρ is$(\mathbb{R}^m \setminus \{0\}) \times (\mathbb{R}^n \setminus \{0\})$.

A.1.2 Maximum of ρ

Aside from the constraints, the domain of ρ is open, so the existence of the maximum requires a justification. Since ρ is invariant by scalar multiplication of its arguments, for every non-zero $\boldsymbol{a} \in \mathbb{R}^m$ and $\boldsymbol{\beta} \in \mathbb{R}^n$, it holds $\rho(\alpha, \beta) = \rho\left(\frac{\alpha}{\| \alpha \|}, \frac{\beta}{\| \beta \|}\right)$, therefore maximizing ρ on the whole domain is equivalent to maximizing it restricted to $\left\{\left(\widetilde{\alpha}, \widetilde{\beta}\right) \in \mathbb{R}^{m+n} | \quad \| \widetilde{\alpha} \| = \| \widetilde{\beta} \| = 1\right\}$ which is a closed and bounded set, therefore the continuous function ρ attains its maximum. Moreover, since ρ is differentiable, its maximum can be found among the critical points, i.e., solutions of $\nabla_\rho = 0$. Further, the objective function in unconstrained CCA problem is concave, thus the critical points found can only be at global maximum. For constrained CCA problem, the constraints define closed subsets in which the existence of the maximum is justified.

A.2. Proof of the equivalence between cCCA and cLS (Eq. (8))

Note that cLS has the extra constraint $\boldsymbol{a}'\boldsymbol{S}_{\boldsymbol{YY}}\boldsymbol{a} = \mathbf{1}$, that are not in cCCA. Since ρ is invariant under scalar multiplication of its arguments, given any $\boldsymbol{a}$ in the domain of ρ one can simply divide its coordinates by $\sqrt{a' S_{YY} \alpha}$, which then satisfies the constraints.

It can be assumed that when cCCA includes the constraint $\boldsymbol{a}'\boldsymbol{S}_{\boldsymbol{YY}}\boldsymbol{a} = \mathbf{1}$ and condition $\boldsymbol{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}\boldsymbol{a}$, both problems have the same feasible solutions. If $(\boldsymbol{a}^*, \boldsymbol{\beta}^*)$ is a feasible solution, then it is a global maximum of cCCA if and only if it is a global minimum of cLS. In the following we are proving this statement.

Let $(\boldsymbol{a}, \boldsymbol{\beta})$ be another feasible solution. We want to show that:

$$\| \boldsymbol{Y}\alpha^* - \boldsymbol{X}\beta^* \|^2 \leq \| \boldsymbol{Y}\alpha - \boldsymbol{X}\beta \|^2 \iff \frac{\alpha^{*\prime}\boldsymbol{S}_{\boldsymbol{YX}}\beta^*}{\sqrt{\alpha^{*\prime}\boldsymbol{S}_{\boldsymbol{YY}}\alpha^*\beta^{*\prime}\boldsymbol{S}_{\boldsymbol{XX}}\beta^*}} \geq \frac{\alpha'\boldsymbol{S}_{\boldsymbol{YX}}\beta}{\sqrt{\alpha'\boldsymbol{S}_{\boldsymbol{YY}}\alpha\beta'\boldsymbol{S}_{\boldsymbol{XX}}\beta}}. \quad \text{(A. 1)}$$

Using the definition of the norm for an arbitrary vector $\boldsymbol{z} \in \mathbb{R}$, i.e. $\| \boldsymbol{z} \|^2 = \boldsymbol{z}'\boldsymbol{z}$, we obtain

$$\begin{aligned} \| \boldsymbol{Y}\alpha^* - \boldsymbol{X}\beta^* \|^2 &= (\boldsymbol{Y}\alpha^*)'(\boldsymbol{Y}\alpha^*) + (\boldsymbol{X}\beta^*)'(\boldsymbol{X}\beta^*) - (\boldsymbol{Y}\alpha^*)'(\boldsymbol{X}\beta^*) - (\boldsymbol{X}\beta^*)'(\boldsymbol{Y}\alpha^*) \\ &= \alpha^{*\prime}\boldsymbol{Y}'\boldsymbol{Y}\alpha^* + \beta^{*\prime}\boldsymbol{X}'\boldsymbol{X}\beta^* - 2\alpha^{*\prime}\boldsymbol{Y}'\boldsymbol{X}\beta^* \end{aligned}$$

(A. 2)

and similarly

$$\| \boldsymbol{Y}\alpha - \boldsymbol{X}\beta \|^2 = \alpha'\boldsymbol{Y}'\boldsymbol{Y}\alpha + \beta'\boldsymbol{X}'\boldsymbol{X}\beta - 2\alpha'\boldsymbol{Y}'\boldsymbol{X}\beta. \quad \text{(A. 3)}$$

Because of the constraints $\boldsymbol{a}'\boldsymbol{S}_{\boldsymbol{YY}}\boldsymbol{a} = 1$ and the facts $S_{YY} = \frac{1}{N-1}\boldsymbol{Y}'\boldsymbol{Y}$, we obtain $\boldsymbol{a}'\boldsymbol{Y}'\boldsymbol{Y}\boldsymbol{a} = N - 1$ for any $\boldsymbol{a}$. Substituting $\boldsymbol{\beta}$ with the least square solution $(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}\boldsymbol{a}$ (see section 2.1.2), $\boldsymbol{a}'\boldsymbol{Y}'\boldsymbol{Y}\boldsymbol{a}$ with $N - 1$, Eq. (A.1) is reduced to

$$\begin{aligned} (N-1) - \alpha^{*\prime}\boldsymbol{Y}'\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{Y}\alpha^* &\leq (N-1) - \alpha'\boldsymbol{Y}'\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{Y}\alpha \\ \Longleftrightarrow \frac{\alpha^{*\prime}\boldsymbol{S}_{\boldsymbol{YX}}\beta^*}{\sqrt{\beta^{*\prime}\boldsymbol{S}_{\boldsymbol{XX}}\beta^*}} &\geq \frac{\alpha'\boldsymbol{S}_{\boldsymbol{YX}}\beta}{\sqrt{\beta'\boldsymbol{S}_{\boldsymbol{XX}}\beta}}. \end{aligned}$$

(A. 4)

Further simplifying $\frac{\alpha'\boldsymbol{S}_{\boldsymbol{YX}}\beta}{\sqrt{\beta'\boldsymbol{S}_{\boldsymbol{XX}}\beta}}$, we obtain

$$\begin{aligned} \frac{\alpha'\boldsymbol{S}_{\boldsymbol{YX}}\beta}{\sqrt{\beta'\boldsymbol{S}_{\boldsymbol{XX}}\beta}} &= \frac{\alpha'\frac{1}{N-1}\boldsymbol{Y}'\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{Y}\alpha}{\sqrt{\frac{1}{N-1}\left(\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}\alpha\right)'\left(\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}\alpha\right)}} = \frac{\frac{1}{N-1}\alpha'\boldsymbol{Y}'\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{Y}\alpha}{\sqrt{\frac{1}{N-1}\alpha'\boldsymbol{Y}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}\alpha}} \\ &= \sqrt{\frac{1}{N-1}\alpha'\boldsymbol{Y}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}\alpha}. \end{aligned}$$

(A. 5)

Therefore, relationship in Eq. (A. 4) is reduced to

$$(N-1) - \alpha^{*'}\boldsymbol{Y}'\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{Y}\alpha^{*} \leq (N-1) - \alpha'\boldsymbol{Y}'\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{Y}\alpha$$
$$\iff \frac{1}{N-1}\alpha^{*'}\boldsymbol{Y}'\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{Y}\alpha^{*} \geq \frac{1}{N-1}\alpha'\boldsymbol{Y}'\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{Y}\alpha,$$

(A. 6)

which indeed holds true.

