

# Eigenanatomy Improves Detection Power for Longitudinal Cortical Change

Brian Avants, Paramveer Dhillon, Benjamin M. Kandel, Philip A. Cook, Corey T. McMillan<sup>1</sup>, Murray Grossman<sup>1</sup>, and James C. Gee

<sup>1</sup> Departments of Radiology and Neurology  
University of Pennsylvania, Philadelphia, PA 19104

**Abstract.** We contribute a novel and interpretable dimensionality reduction strategy, *eigenanatomy*, that is tuned for neuroimaging data. The method approximates the eigendecomposition of an image set with basis functions (the eigenanatomy vectors) that are *sparse*, *unsigned* and are *anatomically clustered*. We employ the eigenanatomy vectors as anatomical predictors to improve detection power in morphometry. Standard voxel-based morphometry (VBM) analyzes imaging data voxel-by-voxel—and follows this with cluster-based or voxel-wise multiple comparisons correction methods to determine significance. Eigenanatomy reverses the standard order of operations by first clustering the voxel data and then using standard linear regression in this reduced dimensionality space. As with traditional region-of-interest (ROI) analysis, this strategy can greatly improve detection power. Our results show that eigenanatomy provides a principled objective function that leads to localized, data-driven regions of interest. These regions improve our ability to quantify biologically plausible rates of cortical change in two distinct forms of neurodegeneration. We detail the algorithm and show experimental evidence of its efficacy.

## 1 Introduction

In machine learning, interpretable data decompositions are termed “parts-based representations” because they transform unstructured data into interpretable pieces [1–3]. Recent work in machine learning points to the fact that exploiting problem-specific information can improve parts-based representations [4–6]. Uninformed, generic matrix decomposition methods, e.g. standard principal component analysis (PCA), may be difficult to interpret because the solutions will produce vectors that are everywhere non-zero, i.e. involve the whole brain rather than its parts. Sparse methods have sought to resolve this issue [2, 3, 7–10]. However, these recent sparse multivariate methods are anatomically uninformed.

In this work, we employ a novel data-driven framework, related to the methods above, to delineate cortical networks wherein longitudinal atrophy patterns in Alzheimer’s disease (AD) and frontotemporal lobar degeneration (FTLD) differ from controls. Our novel image processing framework is open-source, unbiased with respect to registration and segmentation [11, 12] and is formulated spatiotemporally, as described in [13]. At the statistical level, our method is unbiased

in that it uses the intrinsic covariation of the dataset to parcellate the cortex into coherent regions and, in this reduced space, we gain sensitivity over full voxel-wise testing with voxel-based morphometry (VBM) [14]. Dimensionality reduction is critical for datasets that are relatively small and yet which quantify valuable and difficult to collect measurements in uncommon populations of subjects, as in FTLD. Critically, our novel dimensionality reduction method provides an objective function that optimally maps the classical singular value decomposition (SVD) eigenvectors that are both *signed* and *global* into *spatially localized, sparse, unsigned* pseudo-eigenvectors (or eigenanatomy). To our knowledge, the specific objective function used in eigenanatomy is the first to formulate an unsigned sparse decomposition with explicit guidance by the SVD solution and with anatomically informed regularization. We apply this framework in a cohort that is diagnosed by CSF biofluid biomarkers with high specificity and sensitivity [15]. This well-defined cohort provides an excellent test-bed for this new algorithm as we expect to identify specific patterns of cortical atrophy within regions known to be affected in FTLD and AD. At the same time, the relatively small cohort and challenges of longitudinal mapping may make this quantification difficult to achieve with VBM. We also show that eigenanatomy produces more powerful predictors when compared to related methods: VBM, classic SVD and penalized matrix decomposition (PMD) [16]. PMD is freely available and provides a sparse approach to PCA.

