Anatomically-Constrained PCA for Image Parcellation

Paramveer S. Dhillon⋆, James C. Gee†, Lyle Ungar⋆, Brian Avants†
⋆Department of Computer & Information Science †Department of Radiology
University of Pennsylvania, Philadelphia, PA U.S.A

Abstract—**Traditionally clinicians and medical researchers have been using either totally data driven approaches like PCA/CCA/ICA or ROI based analysis for exploratory analysis of brain images. However, PCA/CCA/ICA based approaches suffer from lack of interpretability of results and on the other hand ROI based approaches are too rigid and wrongly assume that the signal lies totally within a predefined region. In this paper, we propose a novel approach which stands in stark contrast with both these approaches as it borrows strength from both these paradigms and leads to statistically refined definitions of ROIs based on information from data. Our approach, called Anatomically Constrained PCA (AC-PCA) provides a principled way of incorporating prior information in the form of probabilistic or binary ROIs while still allowing the data to softly modify the original ROI definitions. Experimental results on cortical thickness images show the superiority of AC-PCA for MCI classification compared to ROI and unconstrained PCA (a totally data based approach).**

Keywords-**ROI, prior, PCA**

I. INTRODUCTION AND RELATED WORK

Over the last few decades there have been significant breakthroughs in medical imaging machinery which has led to an increase in the amount and diversity of data being available e.g. structural and functional modalities, neurocognitive batteries, genetics, and environmental measurements etc. This has lead to a substantial interest in using sophisticated statistical methods to analyze and explore this data. Methods like Principal Component Analysis (PCA), Independent Component Analysis (ICA), Canonical Correlation Analysis (CCA) and their robust and sparse variants [1], [2] have been the workhorse of brain and neuro imaging fields as they provide key insights into the data in a totally data driven way.

However, one potential pitfall of these approaches, which has made clinicians and other researchers cautious with their use is the lack of control over the areas they highlight. Since, they work in a totally data driven way, the end clinician/researcher has little control over the areas of brain they chose. Clinicians usually have some form of prior knowledge as to which areas may contain the signal they are looking for, for example, someone studying fronto-temporal dementia would expect some or most of the signal to lie in frontal cortex, however if the voxels in frontal cortex don't explain the variance in data these approaches won't highlight them.

So, this has led the clinicians and researchers to work with

Input Images

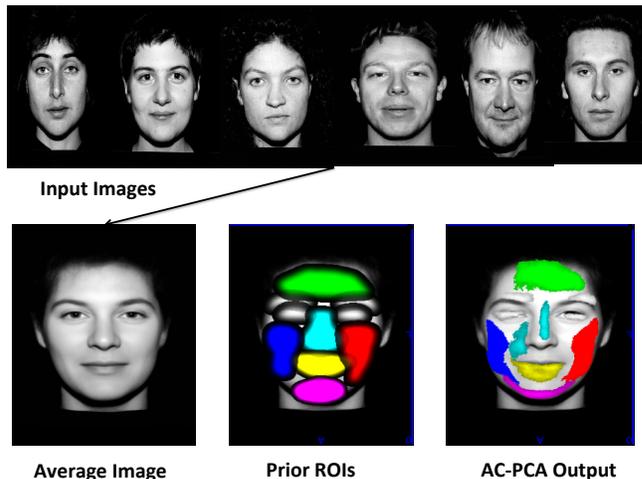**Average Image**   **Prior ROIs**   **AC-PCA Output**

Figure 1. AC-PCA pipeline: 1) Take in raw images 2). Register them and create an average image 3). Define Prior ROIs 4). Run (AC-PCA) which weighs both data and prior ROIs

totally prior driven approaches e.g. ROI (Region of Interest) analysis [3], where they chose a pre-defined region manually or based on past study (for instance *Brodmann Areas*) and study it exclusively. The assumption is that the ROI is correct and contains all relevant signal. However, important signal may have slightly different boundaries than the scientist's conception. The data representation (or spatially varying noise) may also lead to strong or weak signal within different parts of the ROI. Such dataset specific information is not taken into account by a traditional ROI. When effects are localized to the selected region, and that region is well-defined, an ROI analysis may provide the most sensitive testing method. However, some conditions involve a network of regions that may not be fully identified. In such cases—in addition to the general case of an exploratory analysis in a small dataset—dimensionality reduction may provide advantages over an ROI analysis or mass univariate voxel-based morphometry [4].

