PARTIAL SPARSE CANONICAL CORRELATION ANALYSIS (PSCCA) FOR POPULATION STUDIES IN MEDICAL IMAGING

Paramveer S. Dhillon[⋆]*, Brian Avants*[†]*, Lyle Ungar*[⋆]*, James C. Gee*[†]

[⋆]Department of Computer & Information Science [†]Department of Radiology
University of Pennsylvania, Philadelphia, PA U.S.A

ABSTRACT

We propose a new multivariate method, partial sparse canonical correlation analysis (PSCCA), for computing the statistical comparisons needed by population studies in medical imaging. PSCCA is a multivariate generalization of linear regression that allows one to statistically parameterize imaging studies in terms of multiple views of the population (e.g., the full collection of measurements taken from an image set along with batteries of cognitive or genetic data) while controlling for nuisance variables. This paper develops the theory of PSCCA, provides an algorithm and illustrates PSCCA performance on both simulated and real datasets. We show, as a first application and evaluation of this new methodology, that PSCCA can improve detection power over mass univariate approaches while retaining the interpretability and biological plausibility of the estimated effects. We also discuss the strengths, limitations and future potential of this methodology.

Index Terms— Multivariate modeling, Medical Imaging, Spectral Methods

The number of neuroimaging studies published annually has doubled from 9,938 in 2000-2001 to 19,676 in 2009-2010 (`http://www.ncbi.nlm.nih.gov/pubmed/`). This growth has been accompanied by increasing diversity in the types of data being collected; Imaging studies now often include not only various structural and functional modalities but also neurocognitive batteries, genetics, and environmental measurements. However, the statistical methods have changed relatively little over the past twenty years – until very recently (e.g., [1]). The increasing size of imaging datasets and the concomitant desire for performing integrative studies across modalities points to the need for new multivariate statistical methods that elegantly handle large, multi-view datasets. These methods should retain or even improve detection power over traditional mass-univariate (MU) models such as statistical parametric mapping (SPM) which uses the univariate form of the general linear model (GLM). Repeatedly applying the univariate GLM (or linear regression) at each voxel leads to loss of detection power due to the well-known multiple comparisons problem.

Canonical Correlation Analysis (CCA) [2] is a traditional multivariate generalization of standard linear regression. CCA inherently avoids the multiple-comparisons penalty associated with MU methods by symmetrically maximizing the correlation between the full matrices representing two views of the data (here denoted $\mathbf{Y}$ and $\mathbf{X}$). The matrix $\mathbf{X}$ might represent a tabulation of all demographic data, including genetics, diagnosis, behavioral measures, age, etc. while $\mathbf{Y}$ may be a matrix of all the imaging measurements. In contrast, traditional univariate models only allow the predicted value to be a vector while the predictors may be a matrix.

Recently, sparse (or penalized) canonical covariance analyses (SCCovA) [3, 4, 5] have been proposed as an approximation to CCA specifically for the high dimensional ($p \gg n$) setting.[1] The *sparseness* in penalized methods improves interpretability by including in the model only the most important variables from the large set of p (and/or q) predictors. From a medical imaging researcher's perspective, the benefit is that only the most predictive variables (e.g. parts of the brain) will emerge in the results provided by a penalized statistical tool. Hence, brain regions are highlighted in a way that is similar to SPM. Furthermore, regions selected by SCCovA (or similarly sparse canonical correlation analysis (SCCA)) are treated statistically as a collective (or 'network') as opposed to MU methods which treat each predictor as an independent variable.

Despite prior studies using SCCovA and SCCA [6], we are unaware of previous work that studies factoring ("partialling") out nuisance variables within the penalized CCA framework. While this problem is addressed in the $p < n$ setting by partial canonical correlation analysis (PCCA)[7], no penalized formulation has yet been proposed.

This paper contributes the theory of Partial Sparse CCA (PSCCA) along with a novel and efficient iterative algorithm for PSCCA. PSCCA (like CCA) performs a global multivariate test of the association between two modalities that quantify a study's subjects while accounting for a third set of nuisance variables. It generalizes linear regression and is inherently, sparsely multivariate in multiple views of the data un-

[1] SCCovA substitutes the identity matrix for within-view covariance matrices and thus analyze cross-covariance structure, not correlation structure. Thus, SCCA (unlike SCCovA) does not depend on how the observations are scaled.

like MU and standard support vector machines (SVM).