## 2 Methods

The class of methods encompassing non-negative matrix factorization (NMF) [1, 3, 17, 18], sparse principal components analysis (SPCA) [2, 19, 16, 20] and singular value decomposition [21] form the basis for the approach proposed here. More formally, define a  $n \times p$  (rows by columns) matrix  $\mathbf{X}$  where each row derives from an observed subject image such that the collection of images is given as vectors  $\{x_1, \dots, x_n\}$  with each vector  $x_i$  containing  $p$  entries. First, we denote each eigenanatomy component (a pseudo-eigenvector) as  $\mathbf{v}_i$  where  $i$  is ordered such that each eigenanatomy from  $\mathbf{v}_1$  to  $\mathbf{v}_m$  provides a decreasing contribution to the variance of matrix  $\mathbf{X}$ . We define eigenanatomy pseudo-eigenvectors as *sparse*. Sparseness means that some entries in  $\mathbf{v}_j$  will be zero.

The classic singular value decomposition (SVD) may be used to reduce the dimensionality of this data by decomposing the dataset into the eigenvectors of  $C_p = \mathbf{X}^T \mathbf{X}$  and  $C_n = \mathbf{X} \mathbf{X}^T$  (the right and left singular vectors, respectively). The relationship between a  $C_p$  and a  $C_n$  eigenvector is given by  $\mathbf{X} \mathbf{v}^p = \mathbf{v}^n$  and  $\mathbf{X}^T \mathbf{v}^n = \mathbf{v}^p$ . The  $n$  eigenvectors from  $C_p$  and  $C_n$  may be used to reconstruct the matrix  $\mathbf{X}$  by  $\sum_i \mathbf{v}_i^n \otimes \mathbf{v}_i^p \lambda_i$  where the  $\lambda_i$  denotes the  $i^{th}$  eigenvalue and  $\otimes$  the outer product. We now detail the algorithm that comprises the main contribution of this work and which is available as part of the *sccan* program within Advanced Normalization Tools (ANTs) [13].

Our goal is to approximate the matrix  $\mathbf{X}$  with its right and left singular vectors but where the right singular vector is *sparse*. We might, then, minimize:

$$\|\mathbf{X} - \sum_i \mathbf{v}_i^n \otimes \mathbf{v}_i^{sp} \lambda_i\|^2 \tag{1}$$

where the  $\mathbf{v}_i^{sp}$  denotes the  $i^{th}$  *sparse* right singular vector. It is known that the optimal value for  $\mathbf{v}_i^{sp}$  is exactly  $\mathbf{X}^T \mathbf{v}_i^n$  ( if we relax sparseness constraints ). Using the fact that  $\mathbf{X} \mathbf{v}^p = \mathbf{v}^n$  and  $\mathbf{X}^T \mathbf{v}^n = \mathbf{v}^p$ , we therefore reformulate the objective in a slightly simpler form and seek to directly find  $\mathbf{v}_i^{sp}$  by optimizing:

$$\arg \min_{\mathbf{v}_i^{sp}} \|\mathbf{X} \mathbf{v}_i^{sp} - \mathbf{v}_i^n\|^2. \tag{2}$$

This optimization problem is quadratic without sparseness constraints and easily solved by conjugate gradient through the normal equations  $\|\mathbf{X}^T \mathbf{X} \mathbf{v}_i^{sp} - \mathbf{X}^T \mathbf{v}_i^n\|^2$ .

Now, note that the vector  $\mathbf{X}^T \mathbf{v}_i^n = \mathbf{v}_i^p$  might have both positive and negative values. As with non-negative matrix factorization, we seek a decomposition that is unsigned. Thus, an optimal solution that minimizes  $\|\mathbf{X} \mathbf{v}_i^{sp} - \mathbf{v}_i^n\|^2$  will need to model both signs. We therefore make a second adjustment by modeling the positive and negative components of  $\mathbf{v}_i^p$  separately.

Each eigenvector may be written in an expanded form via the use of indicator functions which are diagonal matrices with binary entries. For instance, if  $\mathbf{v}$  contains entries  $[-2, -1, 0, 1, 2]$ , then the positive indicator function is  $I^+ = [0, 0, 0, 1, 1]$  and the negative indicator function is  $I^- = [1, 1, 0, 0, 0]$ .  $\mathbf{v}$  may then be expressed as  $\mathbf{v} = I^+ \mathbf{v} + I^- \mathbf{v} = \mathbf{v}^+ + \mathbf{v}^-$ . We use these indicator functions to separate the positive and negative components of our objective such that the optimization in equation 2 becomes,

$$\arg \min_{\mathbf{v}_i^{sp+}, \mathbf{v}_i^{sp-}} \|\mathbf{X}^T \mathbf{X} \mathbf{v}_i^{sp+} - \mathbf{v}_i^{p+}\|^2 + \|\mathbf{X}^T \mathbf{X} \mathbf{v}_i^{sp-} - \mathbf{v}_i^{p-}\|^2 \tag{3}$$