In this paper, we propose a method that combines strengths from focused ROI analysis and the unbiased and exploratory advantages of dimensionality reduction tools such as PCA. Our approach, called Anatomically Constrained PCA (AC-PCA), allows an initial binary or probabilistic ROI to adapt to the underlying covariation within the data and thus has advantages of PCA. At the same

time, AC-PCA maintains proximity to (and the locality of) the original region and thus has advantages of the standard ROI approach. AC-PCA also maintains positivity in the estimated anatomically-constrained eigenvector, thereby keeping ROI interpretability. This allows us to modify the definitions of labels to capture the variation in dataset while still staying close to the initial ROI definitions. AC-PCA therefore produces labelings with soft weighted averages and as we show in the experimental section, are more sensitive to the underlying brain data than a standard ROI.

Our approach provides a principled way of incorporating priors in a totally data driven approach based on PCA. Our optimization objective provides a tradeoff between 1). staying close to the initial ROI definitions and 2). allowing data to lead the exploratory analysis by explaining variance through PCA. A good way to think about this is as ROI definitions forcing us to be *conservative* and staying close to the initial definitions e.g. *Brodmann's* areas; on the other hand the PCA component gives us *liberty* to be more exploratory and just trust the given data. The tradeoff between the two competing paradigms is defined by user tunable (prior strength) parameter, which is chosen via cross validation on a separately held validation set. Figure 1 gives an example of proof of concept of our approach using faces and also explain our general experimental pipeline used throughout this paper.

Recently, [5] and [6] have proposed constrained versions of ICA for EEG and fMRI applications respectively, however we are not aware of any previous work on anatomically constrained PCA.

The remaining paper is organized as follows; in next section we provide the details of AC-PCA and provide an algorithm which implements our approach. In Section 3, we provide experimental results on real datasets showing the efficacy of our approach and its superiority to both the alternatives mentioned above.

II. ANATOMICALLY CONSTRAINED PRINCIPAL COMPONENT ANALYSIS: AC-PCA

Lets assume that we are given a data matrix $\mathbf{X}$ where each row represents a patient (total $\mathbf{n}$ of them) and each column represents a voxel in image (total $\mathbf{p}$ of them). In practice, we usually have $\mathbf{p} \gg \mathbf{n}$. Also, lets assume that we have a prior matrix $\mathbf{M}(\lambda)$ (function of a parameter λ which controls its strength) whose each row corresponds to a separate prior e.g. different Brodmann areas (total $\mathbf{k}$ of them) and each column (total $\mathbf{p}$ of them) contains the probability of a particular voxel belonging to that prior.

A totally data driven approach like vanilla unconstrained PCA would ignore the prior matrix $\mathbf{M}$ completely and just try to find combinations of voxels that explain the most variance. The objective it optimizes is:

$$\mathbf{v_i^*} = \underset{\mathbf{v_i}, \|\mathbf{v_i}\|=1}{\operatorname{argmax}} \mathbf{v_i}^\top \mathbf{X}^\top \mathbf{X} \mathbf{v_i} \qquad (1)$$

where $\mathbf{v_i^*}$ is the ith direction of maximum variance (called eigenvector) of data matrix $\mathbf{X}$. These vectors when plotted on the image highlight regions which are relevant for explaining variance in the data. Since, we have $\mathbf{p} \gg \mathbf{n}$, the above is an ill-posed problem and we need sparsity in eigenvectors, which gives rise to Sparse PCA [1]. The objective is exactly the same as Equation 1 except that now we have an additional sparsity penalty ($\|\mathbf{v_i}\|_1$) on eigenvectors.

Our proposed algorithm AC-PCA builds on Equation 1 and adds an additional term to the objective for incorporating prior information. Basically, it says that instead of finding the eigenvectors of the data covariance matrix (as a totally data driven approach would do), find the eigenvectors of the transformed data covariance matrix obtained by projecting it to the prior space. An important consequence of this is that we are not confining our data driven priors to lie in the original ROIs rather we are encouraging them to find ways to explain data variance but in this new "prior projected" space. Also, note that instead of taking the prior as is, we smooth it as $M_i^s = \exp(\frac{M_i - 1}{\lambda})$ before using it in our optimization. The modified objective is given below:

$$\mathbf{v_i^*} = \underset{\mathbf{v_i}, \|\mathbf{v_i}\|=1}{\operatorname{argmax}} \mathbf{v_i}^\top \mathbf{diag}(\mathbf{M_i^s}(\lambda)) \mathbf{C} \mathbf{v_i} - \eta \cdot \|\mathbf{v_i}\|_1 \qquad (2)$$