The general PSCCA formulation has many applications. PSCCA may be applied to almost any statistical scenario in medical imaging studies traditionally handled by SPM. PSCCA is able to identify the subset of the brain most correlated with non-imaging variable(s) of interest (for instance, a cognitive battery) while factoring out confounding effects (age, gender). Alternatively, we may apply PSCCA to the case where both views of the data are high-dimensional, for instance, to identify correlations between different imaging modalities independently from covariates such as scanner, gender, etc. PSCCA thus enables complex studies of multiple view data that contains many more variables than observations.

1. BRIEF REVIEW: CCA AND SPARSE CCA (SCCA)

More specifically, given a set of n paired observation vectors $\{(y_1, x_1), ..., (y_n, x_n)\}$–in our case the two matrices are the quantitative imaging measurement ($\mathbf{Y}$) and age, gender, diagnosis ($\mathbf{X}$) matrices –we would like to simultaneously find the directions $\boldsymbol{\phi}_\mathbf{Y}$ and $\boldsymbol{\phi}_\mathbf{X}$ that maximize the correlation of the projections of $\mathbf{Y}$ onto $\boldsymbol{\phi}_\mathbf{Y}$ with the projections of $\mathbf{X}$ onto $\boldsymbol{\phi}_\mathbf{X}$. This is expressed as

$$\rho = \max_{\phi_Y, \phi_X} \frac{\boldsymbol{\phi}_\mathbf{X}^T \boldsymbol{\Sigma}_{\mathbf{XY}} \boldsymbol{\phi}_\mathbf{Y}}{\sqrt{\boldsymbol{\phi}_\mathbf{X}^T \boldsymbol{\Sigma}_{\mathbf{XX}} \boldsymbol{\phi}_\mathbf{X}} \sqrt{\boldsymbol{\phi}_\mathbf{Y}^T \boldsymbol{\Sigma}_{\mathbf{YY}} \boldsymbol{\phi}_\mathbf{Y}}} \tag{1}$$

where $\boldsymbol{\Sigma}_{\mathbf{XX}}$, $\boldsymbol{\Sigma}_{\mathbf{YY}}$ and $\boldsymbol{\Sigma}_{\mathbf{XY}}$ are the auto and cross covariance matrices i.e. $\mathbf{X}^T\mathbf{X}$, $\mathbf{Y}^T\mathbf{Y}$ and $\mathbf{X}^T\mathbf{Y}$, respectively. The above objective can also be thought of as maximizing the numerator $\boldsymbol{\phi}_\mathbf{X}^T \boldsymbol{\Sigma}_{\mathbf{XY}} \boldsymbol{\phi}_\mathbf{Y}$ subject to $\boldsymbol{\phi}_\mathbf{X}^T \boldsymbol{\Sigma}_{\mathbf{XX}} \boldsymbol{\phi}_\mathbf{X} = 1$ and $\boldsymbol{\phi}_\mathbf{Y}^T \boldsymbol{\Sigma}_{\mathbf{YY}} \boldsymbol{\phi}_\mathbf{Y} = 1$

Now, define change of basis as:

$$\boldsymbol{\psi}_\mathbf{X} = \boldsymbol{\Sigma}_{\mathbf{XX}}^{1/2} \boldsymbol{\phi}_\mathbf{X}, \qquad \boldsymbol{\psi}_\mathbf{Y} = \boldsymbol{\Sigma}_{\mathbf{YY}}^{1/2} \boldsymbol{\phi}_\mathbf{Y} \tag{2}$$

Then, substituting (2) in (1) we get

$$\rho = \max_{\boldsymbol{\psi}_\mathbf{Y}, \boldsymbol{\psi}_\mathbf{X}} \frac{\boldsymbol{\psi}_\mathbf{X}^T \boldsymbol{\Sigma}_{\mathbf{XX}}^{-1/2} \boldsymbol{\Sigma}_{\mathbf{XY}} \boldsymbol{\Sigma}_{\mathbf{YY}}^{-1/2} \boldsymbol{\psi}_Y}{\|\boldsymbol{\psi}_\mathbf{X}\| \|\boldsymbol{\psi}_\mathbf{Y}\|} \tag{3}$$