This minimization problem forms the basis for our novel approach to computing eigenanatomy, i.e. anatomically localized approximations to the eigenvectors of an anatomical imaging dataset. Derivation of sparse eigenanatomy is shown in Figure 1.

As noted above, we seek sparse and interpretable solutions. We define a sparse vector as one which minimizes a  $l_0$  or  $l_1$  penalty term i.e. has either a user-specified number of non-zero entries ( $l_0$ ) or absolute sum ( $l_1$ ). Although the  $l_1$  penalty has advantages [2], we use the  $l_0$  penalty because it specifies the fraction of the vector that is allowed to be non-zero. The sparseness restriction is therefore easily interpreted by users of the eigenanatomy method. Eigenanatomy seeks to identify sparse functions  $\mathbf{v}_i^{sp+}$  and  $\mathbf{v}_i^{sp-}$  that closely approximate the eigenvectors in  $n$ -space, i.e.  $\mathbf{v}_n^i = \mathbf{X} \mathbf{v}_i^p$ . The objective function, again employing the normal equations, is then:

$$\arg \min_{\mathbf{v}_i} \sum_{i=1}^n \|C_p \mathbf{v}_i^{sp+} - \mathbf{v}_i^{p+}\|^2 + \|C_p \mathbf{v}_i^{sp-} - \mathbf{v}_i^{p-}\|^2 \tag{4}$$

subject to:  $\|\mathbf{v}_i^{sp+}\|_0 = \|\mathbf{v}_i^{sp-}\|_0 = \gamma,$

**Algorithm 1.** SNLCG optimization for eigenanatomy.

---

```

Input  $\mathbf{X}$ , the eigenvectors of  $\mathbf{X}\mathbf{X}^T$  and  $\gamma$ , the sparseness parameter.
for all  $\mathbf{v}_i \in \{\mathbf{v}_1, \dots, \mathbf{v}_{n-1}\}$  do
   $\mathbf{v}_i \leftarrow \mathbf{X}^T \mathbf{v}_i^n$                                 ▷ Get the  $p$ -space eigenvector from the  $\mathbf{v}_i^n$ .
  Compute  $\mathbf{v}_i^+$  and  $\mathbf{v}_i^-$ .                                ▷ Find the + and - representation of  $\mathbf{v}_i$ .
   $\mathbf{v}_i^{s+} \leftarrow \mathbf{v}_i^{s-} \leftarrow \frac{1}{p}$                 ▷ Initialize the sparse + and - vectors.
   $\mathbf{v}_i^{s+} \leftarrow \text{SNLCG}(\mathbf{X}, \mathbf{v}_i^{s+}, \mathbf{v}_i^+, \gamma)$         ▷ SNLCG + minimization.
   $\mathbf{v}_i^{s-} \leftarrow \text{SNLCG}(\mathbf{X}, \mathbf{v}_i^{s-}, \mathbf{v}_i^-, \gamma)$         ▷ SNLCG - minimization.
   $\mathbf{v}_i^{s-} \leftarrow \mathbf{v}_i^{s-} * (-1)$                         ▷ Reset  $\mathbf{v}_i^{s-}$  to be positive.
end for