where $\mathbf{M_i^s}(\lambda)$ is the "smoothed" prior corresponding to the ith eigenvector and is itself a vector of size $(\mathbf{1} \times \mathbf{p})$ and $\mathbf{C}$ is the data covariance matrix ($\mathbf{X}^\top \mathbf{X}$). The **diag** operator converts the $\mathbf{1} \times \mathbf{p}$ vector into a $\mathbf{p} \times \mathbf{p}$ matrix and is used just for notational convenience. λ is a user tunable parameter which controls the tradeoff between the influence of data and prior and should be typically tuned on a held out validation set in the absence of other knowledge. Its easy to see that smaller values of λ suggest that we trust prior more and as λ is increased we want our eigenvectors $\mathbf{v}$ to be influenced more and more by the data. Also, it is important to note that we set the sparsity parameter η automatically as $\frac{Number\ of\ voxels\ in\ Prior\ ROI}{Total\ Number\ of\ Voxels\ in\ Brain}$.

Figure 2 uses example face data to show the effect of varying the λ parameter on the AC-PCA performance. As mentioned earlier, we choose λ by cross-validation on a validation set.

A. An Algorithm for AC-PCA

Since our data matrices are huge, typical number of voxels in image $\mathbf{p} \approx 100000$, simple methods which use routines from R, MATLAB don't scale that well. Instead, we propose a fast and novel iterative algorithm in C++ (publicly available) for solving the above optimization problem. The algorithm is presented below in Algorithm 1.
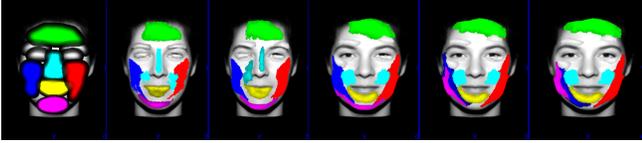
Figure 2. (AC-PCA) outputs for $\lambda = [0.0, 0.2, 0.5, 2.0, 5.0, 10.0]$ from left to right. Note that $\lambda = 0$ corresponds to using only prior ROIs and as λ is increased effect of data increases. This is evident from the fact that with increasing λ our priors especially around *nose, eyes and eyebrows* are getting washed away.

Algorithm 1 Anatomically Constrained Principal Component Analysis: AC-PCA

1: Normalize the data matrix $\mathbf{X}$ by mean centering it.
2: Select the prior scale parameter λ.
3: Uniformly initialize the k eigenvectors $\boldsymbol{v}_{1,\dots,k} \leftarrow 1/p$ and set $j = 0$.
4: **while** $\|\mathbf{v}_{1,\dots,k}^{j+1} - \mathbf{v}_{1,\dots,k}^{j}\|) < \epsilon$ **do**
5:    **for** i=1 to k **do**
6:       $\mathbf{v}_i^j \leftarrow \mathbf{X}\mathbf{v}_i^{j-1}$ //Implements modified Arnoldi Iteration algorithm
7:       $\mathbf{v}_i^j \leftarrow \mathbf{X}^\top \mathbf{v}_i^j$
8:       $\mathbf{v}_i^j \leftarrow \mathbf{v}_i^j \cdot \mathbf{M}_i^s$ // Update the eigenvector estimate with prior information.
9:       Orthogonalize the eigenvector with respect to all other eigenvectors:
$$\mathbf{v}_i^j \leftarrow \mathbf{v}_i^j - \sum_{l<i} \frac{\mathbf{v}_i^j \mathbf{v}_l^j}{\mathbf{v}_l^j \mathbf{v}_l^j} \mathbf{v}_l^j$$
10:       Soft-Max Sparseness: $\mathbf{v}_i^j \leftarrow (\|\mathbf{v}_i^j\| - max(\mathbf{v}_i^j) * \eta)_+ Sign(\mathbf{v}_i^j)$ //Note that η is set automatically as described earlier.
11:       Normalize the eigenvector $\mathbf{v}_i^j \leftarrow \mathbf{v}_i^j/\|\mathbf{v}_i^j\|$
12:    **end for**
13:    j $\leftarrow$ j+1
14: **end while**

III. Experimental Results

In this section we show the performance of our approach on cortical thickness images. Our data consists of images from 222 individuals with equal number of males and females of whom 122 were diagnosed clinically with Mild Cognitive Impairment (MCI) and the remaining 100 were normal controls. The average age of the cohort was 71.33. All images were acquired with a Siemens Trio 3.0 Tesla MRI scanner. The analysis of T1 images was done using publicly available Advanced Normalization Tools (ANTS, http://www.picsl.upenn.edu/ANTS/) and the associated pipelining framework PipeDream (http://sourceforge.net/projects/neuropipedream/) which mapped each subject to an existing, elderly/neurodegenerative population template, built from images acquired from the same scanner and imaging parameters.