The whitening transform is used to convert covariances to correlations and also to de-correlate auto-correlation matrices. In CCA, this normalizes the data such that the optimization can maximize the cross-correlation. The standard whitening transform is defined as $\mathbf{X}_w = \mathbf{X}\boldsymbol{\Sigma}_{\mathbf{XX}}^{-1/2}$ and $\mathbf{Y}_w = \mathbf{Y}\boldsymbol{\Sigma}_{\mathbf{YY}}^{-1/2}$. Applying the whitening transform to (3)

$$Corr(\mathbf{X}_w \boldsymbol{\psi}_\mathbf{X}, \mathbf{Y}_w \boldsymbol{\psi}_\mathbf{Y}) = \rho = \max_{\boldsymbol{\psi}_\mathbf{Y}, \boldsymbol{\psi}_\mathbf{X}} \frac{\boldsymbol{\psi}_\mathbf{X}^T \boldsymbol{\Sigma}_{\mathbf{X}_w \mathbf{Y}_w} \boldsymbol{\psi}_Y}{\|\boldsymbol{\psi}_\mathbf{X}\| \|\boldsymbol{\psi}_\mathbf{Y}\|} \tag{4}$$

where $\boldsymbol{\Sigma}_{\mathbf{X}_w \mathbf{Y}_w} = \mathbf{X}_w^T \mathbf{Y}_w$.

As mentioned earlier, CCA results in vectors $\boldsymbol{\psi}_\mathbf{X}$, $\boldsymbol{\psi}_\mathbf{Y}$ that are not sparse, and these vectors are not unique if $p > n$. In most biomedical imaging applications, p is large and, one needs to find a linear combination of the variables in $\mathbf{X}_w$ and $\mathbf{Y}_w$ that has large correlation but is also sparse in the variables that enter the model.

While several researchers propose sparse formulations of canonical covariance analysis [3, 4, 5], none of the these methods handle confounding variables—a highly desirable modeling property for many biomedical and neuroimaging applications. In the next section, we detail the PSCCA solution to this problem.

2. PSCCA (PARTIAL SPARSE CANONICAL CORRELATION ANALYSIS)

As described earlier, let $\mathbf{X}$ be the matrix with columns containing voxels from one set of images of n subjects; $\mathbf{Y}$ is the matrix with columns containing the second set of measurements from the same n subjects and further let $\mathbf{Z}$ be the matrix of confounding variables (age, gender, etc.) for our neuroimaging problem. The second set of measurements may be voxels from another imaging modality, scores from a battery of neuropsychological tests or a much simpler feature such as a binary diagnosis variable. Also, let $\lambda_\mathbf{X}$ and $\lambda_\mathbf{Y}$ ($\in [0, 1]$) (where higher values indicate more sparsity) be the user defined parameters which control the sparsity for either set of the canonical variates. The sparseness parameters can, alternatively, be chosen automatically from the data so as to maximize the correlation (or likelihood) between the canonical variates.

PCCA [7] finds the correlation between $\mathbf{X}$ and $\mathbf{Y}$ after removing ("partialling out") the linear effect of the confounding variables $\mathbf{Z}$. We denote the $\mathbf{X}$ and $\mathbf{Y}$ matrices with effect of $\mathbf{Z}$ "partialled" out as $\mathbf{X}^{\setminus \mathbf{Z}}$ and $\mathbf{Y}^{\setminus \mathbf{Z}}$. Regressing $\mathbf{X}$ against $\mathbf{Z}$, using standard least squares ($\|\mathbf{X} - \mathbf{Z}\boldsymbol{\beta}\|^2$) gives $\boldsymbol{\beta} = \boldsymbol{\Sigma}_{\mathbf{ZZ}}^{-1} \mathbf{Z}^T \mathbf{X}$. Thus, the residual[2] can be written as $\mathbf{X}^{\setminus \mathbf{Z}} = \mathbf{X} - \mathbf{Z}\boldsymbol{\Sigma}_{\mathbf{ZZ}}^{-1} \mathbf{Z}^T \mathbf{X}$. Applying the whitening transform to $\mathbf{Z}$ as $\mathbf{Z}_w = \mathbf{Z}\Sigma_{\mathbf{ZZ}}^{-1/2}$, we get $\mathbf{X}^{\setminus \mathbf{Z}} = \mathbf{X} - \mathbf{Z}_w \mathbf{Z}_w^T \mathbf{X}$. We can write similar equations for the residual when $\mathbf{Y}$ is regressed against $\mathbf{Z}$.