A.3 Critical points of CCA

Without any constraint CCA is defined as:

$$\rho(\alpha,\beta) = \frac{\alpha'\boldsymbol{S}_{\boldsymbol{XY}}\beta}{\sqrt{\alpha'\boldsymbol{S}_{\boldsymbol{YY}}\alpha\beta'\boldsymbol{S}_{\boldsymbol{XX}}\beta}} \quad \text{(A. 7)}$$

where $\boldsymbol{S_{YX}}$, $\boldsymbol{S_{YY}}$ and $\boldsymbol{S_{XX}}$ are the sample average covariance and variance matrices. Taking the derivative of Eq. (A. 7) gives:

$$\nabla_{\alpha}\rho = \left(\frac{\partial\rho}{\partial\alpha_1},\ldots,\frac{\partial\rho}{\partial\alpha_m}\right)' = \frac{\alpha'\boldsymbol{S}_{\boldsymbol{YY}}\alpha\boldsymbol{S}_{\boldsymbol{YX}}\beta - \alpha'\boldsymbol{S}_{\boldsymbol{YX}}\beta\boldsymbol{S}_{\boldsymbol{YY}}\alpha}{(\alpha'\boldsymbol{S}_{\boldsymbol{YY}}\alpha)^{\frac{3}{2}}(\beta'\boldsymbol{S}_{\boldsymbol{XX}}\beta)^{\frac{1}{2}}} \quad \text{(A. 8)}$$

$$\nabla_{\beta}\rho = \left(\frac{\partial\rho}{\partial\beta_1},\ldots,\frac{\partial\rho}{\partial\beta_n}\right)' = \frac{\beta'\boldsymbol{S}_{\boldsymbol{XX}}\beta\boldsymbol{S}_{\boldsymbol{XY}}\alpha - \beta'\boldsymbol{S}_{\boldsymbol{XY}}\alpha\boldsymbol{S}_{\boldsymbol{XX}}\beta}{(\beta'\boldsymbol{S}_{\boldsymbol{XX}}\beta)^{\frac{3}{2}}(\alpha'\boldsymbol{S}_{\boldsymbol{YY}}\alpha)^{\frac{1}{2}}}. \quad \text{(A. 9)}$$

Setting the derivatives Eq. (A.8) and Eq. (A.9) equal to 0 yields

$$\alpha'\boldsymbol{S}_{\boldsymbol{YY}}\alpha\boldsymbol{S}_{\boldsymbol{YX}}\beta = \alpha'\boldsymbol{S}_{\boldsymbol{YX}}\beta\boldsymbol{S}_{\boldsymbol{YY}}\alpha \quad \text{(A. 10)}$$

$$\beta'\boldsymbol{S}_{\boldsymbol{XX}}\beta\boldsymbol{S}_{\boldsymbol{XY}}\alpha = \beta'\boldsymbol{S}_{\boldsymbol{XY}}\alpha\boldsymbol{S}_{\boldsymbol{XX}}\beta. \quad \text{(A. 11)}$$

Since $\boldsymbol{S_{XX}}$ and $\boldsymbol{S_{YY}}$ are positive-definite with probability 1, they are invertible. From Eq. (A. 11), we solve for $\boldsymbol{\beta}$ as

$$\beta = \frac{\beta'\boldsymbol{S}_{\boldsymbol{XX}}\beta}{\beta'\boldsymbol{S}_{\boldsymbol{XY}}\alpha}\boldsymbol{S}_{\boldsymbol{XX}}^{-1}\boldsymbol{S}_{\boldsymbol{XY}}\alpha. \quad \text{(A. 12)}$$

Substituting $\boldsymbol{\beta}$ from Eq. (A. 12) in Eq. (A. 10) yields

$$\frac{\alpha' \boldsymbol{S_{YY}} \alpha}{\alpha' \boldsymbol{S_{YX}} \beta} \boldsymbol{S_{YY}^{-1}} \boldsymbol{S_{YX}} \frac{\beta' \boldsymbol{S_{XX}} \beta}{\beta' \boldsymbol{S_{XY}} \alpha} \boldsymbol{S_{XX}^{-1}} \boldsymbol{S_{XY}} \alpha = \alpha. \quad \text{(A. 13)}$$

Note ρ is defined as Eq. (A. 7), which also equals to $\frac{\beta' \boldsymbol{S_{XY}} \alpha}{\sqrt{\alpha' \boldsymbol{S_{YY}} \alpha \beta' \boldsymbol{S_{XX}} \beta}}$. Therefore, we further simplify Eq. (A. 13) to

$$\boldsymbol{S_{YY}^{-1}} \boldsymbol{S_{YX}} \boldsymbol{S_{XX}^{-1}} \boldsymbol{S_{XY}} \alpha = \rho^2 \alpha. \quad \text{(A. 14)}$$

Hence, the critical points $\boldsymbol{\alpha}$ are found among the eigenvectors of matrix $\boldsymbol{S_{YY}^{-1}} \boldsymbol{S_{YX}} \boldsymbol{S_{XX}^{-1}} \boldsymbol{S_{XY}}$ and ρ^2 is the corresponding Eigen Value.

A.4 Solution when $p = 1$

The cCCA $p = 1$ problem for arbitrary ψ becomes:

$$\text{cCCA:} \quad \max \frac{\alpha' \boldsymbol{S_{YX}} \beta}{\sqrt{\alpha' \boldsymbol{S_{YY}} \alpha \beta' \boldsymbol{S_{XX}} \beta}} \quad \text{w. r. t.} \quad \begin{cases} \alpha_1 \geq 0, \ldots, \alpha_m \geq 0 \\ \alpha_1 \geq \psi \sum_{k=2}^{m} \alpha_k \end{cases}. \quad \text{(A. 15)}$$

It is possible to linearly transform the cCCA constraint to a simple non-negative constraint by defining

$$\alpha = \boldsymbol{M\Phi}; \quad \Phi \geq 0. \quad \text{(A. 16)}$$

where $\boldsymbol{\Phi}$ is the new spatial weights vector and the matrix $\boldsymbol{M}$ for this transform is:

$$\boldsymbol{M} = \begin{pmatrix} \psi & \psi & \cdots & \psi & \psi \\ 0 & 1 & & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}. \quad \text{(A. 17)}$$

If $\boldsymbol{\Phi}$ satisfies the non-negative constraint then $\boldsymbol{\alpha}$ satisfies the $p = 1$ constraint. Notice that the weight of the center voxel α_1 equals to ψ times the summation of all the elements in $\boldsymbol{\Phi}$. Therefore, the first element of $\boldsymbol{\Phi}$ can also be 0, and $\boldsymbol{\alpha} = \boldsymbol{M\Phi}$ still satisfies the $p = 1$ constraint. This situation needs to be considered separately.