```

---

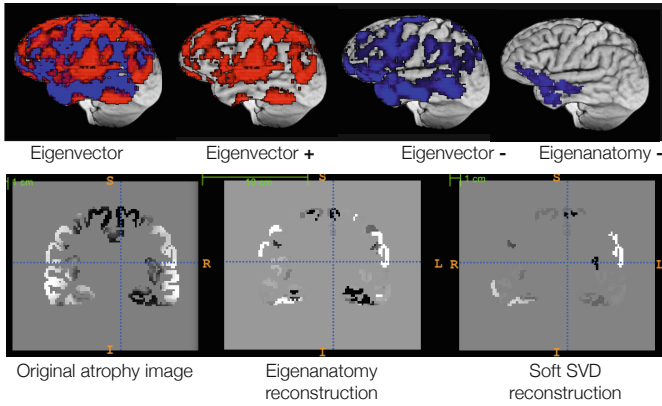
where  $\gamma$  defines the desired level of sparseness for each eigenanatomy vector. Eigenanatomy therefore produces  $2*n$  sparse pseudo-eigenvectors whose product with  $\mathbf{X}$  may be used as predictors in standard linear regression. Importantly, because these vectors are unsigned, they may be interpreted as *weighted averages of the input data*.

Sparseness can be enforced by a soft-thresholding algorithm as in [2, 16]. We denote this function as  $S(\mathbf{v}, \gamma)$  and (in an *ad hoc* manner) allow it to also reject isolated voxels of the eigenanatomy vector that are non-contiguous (i.e. we provide a cluster threshold as in VBM). Minimization problems involving the  $l_0$  penalty are  $np$ -hard. The relaxed form of this objective function (i.e. without the sparseness constraint) is purely quadratic and can easily be solved by a conjugate gradient method. Thus, we propose a new sparse, nonlinear conjugate gradient (SNLCG) method as a minimization procedure for the eigenanatomy objective function to deal with the nonlinearities induced by the  $S$  function and  $l_0$  constraint. The additional advantage of SNLCG is that its solutions approach the quadratic minimum as sparseness constraints are relaxed. We detail the minimization algorithms for the eigenanatomy objective function (equation 4) in algorithms 1 and 2. The algorithms are also available in an open-source **R** package for free use and within the ANTs toolkit.

### 3 Results

For all uses of eigenanatomy below, we set the sparseness parameter,  $\gamma$ , to select 5% of the voxels in the cortex. We choose 5% because this provides interpretable clusters of regions in the cortex and yet still allows a reasonable reconstruction of the original data matrix, as shown in Figure 1.

**Reconstruction Error:** We quantify the ability of the eigenanatomy algorithm to reconstruct the original dataset from sparse eigenvectors using equation 1. As a baseline, we compare the full eigenanatomy solution to the reconstruction given by applying the soft-threshold function  $S$  directly to the SVD-derived



**Fig. 1. The eigenanatomy basis functions:** The original eigenvector (far left) has both positive and negative components. These are separated into the positive and negative vector components (middle figures). The sparse eigenanatomy approximation to  $\mathbf{v}^-$  is shown at far right. Because the entries of  $\mathbf{v}^{s-}$  are either zero or negative, the sign of  $\mathbf{v}^{s-}$  can be changed to positive. Thus,  $\mathbf{v}^{s-}$  is an interpretable measurement of the data and provides a weighted average of the original signal. Ultimately, the weighted average of the imaging data provided by  $\mathbf{v}^{s-}$  is used as a predictor in regression. The same is done with  $\mathbf{v}^{s+}$ . In the lower portion of the figure, we see reconstruction results from the eigenanatomy method—see the results section for more explanation.

vectors  $\mathbf{v}_i^p$  without further optimization. We call this “soft-SVD”. We also compare reconstruction error to the eigenanatomy algorithm run with a restriction on the number of iterations in the SNLCG sub-algorithm. These experiments show that the full eigenanatomy algorithm run until convergence (error = 1.251) improves upon both the soft-SVD solution (error = 1.292) and the limited iteration eigenanatomy solution (error = 1.279). Error is measured by the frobenius norm taken between the original matrix and the reconstructed matrix.

**Neuroimaging Data:** Our cohort consists of 61 participants, including 15 patients with AD (7 females), 23 patients with FTLN (14 females), and 23 controls (13 females). All patients were clinically diagnosed by a board-certified neurologist and cerebrospinal fluid confirmation of the underlying pathology was obtained [15]. No significant difference exist between disease duration, age or education in the patient or control groups. Two 3.0T MPRAGE T1-weighted magnetic resonance images were obtained for each subject. The FTLN group (time interval 1.12 years  $\pm$  0.28) trended ( $p < 0.081$ ) towards having a reduced interval when compared to elderly controls (time interval 1.29  $\pm$  0.36 years), as did the AD group (time interval 1.13  $\pm$  0.25 years,  $p < 0.12$ ). The interval between scans was therefore factored out as a nuisance variable.