Now, we can write the complete variance-covariance matrix of the residuals as:

$$\begin{bmatrix} \boldsymbol{\Sigma}_{\mathbf{XX}}^{\setminus \mathbf{Z}} & \boldsymbol{\Sigma}_{\mathbf{XY}}^{\setminus \mathbf{Z}} \\ \boldsymbol{\Sigma}_{\mathbf{YX}}^{\setminus \mathbf{Z}} & \boldsymbol{\Sigma}_{\mathbf{YY}}^{\setminus \mathbf{Z}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T\mathbf{X} - \mathbf{X}^T\mathbf{Z}_w\mathbf{Z}_w^T\mathbf{X} & \mathbf{X}^T\mathbf{Y} - \mathbf{X}^T\mathbf{Z}_w\mathbf{Z}_w^T\mathbf{Y} \\ \mathbf{Y}^T\mathbf{X} - \mathbf{Y}^T\mathbf{Z}_w\mathbf{Z}_w^T\mathbf{X} & \mathbf{Y}^T\mathbf{Y} - \mathbf{Y}^T\mathbf{Z}_w\mathbf{Z}_w^T\mathbf{Y} \end{bmatrix}$$

The PCCA problem can therefore be written as:

$$\rho_{PCCA} = \max_{\boldsymbol{\phi}_\mathbf{Y}, \boldsymbol{\phi}_\mathbf{X}} \frac{\boldsymbol{\phi}_\mathbf{X}^T \boldsymbol{\Sigma}_{\mathbf{XY}}^{\setminus \mathbf{Z}} \boldsymbol{\phi}_Y}{\sqrt{\boldsymbol{\phi}_\mathbf{X}^T \boldsymbol{\Sigma}_{\mathbf{XX}}^{\setminus \mathbf{Z}} \boldsymbol{\phi}_\mathbf{X}} \sqrt{\boldsymbol{\phi}_\mathbf{Y}^T \boldsymbol{\Sigma}_{\mathbf{YY}}^{\setminus \mathbf{Z}} \boldsymbol{\phi}_\mathbf{Y}}} \tag{5}$$

[2] Note that $\mathbf{X}^{\setminus \mathbf{Z}}$ is actually what is called the residual $\mathbf{X} - \mathbf{Z}\boldsymbol{\beta}$ in a least squares regression problem.

After some algebraic manipulation we can write the PCCA objective compactly as

$$\rho_{PCCA} = \max_{\psi_{\mathbf{Y}}, \psi_{\mathbf{X}}} \frac{\psi_{\mathbf{X}}^T \Sigma_{\mathbf{X}_w \mathbf{Y}_w}^{\setminus \mathbf{Z}} \psi_{\mathbf{Y}}}{\|\psi_{\mathbf{X}}\| \|\psi_{\mathbf{Y}}\|} \tag{6}$$

where $\mathbf{X}_w = \mathbf{X}(\Sigma_{\mathbf{XX}}^{\setminus \mathbf{Z}})^{-1/2}$ and $\mathbf{Y}_w = \mathbf{Y}(\Sigma_{\mathbf{YY}}^{\setminus \mathbf{Z}})^{-1/2}$.

Finally, the above objective after incorporating the user specified ℓ_1 sparsity penalties ($\lambda_{\mathbf{X}}$ and $\lambda_{\mathbf{Y}}$) and under the constraints $\psi_{\mathbf{X}}^T \psi_{\mathbf{X}} = \psi_{\mathbf{Y}}^T \psi_{\mathbf{Y}} = 1$ can be written as:

$$\rho_{PSCCA} = \max_{\psi_{\mathbf{Y}}, \psi_{\mathbf{X}}} \{\psi_{\mathbf{X}}^T \Sigma_{\mathbf{X}_w \mathbf{Y}_w}^{\setminus \mathbf{Z}} \psi_{\mathbf{Y}} - \lambda_{\mathbf{X}} \|\psi_{\mathbf{X}}\|_1 - \lambda_{\mathbf{Y}} \|\psi_{\mathbf{Y}}\|_1\} \tag{7}$$

Our optimization strategy for (7) combines power iteration and soft thresholding to compute the canonical vectors while satisfying the sparsity constraints. The approach, described in the next section, uses an alternating least squares method [8] extended to include sparsity constraints.