A.5 Proof of covariance matrix property

Let ***V,W*** be two random vectors and *T,R* are two transformations. Then, the covariance matrix of ***TU*** and ***RV*** is:

$$S_{TU\ \ RV} = E\left[(T\boldsymbol{U} - E[T\boldsymbol{U}])(R\boldsymbol{V} - E[R\boldsymbol{V}])'\right] \quad \text{(A. 18)}$$

where *E*[] denotes the expectation value of the argument. Eq. (A. 18) can be expanded as:

$$\begin{aligned} S_{TU\ \ RV} &= E\left[(T\boldsymbol{U} - TE[\boldsymbol{U}])(R\boldsymbol{V} - RE[\boldsymbol{V}])'\right] \\ &= TE\left[(\boldsymbol{U} - E[\boldsymbol{U}])(\boldsymbol{V} - E[V])'\right]R \\ &= TS_{UV}R. \end{aligned} \quad \text{(A. 19)}$$

A.6. Formulating objective function of the augmented Lagrangian method

In general, the cCCA problem (Eq. (8)) is to minimize objective function g(***α***) with equality constraint(s) $c_i(\boldsymbol{\alpha}) = 0, \forall i \in E$ and inequality constraint(s) $c_j(\boldsymbol{\alpha}) > 0, \forall j \in I$. In forming the objective function of the augmented Lagrangian method, inequality constraints are converted to equality constraints by adding $s_j \geq 0$ so that $c_j(\boldsymbol{\alpha}) - s_j = 0, \forall j \in I$. Each equality constraint is then added to the original objective function g(***α***) as a quadratic penalty term:

$$F(\alpha;\mu) = g(\alpha) + \frac{\mu}{2}\Sigma_{i\in E}c_i(\alpha)^2 + \frac{\mu}{2}\Sigma_{j\in I}(c_j(\alpha) - s_j)^2, \quad \mu > 0. \quad \text{(A. 20)}$$

Further, Lagrangian multiplier estimates are incorporated into the objective function to reduce the possibility of ill conditioning in the quadratic penalty method. Therefore, the objective function in the unconstrained augmented Lagrangian problem is as follows:

$$L_A(\alpha, \boldsymbol{s}, \lambda;\mu) = F(\alpha;\mu) - \Sigma_{i\in E}\lambda_i c_i(\alpha) - \Sigma_{j\in I}\lambda_j(c_j(\alpha) - s_j), \forall j \in I, \lambda_j \geq 0, \quad \text{(A. 21)}$$

where $L_A(\boldsymbol{\alpha}, \boldsymbol{s}, \boldsymbol{\lambda}; \mu)$ is the standard Lagrangian expression of $F(\boldsymbol{\alpha}; \mu)$ with constraints $c_i(\boldsymbol{\alpha}) = 0, \forall i \in E$ and $c_j(\boldsymbol{\alpha}) - s_j = 0, \forall j \in I$. If we consider ***s*** as the only unknown in $L_A(\boldsymbol{\alpha}, \boldsymbol{s}, \boldsymbol{\lambda}; \mu)$, Eq. (A. 21) is a convex quadratic function which reaches the global minimum at $s_j = \max\left(c_j(\alpha) - \frac{\lambda_j}{\mu}, 0\right), \forall j \in I$. Therefore, $L_A(\boldsymbol{\alpha}, \boldsymbol{s}, \boldsymbol{\lambda}; \mu)$ can be rewritten as $L_A(\boldsymbol{\alpha}, \boldsymbol{\lambda}; \mu)$ in term of our cCCA problem (Eq. (8)) as:

$$\begin{cases} L_A(\alpha;\lambda^t,\mu^t) = g(\alpha) + \sum_{j=1}^{10}\Gamma\left(c_j(\alpha),\lambda_j^t;\mu^t\right) - \lambda_{11}^t c_i(\alpha) + \frac{\mu^t}{2}(c_i(\alpha))^2 \\ \Gamma\left(c_j(\alpha),\lambda_j^t;\mu^t\right) = \begin{cases} -\lambda_j^t c_j(\alpha) + \frac{\mu^t}{2}(c_j(\alpha))^2 & \text{if } c_j(\alpha) - \frac{\lambda_j^t}{\mu^t} \le 0 \\ -\frac{1}{2\mu^t}\left(\lambda_j^t\right)^2 & \text{otherwise} \end{cases} \end{cases},$$

(A. 22)

where the superscript t denotes the t^{th} iteration and the constraints are:

$$\begin{cases} c_j(\alpha) = \alpha' \boldsymbol{S_{YY}} \alpha - 1 = 0, i \in E \\ c_j(\alpha) = a_j \ge 0, \quad j = 1,2,\ldots,9 \\ c_{10}(\alpha) = \alpha_1^p - \psi \sum_{k=2}^{9} \alpha_k^p \ge 0 \end{cases}.$$

(A. 23)

Eq. (A. 22) is then iteratively solved using unconstrained optimization techniques as described in Nocedal and Wright [2006], λ is updated by $\lambda_i^{t+1} = \lambda_i^t - \mu c_i(x), \forall i \in E \cup I$. If $\lambda_j^{t+1} < 0$ for any inequality constraint, λ_j^{t+1} is set to zero automatically.

A.7 Effective Gaussian smoothing filter width

To compare results of cCCA with single voxel and Gaussian smoothing (SV+GS), it is necessary to compute the FWHM of a comparable Gaussian smoothing kernel. CCCA is applied to local 3×3 neighborhoods involving a center voxel and its 8 neighbors. We solve cCCA with 256 sub-problems, where all the possible configurations are treated with equal probability. Therefore, we obtain the average configuration size s:

$$s = \frac{\sum_{k=1}^{9}\binom{8}{k-1}k}{\sum_{k=1}^{9}\binom{8}{k-1}} = 2.24 \text{ pixel}$$

(A. 24)

and the comparable Gaussian filter width is then

$$\text{FWHM} = s * \text{pixel size} = 3.8 \text{ mm}.$$

(A. 25)

This value is used in the mass-univariate analysis with Gaussian smoothing (SV+GS).

References

Abadie J, Carpentier J. Generalization of the Wolfe reduced gradient method to the case of nonlinear constraints. Optimization. 1969; 37:47.

Akaho, S. In Proceedings of the International Meeting of the Psychometric Society (IMPS2001). Springer-Verlag; 2006. A Kernel Method for Canonical Correlation Analysis.

Bland JM, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. The lancet. 1986; 327(8476):307–310.

Breakspear M, Brammer M, Bullmore E, Das P, Williams L. Spatiotemporal wavelet re-sampling for functional neuroimaging data. Hum Brain Mapp. 2004; 23:1–25. [PubMed: 15281138]

Broomhead DS, Lowe D. Radial basis functions, multi-variable functional interpolation and adaptive networks (No. RSRE-MEMO-4148). Royal Signals and Radar Establishment Malvern (United Kingdom). 1988

Bullmore E, Long C, Suckling J, Fadili J, Calvert G, Zelaya F, Carpenter T, Brammer M. Colored noise and computational inference in neurophysiological (fMRI) time series analysis: Resampling methods in time and wavelet domains. Hum Brain Mapp. 2001; 12:61–78. [PubMed: 11169871]

Calhoun VD, Stevens MC, Pearlson GD, Kiehl KA. fMRI analysis with the general linear model: removal of latency-induced amplitude bias by incorporation of hemodynamic derivative terms. NeuroImage. 2004; 22(1):252–257. [PubMed: 15110015]

Calhoun VD, Liu J, Adali T. A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data. NeuroImage. 2009; 45(1):S163–S172. [PubMed: 19059344]

Calhoun VD, Sui J. Multimodal fusion of brain imaging data: A key to finding the missing link (s) in complex mental illness. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging. 2016; 1(3):230–244. [PubMed: 27347565]

Cordes D, Jin M, Curran T, Nandy R. The smoothing artifact of spatially constrained canonical correlation analysis in functional MRI. Journal of Biomedical Imaging. 2012a; 2012:3.