We used two cortical label (probabilistic ROIs) definitions for our experiments 1). Non-Rigid Image Registration Evaluation Project (NIREP), (http://www.nirep.org/), 32 labels in total and 2). LONI Probabilistic Brain Atlas (LPBA40) [7], 55 labels in total.
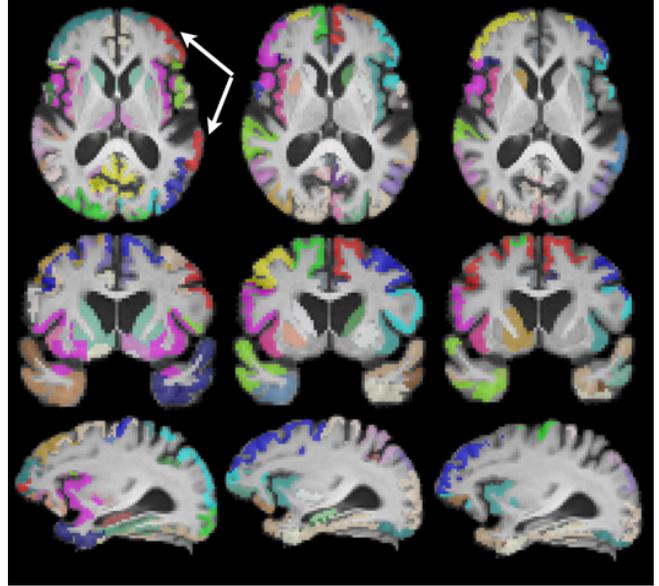

Figure 3. Results for LPBA40 labels. Left to right- Unconstrained PCA, prior LPBA labels, AC-PCA labels. The arrows show left frontal gyrus and the left superior temporal gyrus. The value of the prior strength λ for AC-PCA is 0.2, chosen on validation set. Note that not all the 55 anatomical priors can be seen in this slice.

We ran AC-PCA independently for each label set with varying values of prior strengths, $\lambda = [0.1, 0.2, 0.5, 1.0, 2.0, 5.0, 10.0, 20.0]$.

A. Qualitative Results

The figures below show slices of axial and sagittal section labelings for unconstrained PCA, true cortical labels (NIREP and LPBA40) and prior constrained PCA (AC-PCA).

Note that in the unconstrained PCA, the same label is assigned to parts of the left frontal gyrus and the left superior temporal gyrus (shown by arrows in figure) since it has no notion of anatomy; on the other hand, AC-PCA, since it is anatomical prior driven does not join distant structures which aids interpretability.

Another interesting observation is that both AC-PCA and unconstrained PCA have assigned the same label to corresponding regions in the left and right hemispheres. This occurs less frequently in AC-PCA as we have constrained it via anatomical priors.

B. Classification of MCI vs Controls

We hypothesize that AC-PCA will improve over basic ROIs on (MCI vs Controls) classification results by allowing the summary measurements derived from the imaging data to adapt to the underlying signal. We also hypothesize that AC-PCA locality will improve over unconstrained PCA-based classification.

We use the projections (eigenvectors projected onto the data matrix) resulting from unconstrained PCA, AC-PCA and prior cortical labels to distinguish individuals with Mild
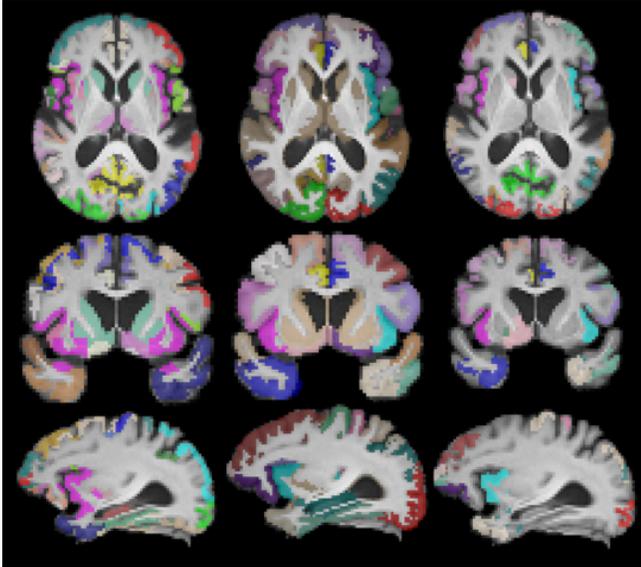
Figure 4. Results for NIREP labels. Left to right- Unconstrained PCA, prior NIREP labels, AC-PCA labels. The value of the prior strength λ for AC-PCA is 0.5, chosen on validation set. Note that not all the 32 anatomical priors can be seen in this slice.