2.1. PSCCA Algorithm

Following [8], we propose a power iteration based algorithm for PSCCA for the general problem of finding principal eigenvectors of the matrices. This numerical approach does not require one to ever explicitly form the full $\mathbf{X}_w^T \mathbf{Y}_w$ matrix and is therefore appropriate for large datasets where the number of columns in both views may count in the millions or more. In all steps below, we employ the pseudoinverse when needed. In addition, the function $(x)_+$ is equal to x if $x \geq 0$ and 0 if $x < 0$ and

$$Sign(x) = \begin{cases} -1, & \text{if } x < 0 \\ 0, & \text{if } x = 0 \\ 1, & \text{if } x > 0 \end{cases} \tag{8}$$

Note that positivity or negativity constraints on the $\psi_{\mathbf{X}}, \psi_{\mathbf{Y}}$ may be trivially included with a minor modification to Algorithm 1. We use permutation testing on $\mathbf{X}$, $\mathbf{Y}$ to assess significance where the test statistic is the partial correlation between the two main views.

3. RESULTS

The code for the PSCCAN implementation, the simulation study and the neuroimaging study will be made available at publication time.

3.1. Simulations

Define a "true" linear signal vector with n entries, $\boldsymbol{v}$, such that the value of each entry is $\boldsymbol{v}_i = i/n$ where i indexes the vector. A second signal is a vector drawn from a zero mean unit variance Gaussian distribution, $\boldsymbol{g}_x$ with p entries. The first view is then $\mathbf{X} = \boldsymbol{v}^T \boldsymbol{g}_x$ and we similarly generate $\mathbf{Y}$ with $n \times q$ entries. We optionally add noise to both views. In 100 low-noise simulations, SCCA produces a significant association. However, when we use $\mathbf{Z} = \boldsymbol{v}+$ *noise* as a confounding covariate in PSCCA on $\mathbf{X}$ and $\mathbf{Y}$, then no significant association exists. Both results are as expected and provide a sanity check on our theory and implementation. The second experimental validation of our implementation and theory generates $\mathbf{X}$ and $\mathbf{Y}$ where the first $p/2, q/2$ columns are derived from $\boldsymbol{v}$. The second $p/2, q/2$ columns in $\mathbf{X}, \mathbf{Y}$ are derived from a different "true" signal ($\boldsymbol{v}_2$) with a less strong linear relationship than in the first half of the matrices. Thus, when we use SCCA with sparseness $\lambda_{\mathbf{X}} = \lambda_{\mathbf{Y}} = 0.25$, the first half of the matrix is selected. PSCCA selects the second half of the matrix when $\mathbf{Z}$ is used as confounding covariate. Both are significant across permutations. Due to noise, in some simulations, a few entries from the first half of the matrix may enter the model with low weight. If we add a column containing signal derived from $\boldsymbol{v}_2$ to $\mathbf{Z}$ then, as predicted, PSCCA results become insignificant. Figure 1 shows the vectors $\phi_{\mathbf{X}}$ selected by SCCA and PSCCA on the same input data where PSCCA uses $\mathbf{Z}$ (derived from $\boldsymbol{v}$ alone) as confounding covariate.