Cordes D, Jin M, Curran T, Nandy R. Optimizing the performance of local canonical correlation analysis in fMRI using spatial constraints. Hum Brain Mapp. 2012b; 33:2611–2626. [PubMed: 23074078]

Das S, Sen PK. Restricted canonical correlations. Linear Algebra and its Applications. 1994; 210:29–47.

Dickerson BC, Salat DH, Bates JF, Atiya M, Killiany RJ, Greve DN, Dale AM, Stern CE, Blacker D, Albert MS, Sperling RA. Medial temporal lobe function and structure in mild cognitive impairment. Annals of Neurology. 2004; 56(1):27–35. [PubMed: 15236399]

Dickerson BC, Eichenbaum H. The episodic memory system: Neurocircuitry and disorders. Neuropsychopharmacology. 2010; 35(1):86–104. [PubMed: 19776728]

Dong L, Zhang Y, Zhang R, Zhang X, Gong D, Valdes-Sosa PA, Yao D. Characterizing nonlinear relationships in functional imaging data using eigenspace maximal information canonical correlation analysis (emiCCA). NeuroImage. 2015a; 109:388–401. [PubMed: 25592998]

Dong L, Wang P, Bin Y, Deng J, Li Y, Chen L, Luo C, Yao D. Local Multimodal Serial Analysis for Fusing EEG-fMRI: A New Method to Study Familial Cortical Myoclonic Tremor and Epilepsy. IEEE Transactions on Autonomous Mental Development. 2015b; 7(4):311–319.

Fawcett T. An introduction to ROC analysis. Pattern Recognition Letters. 2006; 27(8):861–874.

Fletcher R, Reeves CM. Function minimization by conjugate gradients. The Computer Journal. 1964; 7(2):149–154.

Fisher, RA. Statistical Methods for Research Workers. Oliver and Boyd; Edinburgh: 1925.

Friman O, Cedefamn J, Lundburg P, Borga M, Knutsson H. Detection of neural activity in functional MRI using canonical correlation analysis. Magn Reson Med. 2001; 45:323–330. [PubMed: 11180440]

Friman O, Borga M, Lundberg P, Knutsson H. Adaptive analysis of fMRI data. NeuroImage. 2003; 19:837–845. [PubMed: 12880812]

Fischl B. FreeSurfer. NeuroImage. 2012; 62(2):774–781. [PubMed: 22248573]

Friston KJ, Holmes AP, Worsley KJ, Poline JP, Frith CD, Frackowiak RS. Statistical parametric maps in functional imaging: a general linear approach. Hum Brain Mapp. 1994; 2(4):189–210.

Giavarina D. Understanding Bland Altman analysis. Biochemia Medica. 2015; 25(2):141–151. [PubMed: 26110027]

Glover GH, Law CS. Spiral-in/out BOLD fMRI for increased SNR and reduced susceptibility artifacts. Magn Reson Med. 2001; 46(3):515–522. [PubMed: 11550244]

Hardoon DR, Mourao-Miranda J, Brammer M, Shawe-Taylor J. Unsupervised analysis of fMRI data using kernel canonical correlation. NeuroImage. 2007; 37:1250–1259. [PubMed: 17686634]

Haykin, SS. Neural Networks and Learning Machines. Vol. 3. Pearson Education, Inc.; Upper Saddle River, NJ, USA: 2009.

Hestenes MR. Multiplier and gradient methods. Journal of Optimization Theory and Applications. 1969; 4(5):303–320.

Holmes AP, osephs O, Büchel C, Friston KJ. Statistical modelling of low-frequency confounds in fMRI. Proc 3rd Int.Conf. Func. Mapp. Hum. Brain. 1997

Hotelling H. Relations between two sets of variates. Biometrika. 1936; 28:321–377.

Iglesias JE, Augustinack JC, Nguyen K, Player CM, Player A, Wright M, Roy N, Frosch MP, Mckee AC, Wald LL, Fischl B, Van Leemput K. A computational atlas of the hippocampal formation using ex vivo, ultra-high resolution MRI: application to adaptive segmentation of in vivo MRI. NeuroImage. 2015; 115:117–137. [PubMed: 25936807]

Jin M, Nandy R, Curran T, Cordes D. Extending local canonical correlation analysis to handle general linear contrasts for fMRI data. Journal of Biomedical Imaging. 2012a; 2012:4.

Jin M, Curran T, Nandy R, Cordes D. Fast constrained canonical correlation analysis for fMRI. Visualization, Image Processing and Computation in Biomedicine,. 2012b DOI: 10.1615 /. 2012004964.

Jin, M., Curran, T., Nandy, R., Cordes, D. Organization of Human Brain Mapping (OHBM). United States: Jun. 2015 Prediction of aMCI subjects using local fMRI activation patterns.

Lasdon LS, Waren AD, Jain A, Ratner M. Design and testing of a generalized reduced gradient code for nonlinear programming. ACM Transactions on Mathematical Software (TOMS). 1978; 4(1): 34–50.

Levin-Schwartz Y, Song Y, Schreier PJ, Calhoun VD, Adalı T. Sample-poor estimation of order and common signal subspace with application to fusion of medical imaging data. NeuroImage. 2016; 134:486–493. [PubMed: 27039696]

Luenberger, DG., Ye, Y. Linear and nonlinear programming. Vol. 2. Addison-Wesley; Reading, MA: 1984.

Mardia KV, Kent JT, Bibby JM. Multivariate analysis. Academic press. 1979

Monteiro, JM., Rao, A., Ashburner, J., Shawe-Taylor, J., Mourão-Miranda, J. Multivariate effect ranking via adaptive sparse PLS. In Pattern Recognition in Neuroimaging (PRNI); 2015 International Workshop, IEEE; Jun 25-28. 2015

Nandy R, Cordes D. Novel nonparametric approach to canonical correlation analysis with applications to low CNR functional MRI data. Magn Reson Med. 2003; 50:354–365. [PubMed: 12876712]

Nandy R, Cordes D. Improving the spatial specificity of canonical correlation analysis in fMRI. Magn Reson Med. 2004; 52:947–952. [PubMed: 15389937]

Nandy R, Cordes D. A semi-parametric approach to estimate the family-wise error rate in fMRI using resting-state data. NeuroImage. 2007; 34:1562–1576. [PubMed: 17196400]

Nocedal, J., Wright, S. Numerical optimization. Springer Science & Business Media; 2006.

Nocedal, J., Yuan, YX. Combining trust region and line search techniques. Springer US; 1998. p. 153-175.

Pedregosa, F., Eickenberg, M., Thirion, B., Gramfort, A. PRNI '13 Proceedings of the 2013 International Workshop on Pattern Recognition in Neuroimaging; 2013. p. 165-169.

Petersen RC, Doody R, Kurz A, Mohs RC, Morris JC, Rabins PV, Ritchie K, Rossor M, Thal L, Winblad B. Current concepts in mild cognitive impairment. Archives of neurology. 2001; 58(12): 1985–1992. [PubMed: 11735772]

Ragnehed M, Engström M, Knutsson H, Söderfeldt B, Lundberg P. Restricted canonical correlation analysis in functional MRI—validation and a novel thresholding technique. Journal of Magnetic Resonance Imaging,. 2009; 29(1):146–154. [PubMed: 19097096]

Raichle ME, Snyder AZ. A default mode of brain function: a brief history of an evolving idea. NeuroImage. 2007; 37(4):1083–1090. [PubMed: 17719799]

Rockafellar RT. Monotone operators and the proximal point algorithm. SIAM Journal on Control and Optimization. 1976; 14(5):877–898.

Rosenbrock H. An automatic method for finding the greatest or least value of a function. The Computer Journal. 1960; 3(3):175–184.