**Detection Power in Comparison to VBM, SVD and PMD:** In this section, we employ eigenanatomy to compare the ability to detect group differences in cortical atrophy rate between FTLN subjects and controls as well as AD

**Algorithm 2.** SNLCG sub-algorithm for eigenanatomy.

---

```

Input  $\mathbf{X}, \mathbf{v}_i^-, \mathbf{v}_i^{s-}, \gamma$ .
 $\mathbf{b} \leftarrow \mathbf{v}_i^-$ 
 $\mathbf{x}_k \leftarrow \mathbf{v}_i^{s-}$ 
 $\mathbf{r}_k \leftarrow (\mathbf{b} - \mathbf{X}^T(\mathbf{X} \mathbf{x}_k))$   $\triangleright$  Use the gradient of the quadratic term.
 $\mathbf{p}_k \leftarrow \mathbf{r}_k$ 
 $\Delta E \leftarrow \infty$ 
while  $\Delta E > 0$  do
   $\alpha_k \leftarrow \langle \mathbf{p}_k, \mathbf{X}^T \mathbf{X} \mathbf{x}_k \rangle$   $\triangleright \langle \cdot, \cdot \rangle$  denotes inner product.
   $\mathbf{x}_{k+1} \leftarrow \mathbf{x}_k + \alpha_k \mathbf{p}_k$ 
   $\mathbf{x}_{k+1} \leftarrow S(\mathbf{x}_{k+1}, \gamma)$ 
   $\triangleright$  Project  $\mathbf{x}_{k+1}$  to the sparse solution space given by the objective function.
   $\mathbf{r}_{k+1} \leftarrow S(\mathbf{b} - \mathbf{X}^T(\mathbf{X} \mathbf{x}_{k+1}), \gamma)$   $\triangleright$  Use the gradient of the quadratic term.
   $\triangleright$  Project  $\mathbf{r}_{k+1}$  to the sparse solution space given by the objective function.
   $\beta_k \leftarrow \|\mathbf{r}_{k+1}\|^2 / \|\mathbf{r}_k\|^2$   $\triangleright$  Standard conjugate gradient definitions below.
   $p_{k+1} = r_{k+1} + \beta_k * p_k$ 
   $\Delta E \leftarrow \|\mathbf{r}_k\| - \|\mathbf{r}_{k+1}\|$ 
   $\mathbf{r}_k \leftarrow \mathbf{r}_{k+1}; \mathbf{x}_k \leftarrow \mathbf{x}_{k+1}; \mathbf{p}_k \leftarrow \mathbf{p}_{k+1}$ 
end while

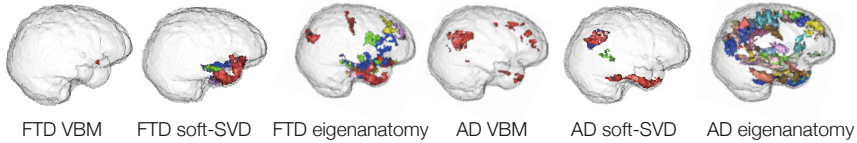
```

---

subjects and controls. This analysis shows specificity of the approach and biological plausibility in two different neurodegenerative disorders.

We passed the same input dataset to all methods. The data consisted of unbiased voxel-wise measures of annualized atrophy rate in the cortex of all patients normalized to a group template, as described elsewhere [13]. The regression model employed for all methods is summarized ( in  $\mathbf{R}$  syntax) as: *atrophy-rate*  $\approx 1 + \text{diagnosis} + \text{education} + \text{interval-between-images} + \text{disease-duration} + \text{gender}$ . “Diagnosis” is the predictor of interest i.e. we test whether the presence of disease predicts atrophy rate given the presence of the covariates. The *atrophy-rate* is either a vector of voxel-wise measures or a basis function projection against the original atrophy rate image matrix  $\mathbf{X}$  i.e.  $\mathbf{X}\mathbf{v}_i^p$ . The latter case is used for classic SVD, PMD, soft-SVD in addition to eigenanatomy. We define significance as a  $q$ -value  $< 0.05$  where a  $q$ -value is a false discovery rate corrected  $p$ -value. For PMD and SVD, we tested  $n$  different atrophy rates ( one for each eigenvector ) where  $n$  is number of subjects. For eigenanatomy and soft-SVD, we used  $2n$  predictors as suggested by design in the algorithm. For VBM, we tested all 50,194 voxels (the number of columns in  $\mathbf{X}$ ).