Algorithm 1 Computing principal eigenvectors for PSCCA

1: Apply the whitening transformation to $\mathbf{Z}$ to get $\mathbf{Z}_w$.
2: Compute $\mathbf{X}^{\setminus \mathbf{Z}}$ and $\mathbf{Y}^{\setminus \mathbf{Z}}$ and the whitened matrices $\mathbf{X}_w$ and $\mathbf{Y}_w$.
3: Select the (fractional) sparsity parameters $\lambda_{\mathbf{X}}$ and $\lambda_{\mathbf{Y}}$
4: Randomly initialize $\psi_{\mathbf{X}}^0$ and $\psi_{\mathbf{Y}}^0$ ($\sim \mathcal{N}(0,1)$) and set $k = 0$.
5: **while** Δ Corr($\boldsymbol{X}_w \psi_{\mathbf{X}}^{k+1}$, $\boldsymbol{Y}_w \psi_{\mathbf{Y}}^{k+1}$) $< \epsilon$ **do**
6: Compute $\psi_{\mathbf{X}}^{k+1} = {\mathbf{X}_w}^T \mathbf{Y}_w \psi_{\mathbf{Y}}^k - {\mathbf{X}_w}^T \mathbf{Z}_w {\mathbf{Z}_w}^T \mathbf{Y}_w \psi_{\mathbf{Y}}^k$
7: Soft-Max Sparseness: $\psi_{\mathbf{X}}^{k+1} \leftarrow (\|\psi_{\mathbf{X}}^{k+1}\| - max(\psi_{\mathbf{X}}^{k+1}) * \lambda_{\mathbf{X}})_+ Sign(\psi_{\mathbf{X}}^{k+1})$
8: Normalize: $\psi_{\mathbf{X}}^{k+1} \leftarrow \frac{\psi_{\mathbf{X}}^{k+1}}{\|\psi_{\mathbf{X}}^{k+1}\|}$
//Repeat Same Procedure for $\psi_{\mathbf{Y}}$
9: Compute $\psi_{\mathbf{Y}}^{k+1} = {\mathbf{Y}_w}^T \mathbf{X}_w \psi_{\mathbf{X}}^{k+1} - {\mathbf{Y}_w}^T \mathbf{Z}_w {\mathbf{Z}_w}^T \mathbf{X}_w \psi_{\mathbf{X}}^{k+1}$
10: Soft-Max Sparseness: $\psi_{\mathbf{Y}}^{k+1} \leftarrow (\|\psi_{\mathbf{Y}}^{k+1}\| - max(\psi_{\mathbf{Y}}^{k+1}) * \lambda_{\mathbf{Y}})_+ Sign(\psi_{\mathbf{Y}}^{k+1})$
11: Normalize: $\psi_{\mathbf{Y}}^{k+1} \leftarrow \frac{\psi_{\mathbf{Y}}^{k+1}}{\|\psi_{\mathbf{Y}}^{k+1}\|}$
12: k ← k+1
13: **end while**

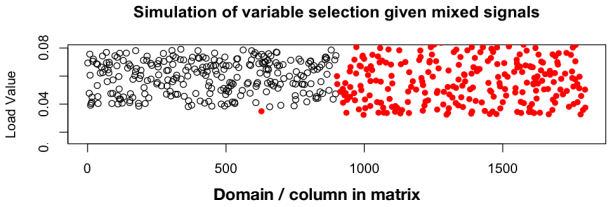

Fig. 1. The black hollow circles show the non-zero entries in $\phi_{\mathbf{X}}$ that are selected by SCCA, that is, the value of the vector $\phi_{\mathbf{X}}$. The red full circles show the non-zero entries in the vector $\phi_{\mathbf{X}}$ that are selected by PSCCA. The $\mathbf{Z}$ signal factors out the confounding signal in the first half of the matrix leaving the second signal of interest in the second half to be the source of the significant association.

3.2. Comparison of regression and PSCCA on OASIS data

Our first evaluation on real data employs PSCCA as a form of multivariate regression between imaging, diagnosis and nuisance variables. We employ a subset of the freely available OASIS dataset to compare PSCCA to mass-univariate linear regression. This subset of the OASIS data contains elderly subjects (n=38) in addition to subjects with Alzheimer's disease (n=31) of both genders (39 F, 30 M) and with ages that range between 62 and 98 years. Our evaluation criterion compares both methods' power to detect the known anatomical distribution of AD-related atrophy in gray matter (hippocampus, cuneus, temporal lobe) [6] where gray matter was segmented and normalized by using standard open source software. We use the whole brain, in template space, as region of interest in order to challenge the power of the MU method relative to the single test performed by multivariate PSCCA. We assume that the researcher has pre-selected the sparseness parameter for the study. We choose λ (sparsity parameter) for the gray matter voxels such that 10% of the ROI (contained in the $\mathbf{X}$ matrix) will be selected by PSCCA. The $\mathbf{Y}$ matrix, in this case, is the diagnosis vector that defines whether a subject is control or patient. The nuisance matrix $\mathbf{Z}$ contains age and gender variables. We run both the MU statistics (via the **R** program) and our own independently developed PSCCA implementation (C++ based, BSD license, open-source) on identical input data. Using false discovery rate (FDR) correction on the regression-based p-values for diagnosis, we find that the minimum q-value is 0.183, thus insignificant after correction. In contrast, PSCCA shows significant effects at the $p = 0.041$ level, 10000 permutations. We visualize the regions that emerge from PSCCA by overlaying the first canonical vector $\psi_{\mathbf{X}}$ on the brain. Figure 2 compares the PSCCA output with the regression results overlaid on the brain at the level of $p = 0.01$ uncorrected.