Shanno DF. On Broyden-Fletcher-Goldfarb-Shanno method. Journal of Optimization Theory and Applications. 1985; 46:87–94.

Siotani, M., Hayakawa, T., Fujikoshi, Y. Modern Multivariate Statistical Analysis. American Sciences Press; Columbus, Ohio: 1985.

Soltanian-Zadeh H, Saigal R, Windham JP, Yagle AE, Hearshen DO. Optimization of MRI protocols and pulse sequence parameters for eigenimage filtering. IEEE Transactions on Medical Imaging. 1994; 13(1):161–175. [PubMed: 18218494]

Squire LR. Memory and the hippocampus: a synthesis from findings with rats, monkeys, and humans. Psychol. Rev. 1992; 99:195–231. [PubMed: 1594723]

Squire LR, Stark CE, Clark RE. The medial temporal lobe. Annu. Rev. Neurosci. 2004; 27:279–306. [PubMed: 15217334]

Sui J, Adali T, Pearlson G, Yang H, Sponheim SR, White T, Calhoun VD. A CCA+ ICA based model for multi-task brain imaging data fusion and its application to schizophrenia. NeuroImage. 2010; 51(1):123–134. [PubMed: 20114081]

Sui, J., Castro, E., He, H., Bridwell, D., Du, Y., Pearlson, GD., Jiang, T., Calhoun, VD. Combination of FMRI-SMRI-EEG data improves discrimination of schizophrenia patients by ensemble feature selection; In 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; IEEE; 2014. p. 3889-3892.

Thompson, B. Encyclopedia of Statistics in Behavioral Science. John Wiley & Sons, Ltd.; 2005. Canonical correlation analysis.

Udovičić M, Baždarić K, Bilić-Zulle L, Petrovečki M. What we need to know when calculating the coefficient of correlation? Biochemia Medica. 2007; 17(1):10–15.

Zeineh MM, Engel SA, Thompson PM, Bookheimer SY. Dynamics of the hippocampus during encoding and retrieval of face-name pairs. Science. 2003; 299:577–580. [PubMed: 12543980]

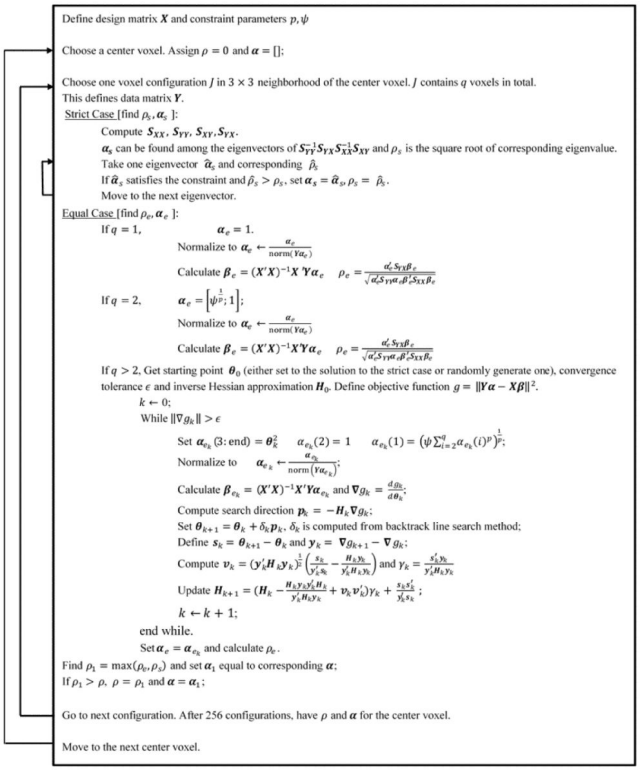

Figure 1.

Flow chart of the proposed algorithm in solving the cCCA model considering 256 configurations. All derivatives ∇g_k are calculated numerically. Note that self-scaling is implemented by parameter γ_k.

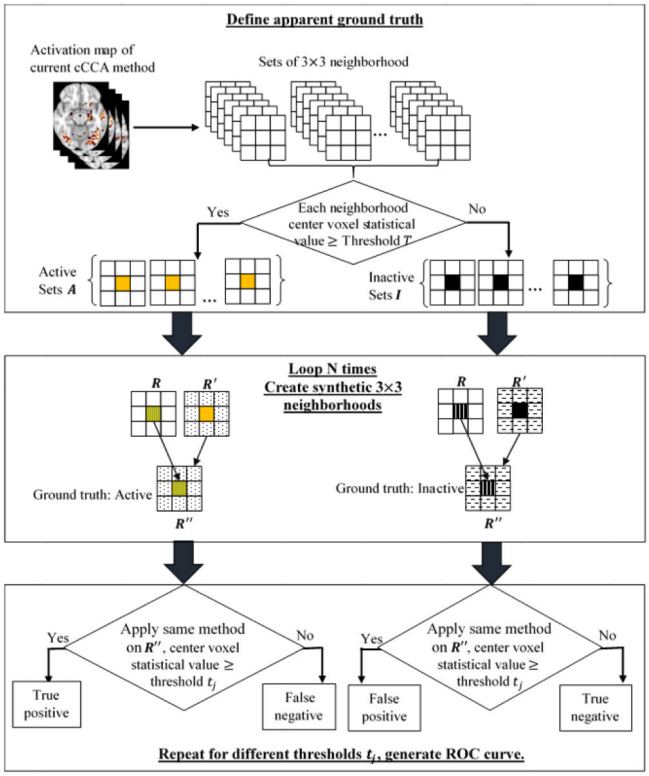

Figure 2.
Steps to generate apparent ROC curves from real fMRI data.

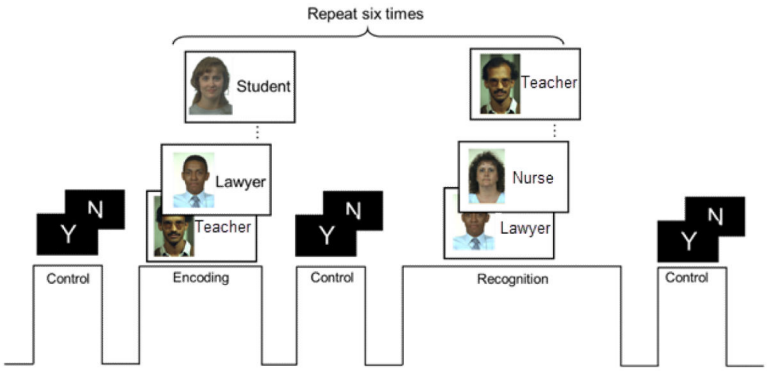

Figure 3.

FMRI episodic memory experiment. The task is to memorize novel faces paired with occupations. During the encoding period, 7 faces paired with occupations, each lasting 3 seconds are displayed, followed by an 11 seconds control period (for distraction purpose). During the recognition period, 14 stimuli, half novel and half identical, are shown in random order, each lasting for 3 seconds. Encoding and recognition periods are repeated 6 times in one session. The paradigm starts and ends with a control period.