Algorithm ($\mu \pm \sigma$)	NIREP	LPBA40
Unconstrained PCA	$62.58\% \pm 4.5\%$	$62.58\% \pm 4.5\%$
ROI Cortical Labels	$66.05 \pm 3.9\%$	$65.95\% \pm 3.6\%$
AC-PCA	$\mathbf{67.96\% \pm 2.3\%}$	$\mathbf{67.15\% \pm 3.2\%}$

Table I
TEST SET CLASSIFICATION ACCURACIES AVERAGED OVER 10000 RUNS
(MORE DETAILS IN TEXT). AC-PCA IS SIGNIFICANTLY
($p - val < 0.0001$ IN PAIRED T-TEST) BETTER THAN UNCONSTRAINED
PCA AND ROI CORTICAL LABELS

Cognitive Impairment (MCI) from normal aging individuals [8]. Since it is a classification problem, we used logistic regression and randomly split the data into training (80%), validation (10%) and testing (10%). Firstly, we use cross validation to determine the best value of prior strength parameter (λ); we train on the training data and test on validation data with varying prior strengths as mentioned above and then choose the best value of prior strength as the one which gave the least classification error on the validation set. The best prior strength value was $\lambda = 0.5$ for NIREP labels and $\lambda = 0.2$ for LPBA40 labels.

Next, with these values of λ we trained AC-PCA on the entire training and validation set and tested on the test set. This whole procedure was repeated 10000 times and the classification accuracies are given in Table 1.

As can be seen from the table, the AC-PCA significantly (paired t-test) outperforms unconstrained PCA as well as the approach which just uses cortical ROI labels hence leading to a classifier which can better distinguish MCI patients from normal controls.

IV. CONCLUSION

In this paper we proposed a novel approach for incorporating prior information in the form of anatomical priors into totally data driven approaches like PCA. We augmented the basic PCA objective to take into account prior anatomical information and proposed a fast modified Arnoldi iteration algorithm to solve the optimization problem. We also illustrated the benefits of our approach over standard unconstrained PCA as well as ROI based analysis by experiments on cortical thickness images which shows the superiority of AC-PCA for MCI classification compared to ROI and unconstrained PCA (a totally data based approach). Besides this, the labels output by AC-PCA readily lend themselves to interpretability by clinicians and other medical researchers as there is a one-one correspondence between the AC-PCA labels and prior anatomical labels, unlike totally data driven approaches like unconstrained PCA.

REFERENCES

[1] Daniela M Witten, Robert Tibshirani, and Trevor Hastie, "A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis.," *Biostatistics*, vol. 10, no. 3, pp. 515–534, Jul 2009.

[2] Brian B Avants, Philip A Cook, Lyle Ungar, James C Gee, and Murray 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.

[3] Russell A. Poldrack, "Region of interest analysis for fMRI.," *Social cognitive and affective neuroscience*, vol. 2, no. 1, pp. 67–70, Mar. 2007.

[4] Andrea Mechelli, Cathy J. Price, Karl J. Friston, and John Ashburner, "Voxel-based morphometry of the human brain: Methods and applications," *Current Medical Imaging Reviews*, vol. 1, pp. 105–113, 2005.

[5] Muhammad Tahir Akhtar and Christopher J. James, "Focal artifact removal from ongoing eeg–a hybrid approach based on spatially-constrained ica and wavelet de-noising.," *Conf Proc IEEE Eng Med Biol Soc*, vol. 2009, pp. 4027–4030, 2009.

[6] Ze Wang, "Fixed-point algorithms for constrained ica and their applications in fmri data analysis.," *Magn Reson Imaging*, vol. 29, no. 9, pp. 1288–1303, Nov 2011.

[7] David W. Shattuck, Mubeena Mirza, Vitria Adisetiyo, Cornelius Hojatkashani, Georges Salamon, Katherine L. Narr, Russell A. Poldrack, Robert M. Bilder, and Arthur W. Toga, "Construction of a 3D probabilistic atlas of human cortical structures.," *NeuroImage*, vol. 39, no. 3, pp. 1064–1080, Feb. 2008.

[8] Luping Zhou, Yaping Wang, Yang Li, Pew-Thian Yap, Dinggang Shen, and ADNI, "Hierarchical anatomical brain networks for mci prediction: Revisiting volumetric measures.," *PLoS One*, vol. 6, no. 7, pp. e21935, 2011.