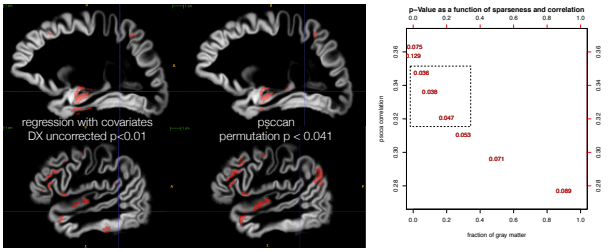

Fig. 2. PSCCAN (right) versus mass-univariate uncorrected statistics (left). Both methods reveal similar areas of the brain. However, the mass-univariate results cannot be considered significant (after FDR correction) due to the multiple comparisons problem. It is possible that another correction method would retain some of the mass-univariate effects but we choose FDR because it is standard and only moderately conservative. We show, at right, the relationship of estimated significance to variations in the sparseness parameter (for the image voxel matrix $\mathbf{X}$) and PSCCA correlation. The significant region is outlined in a dashed box. In a real study, one would only use the pre-selected sparseness parameter.

4. DISCUSSION AND CONCLUSION

In this paper we proposed a new statistical tool that is ideal for multivariate imaging studies. Results on synthetic and real world data (OASIS) further corroborate our hypothesis that PSCCA is able to increase detection power in the presence of covariates and extract biologically plausible, multivariate patterns from neuroimaging data. Specifically, PSCCA reveals significant patterns of difference between elderly and AD subjects that are within brain regions known to be affected by Alzheimer's tauopathy. Although the MU model fails to reveal significant effects, there is notable similarity between regions selected by PSCCA and those voxels in the brain that had uncorrected p-value < 0.01. In our experiments we only use the primary eigenvector from PSCCA; Future work will analyze the effect of including additional eigenvectors and will seek to further investigate alternatives for assessing PSCCA significance in interpretable ways. Finally, as in standard correlation, one should take care to visualize PSCCA results to investigate the potential impact of outliers.

5. REFERENCES

[1] D. Tosun, P. Mojabi, M. W. Weiner, and N. Schuff, "Joint analysis of structural and perfusion MRI for cognitive assessment and classification of Alzheimer's disease and normal aging.," *Neuroimage*, vol. 52, no. 1, pp. 186–197, Aug 2010.

[2] H. Hotelling, "Canonical Correlation Analysis (CCA)," *Journal of Educational Psychology*, 1935.

[3] E. Parkhomenko, D. Tritchler, and J. Beyene, "Sparse Canonical Correlation Analysis with Application to Genomic Data Integration," *Statistical Applications in Genetics and Molecular Biology*, vol. 8, no. 1, 2009.

[4] D. M. Witten, R. Tibshirani, and T. Hastie, "A penalized matrix decomposition, with applications to Sparse Principal Components and Canonical Correlation Analysis," *Biostatistics*, vol. 10, no. 3, pp. 515–534, July 2009.

[5] A. Lykou and J. Whittaker, "Sparse CCA using a Lasso with positivity constraints," *Computational Statistics & Data Analysis*, vol. 54, no. 12, pp. 3144–3157, 2010.

[6] B. B. Avants, P. A. Cook, L. Ungar, J. C Gee, and M. Grossman, "Dementia induces correlated reductions in white matter integrity and cortical thickness: A multivariate neuroimaging study with sparse canonical correlation analysis.," *Neuroimage*, vol. 50, no. 3, pp. 1004–1016, Apr 2010.

[7] N. Timm and J. Carlson, "Part and Bipartial Canonical Correlation Analysis," *Psychometrika*, vol. 41, pp. 159–176, 1976.

[8] G. H. Golub and H. Zha, "The Canonical Correlations of Matrix Pairs and their Numerical Computation," Tech. Rep., Stanford University, Stanford, CA, USA, 1992.