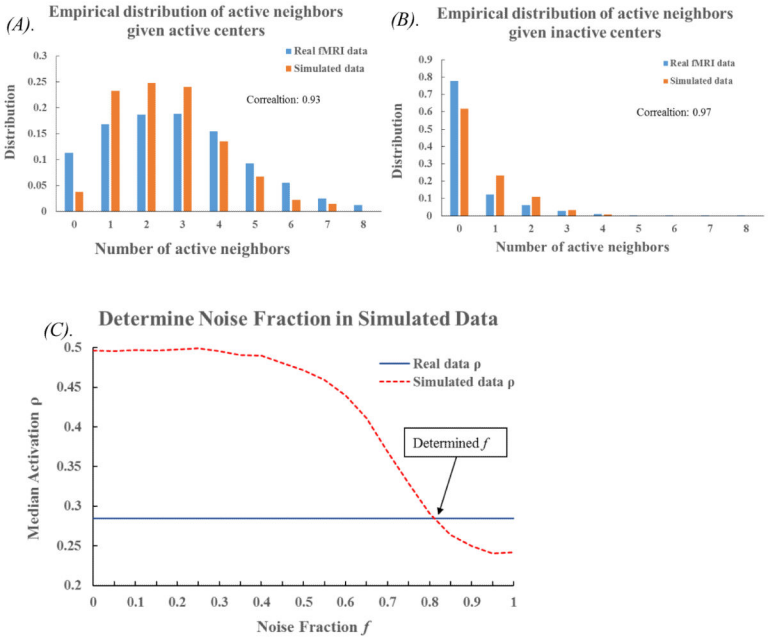

Figure 4.

(A) Empirical distribution of active neighbors with an active center in real fMRI data (blue) and simulated patterns (orange) at the significance level of p<0.05 (uncorrected) after applying single voxel analysis. The Pearson correlation between the two distributions is 0.9285 (p-value<0.001). *(B) E*mpirical distribution of active neighbors with an inactive center in real fMRI data (blue) or simulated patterns (orange). The Pearson correlation between the two distributions is 0.9738 (p-value<0.001). *(C)* Determination of noise fraction *f* in simulated data. The solid blue line represents the average correlation value (ρ) at significance level of p<0.05 (uncorrected) after applying single voxel analysis on real fMRI data; the dashed curve is the average correlation (ρ) value acquired in the same way on simulated data with different noise fraction *f* from 0 to 1. Black arrow indicates where the two median values match.

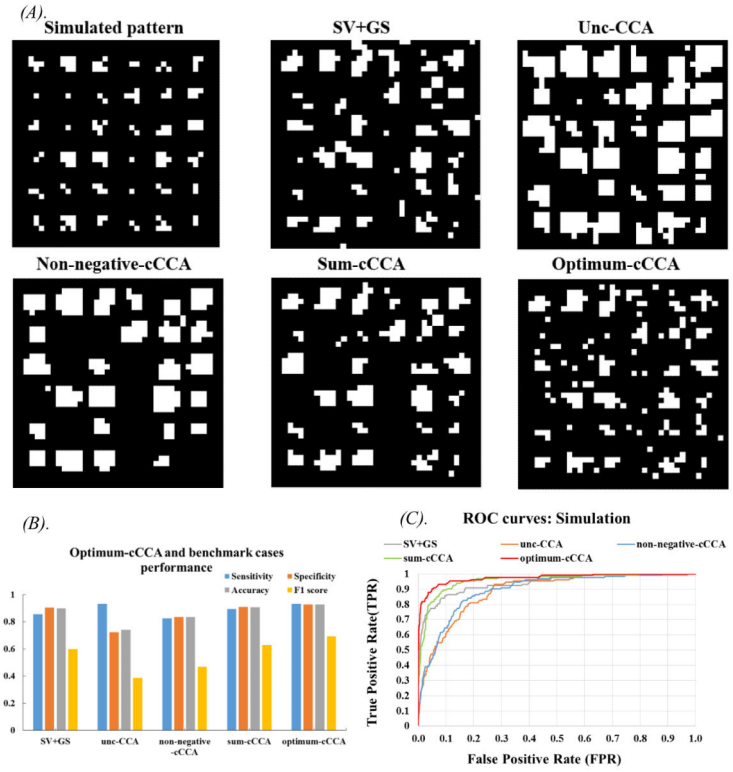

Figure 5.

Results for simulation with noise fraction $f = 0.8$. *(A).* Simulated patterns and spatial activation maps obtained from 5 methods (4 benchmarks and 1 optimum-cCCA), thresholded at the point with lowest total error rate given by (FPR + FNR) . *(B).* Sensitivity (blue), specificity (orange), accuracy (gray) and $F1$ score (yellow), measured by comparing the thresholded activation maps and the simulated pattern. *(C).* ROC curves for the same 5 methods.

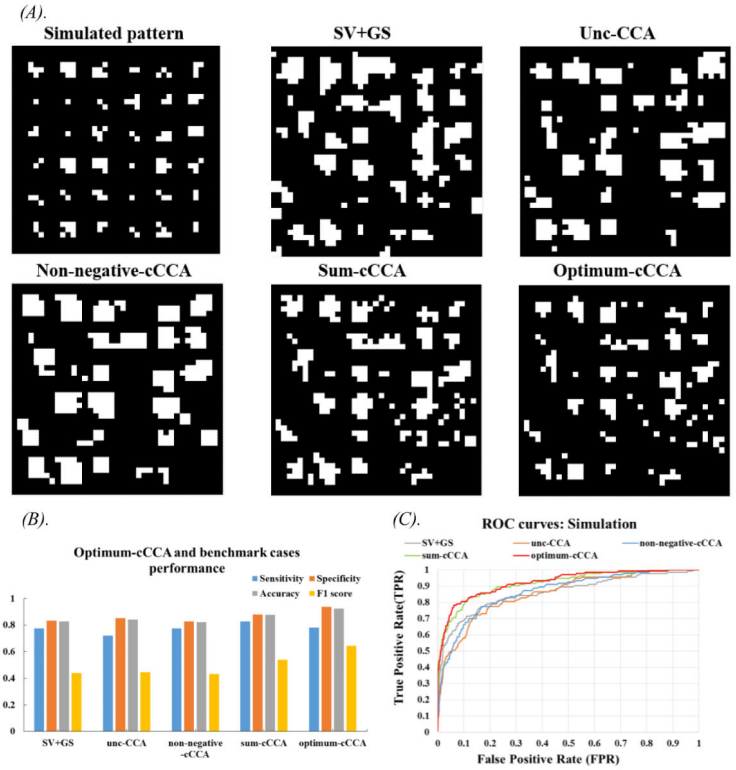

Figure 6.

Results for simulation with noise fraction $f = 0.85$. *(A).* Simulated patterns and spatial activation maps obtained from 5 methods (4 benchmarks and 1 optimum-cCCA), thresholded at the point with lowest total error rate given by min(FPR + FNR) . *(B).* Sensitivity (blue), specificity (orange), accuracy (gray) and $F1$ score (yellow), measured by comparing the thresholded activation maps and the simulated pattern. *(C).* ROC curves for the same 5 methods.

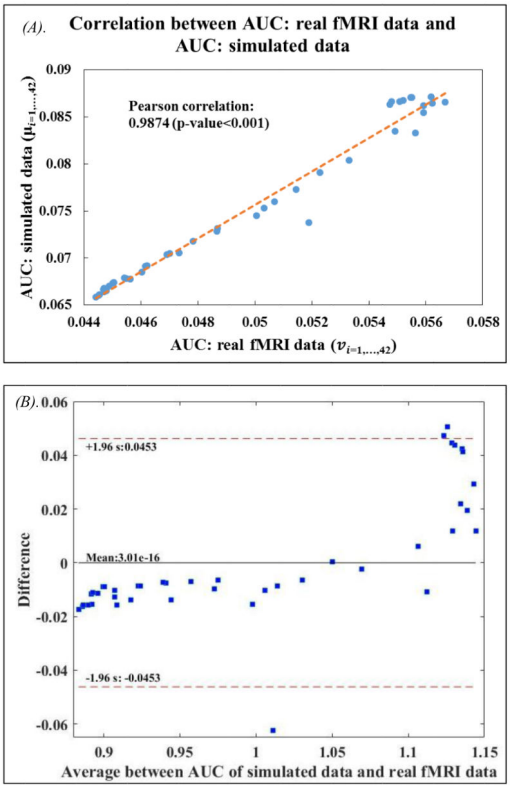

Figure 7.

(A). Correlation plot between AUC (area under the ROC curve integrating over FPR ∈ [0,0.1]), obtained from simulated data ($\mu_{i=1,\ldots,42}$) and real fMRI data ($\nu_{i=1,\ldots,42}$) *(B).* Bland Altman plot for the same two sequences after normalization by each mean value.

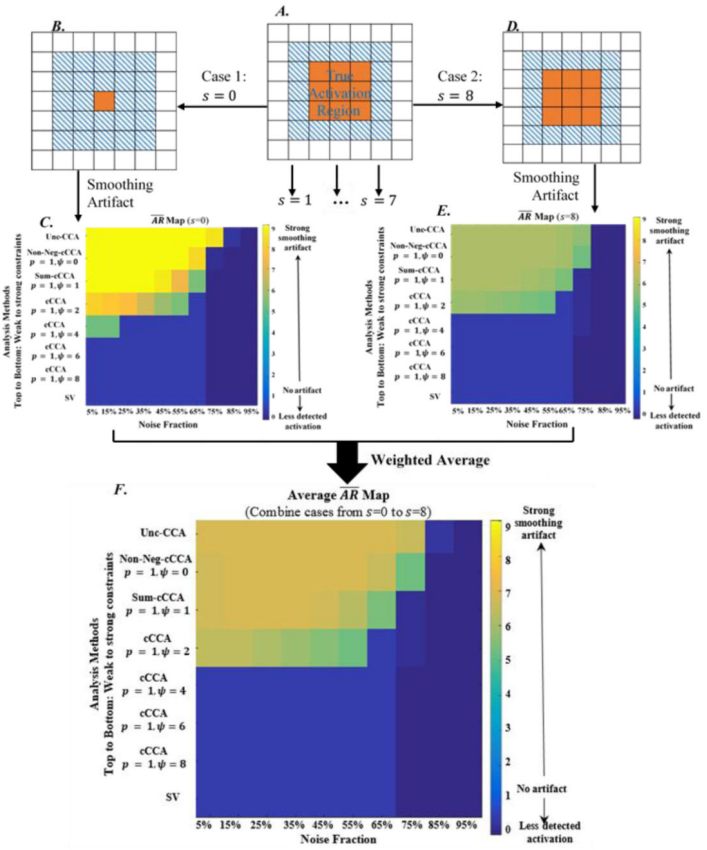

Figure 8.

Smoothing artifact for different analysis methods as a function of noise fraction. *(A).* Simulated 7×7 neighborhoods. True activations are limited to the center 3×3 region (orange part) and all analysis are limited to the center 5×5 area (blue striped part). The center voxel of the true activation region is always kept active and the number of active neighbors in the center 3×3 neighborhoods ranges from 0 to 8. *(B).* One representative case with no active neighbor in the true activation region ($s = 0$), i.e. only the centermost voxel is active (orange area). *(C).* Active ratio $\left(\overline{\mathrm{AR}}\right)$ maps for different analysis methods as a function of noise fraction, averaging over 100 simulated 7×7 $s = 0$ neighborhoods. Horizontal axis labels the noise fraction while vertical axis labels analysis methods, ranked from the weakest to the strongest center dominance constraint in the analysis (top to bottom). (*D).* One representative case with all 8 neighbors are active in the true activation region ($s = 8$). *(E).* Active ratio $\left(\overline{\mathrm{AR}}\right)$ maps for different analysis methods as a function of noise fraction, averaging over 100 simulated 7×7 neighborhoods. *(F).* Average $\overline{\mathrm{AR}}$ map $\left(\overline{\overline{\mathrm{AR}}}\right)$, combining 0 to 8 active neighbor cases in simulated data, obtained with Eq. (22). All activations are thresholded at FWE<0.05. Color bar is from 0 to 9 for *(C), (E)* and *(F).*

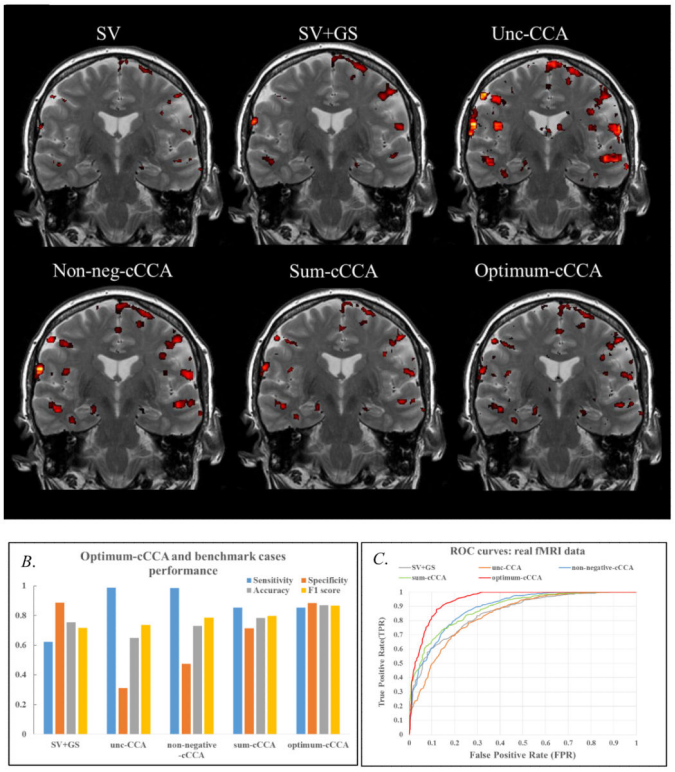

Fig.9.

*(A).*Activation maps with p<0.001 (uncorrected) of a representative subject for contrast Encoding vs Control, obtained from single voxel analysis, four benchmark analysis methods and optimum-cCCA method, overlaid on the subject T2 image. Thresholds are computed from the null distribution for each method. *(B).* Sensitivity (blue), specificity (orange), accuracy (gray) and *F*1 score (yellow), measured by comparing the thresholded activation maps (p<0.05, FWE) and the apparent ground truth. *(C).* ROC curves for the same 5 methods (four benchmark method and optimum-cCCA method).

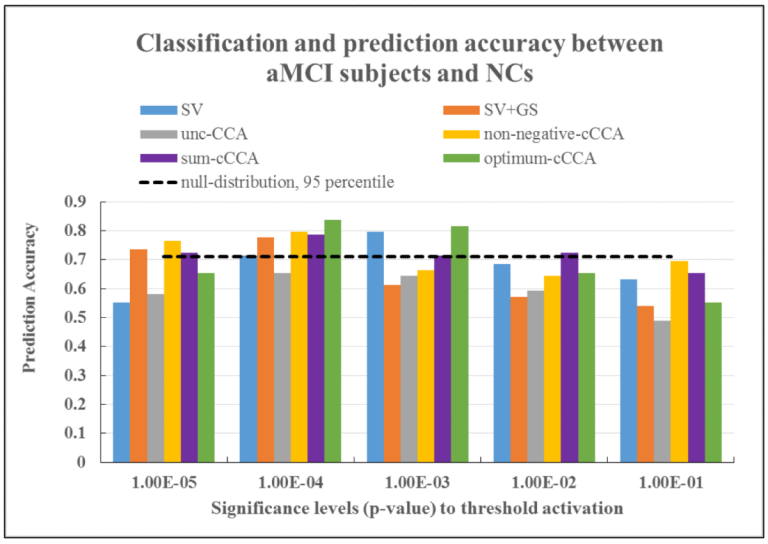

Figure 10.

Classification and prediction accuracy between aMCI subjects and normal controls based on activation percentages obtained from different analysis methods (single voxel analysis: blue bars; SV+GS: orange bars; unc-CCA: grey bars; non-negative-cCCA: yellow bars; sum-cCCA: purple bars and optimum-cCCA: green bars) at various significance levels. Dashed black line represents 95th percentile of the null distribution of the predication accuracy calculated from the permutation test.

Table1

Computation time and accuracy for each method solving the cCCA problem with (p, ψ) = (2,1), averaging over 700 3×3 local neighborhoods.

	Computation Time of one 3×3 region	**Accuracy Prob$\lvert\rho - \rho_{benchmark}\rvert$ < 0.001**	**Accuracy Prob$\lvert\rho - \rho_{benchmark}\rvert$ < 0.01**
(a) BFGS with 1 starting point for entire 3×3 region;	0.13s	85.59%	91.20%
(b) BFGS with 1 starting point for each 256 sub-problems	7.48s	100.00%	100.00%
(c) Augmented Lagrangian with 1 starting point for entire 3×3 region	1.81s	95.41%	96.68%
(d) Reduced Gradient with 1 starting point for entire 3×3 region	0.14s	95.41%	96.56%
(e) Reduced Gradient with 9 starting points for entire 3×3 region	0.74s	99.62%	100.00%

Table 2

Area under ROC (AUC) for simulated data using different analysis methods.

(A). Noise fraction $f = 0.8$

Analysis Method	AUC
SV	0.0858
SV+GS	0.0609
unc-CCA	0.0430
non-negative-cCCA	0.0446

pψ	1	2	4	8	16	32
0.5	0.0736	0.0865	0.0869	0.0866	0.0864	0.0863
1	0.0746	0.0832	0.0865	0.0870	0.0869	0.0867
2	0.0705	0.0758	0.0803	0.0835	0.0854	0.0862
4	0.0678	0.0705	0.0731	0.0754	0.0773	0.0791
8	0.0666	0.0679	0.0691	0.0704	0.0716	0.0729
16	0.0660	0.0667	0.0673	0.0679	0.0686	0.0691
32	0.0658	0.0661	0.0664	0.0667	0.0671	0.0673

(B). Noise fraction $f = 0.85$

Analysis Method	AUC
SV	0.0603
SV+GS	0.0564
unc-CCA	0.0449
non-negative-cCCA	0.0470

pψ	1	2	4	8	16	32
0.5	0.0643	0.0666	0.0627	0.0612	0.0609	0.0608
1	0.0649	0.0683	0.0666	0.0644	0.0628	0.0618
2	0.0626	0.0652	0.0672	0.0678	0.0672	0.0664
4	0.0609	0.0621	0.0635	0.0644	0.0656	0.0665
8	0.0602	0.0607	0.0613	0.0618	0.0624	0.0630
16	0.0599	0.0601	0.0603	0.0606	0.0609	0.0611
32	0.0598	0.0599	0.0600	0.0601	0.0602	0.0603

Note: Analysis methods are the SV analysis, 4 benchmark models and 42 cCCA models with combinations of and $p = 0.5,1,2,4,8,16,32$ and $\psi = 1,2,4,8,16,32$. Sum-cCCA is covered by $(p, \psi) = (1, 1)$. Max-CCA is covered by $(p, \psi) = (32,1)$. Optimum parameter combination of cCCA model is the one that gives highest area under ROC, highlighted in orange. The area under the ROC curve is integrated over the range of FPR = [0, 0.1]. In addition, for cCCA models which give AUCs within 5% less than the optimum-cCCA model, are highlighted in blue.

Table 3

Improvement in detection accuracy of optimum-cCCA model relative to the SV analysis, four benchmark models and max-cCCA model.

SV	SV+GS	unc-cCCA	non-negative-cCCA	sum-cCCA	max-cCCA
1.4%	42.9%	102.4%	95.1%	16.7%	32.3%
13.3%	21.1%	52.2%	45.36%	5.4%	14.2%

Note: Improvement is measured by area under ROC curve integrated over the range of FPR = [0, 0.1].

Table 4

Statistical thresholds at p-value of 1e-5, 1e-4, 1e-3 and 1e-2, obtained from wavelet-resampled resting-state data, F statistics for contrast Encoding v/s Control.

	1e-5	1e-4	1e-3	1e-2
SV	26.88	19.29	12.95	7.58
SV+GS	27.03	20.18	13.77	8.05
unc-CCA	46.97	37.98	29.51	21.05
non-negative-cCCA	38.22	28.45	20.31	13.15
sum-cCCA	33.64	24.38	16.96	10.52
optimum-cCCA	28.40	20.43	13.86	8.19

Table 5

Percentage of active voxels in MTL subregions for different analysis methods.

ROI	Single Voxel	SV+GS	unc-CCA	sum-cCCA	optimum-cCCA
CA1_Left	4.26(8.27)	8.13(9.69)	24.74(21.65)	13.02(14.72)	9.21(12.30)
CA234DG_Left	3.78(5.07)	12.32(14.05)	20.86(17.14)	12.51(12.34)	9.15(9.93)
SUB_Left	6.34(7.02)	14.58(16.33)	29.49(22.68)	18.49(15.42)	13.82(12.32)
ERC_Left	6.47(16.54)	5.78(11.05)	22.50(29.49)	14.54(23.82)	11.69(22.31)
PHC_Left	1.86(2.73)	4.13(5.41)	15.18(13.98)	7.11(6.55)	4.47(4.92)
FUS_Left	10.30(5.75)	16.50(8.69)	30.86(14.68)	22.36(10.73)	18.10(9.27)
CA1_Right	3.17(7.74)	4.24(7.50)	15.80(19.61)	8.52(12.86)	6.43(11.55)
CA234DG_Right	2.17(3.96)	3.70(6.18)	15.36(19.56)	7.49(9.69)	5.06(7.16)
SUB_Right	2.27(3.21)	5.55(5.76)	16.80(17.98)	8.53(9.31)	6.02(6.95)
ERC_Right	2.82(5.52)	3.77(6.12)	14.24(17.99)	8.88(14.09)	6.74(11.91)
PHC_Right	0.67(1.05)	1.88(3.62)	8.21(8.48)	3.68(5.03)	2.71(3.82)
FUS_Right	5.38(4.17)	9.79(7.23)	18.75(13.90)	12.39(7.74)	9.82(6.52)

Note: ROIs are obtained from subject specific masks output from Freesurfer. Percentages of active voxels (i.e. number of active voxels divided by the total number of voxels in ROI) are calculated at different p-value of 0.5, 0.1, 1e-2, 1e-3, 1e-4, 1e-5, (uncorrected). Only active voxels determined for p-value of 0.001 (uncorrected) are listed here as an example. The numbers specify the average over 14 subjects (7 aMCI subjects and 7 NCs) and the standard deviation is listed in parenthesis. CA1: Cornu Ammonis area 1, CA234&DG: combining CA2, CA3, CA4 and Dentate Gyrus, SUB: Subiculum, ERC: Entorhinal Cortex, PHC: Parahippocampal Cortex and FUS: Fusiform Gyrus.