For the FTLD subjects and the AD subjects, when classic SVD projections were used as measures of atrophy rate, *no* significant predictors emerged. The same is true for the PMD method—although we note the caveat that, potentially, a more exhaustive parameter search may have resulted in better PMD results. Both univariate VBM and soft-SVD identify significant effects although the minimum  $q$ -value and extensiveness of both soft-SVD and VBM are far less than eigenanatomy. This is particularly true for the FTLD subjects for which VBM only produces a small 20 voxel cluster that survives correction. The results



**Fig. 2. Statistical comparison:** Eigenanatomy detects the most effects and also with smallest p-values (FTLD:  $q < 0.0015$ , AD:  $q < 0.002$ ) versus the next best soft-SVD (FTLD:  $q < 0.0035$ , AD:  $q < 0.009$ ). Different colors represent different eigenvectors / predictors. Note that multiple disjoint, but related, voxel clusters (i.e. a network) may be involved in an eigenanatomy vector. VBM detects 20 voxels in FTLD ( $q < 0.015$ ). In AD, univariate results are more robust, likely due to the widespread nature of AD atrophy, and 1382 significant voxels ( $q < 0.0270$ ) were detected.

are summarized further in Figure 2. Detailed clinical interpretation is beyond the scope of the paper but the results, overall, are coincident with what is known about these disorders. In particular, the largest atrophy rate in FTLD was in right orbitofrontal cortex. For AD, this region was in the precuneus.

## 4 Conclusion

We detailed the eigenanatomy theory and algorithm and showed that eigenanatomy improves image reconstruction from a sparse set of anatomical basis functions. We showed that eigenanatomy also improves detection power for detecting group differences in longitudinal cortical change relative to SVD, PMD and univariate VBM. We note that showing that a method increases detection power does not comprise sufficient validation. This approach is not limited to longitudinal analysis and may be applied in a variety of morphometry contexts. Future extensions to the basic eigenanatomy algorithm given here will include network interpretations, exploration of alternatives to the SNLCG algorithm, alternative penalty terms and an automation of parameter selection.

## References

1. Lee, D.D., Seung, H.S.: Learning the parts of objects by non-negative matrix factorization. *Nature* 401(6755), 788–791 (1999)
2. Zou, H., Hastie, T., Tibshirani, R.: Sparse principal component analysis. *Journal of Computational and Graphical Statistics* 15(2), 262–286 (2006)
3. Hoyer, P.O., Dayan, P.: Non-negative matrix factorization with sparseness constraints. *Journal of Machine Learning Research* 5, 1457–1469 (2004)
4. Guan, N., Tao, D., Luo, Z., Yuan, B.: Manifold regularized discriminative non-negative matrix factorization with fast gradient descent. *IEEE Trans. Image Process.* 20(7), 2030–2048 (2011)
5. Cai, D., He, X., Han, J., Huang, T.S.: Graph regularized non-negative matrix factorization for data representation. *IEEE Trans. Pattern Anal. Mach. Intell.* (December 2010)

6. Hosoda, K., Watanabe, M., Wersing, H., Krner, E., Tsujino, H., Tamura, H., Fujita, I.: A model for learning topographically organized parts-based representations of objects in visual cortex: topographic nonnegative matrix factorization. *Neural Comput.* 21(9), 2605–2633 (2009)
7. Witten, D.M., Tibshirani, R.: A framework for feature selection in clustering. *J. Am. Stat. Assoc.* 105(490), 713–726 (2010)
8. Friedman, J., Hastie, T., Tibshirani, R.: Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* 33(1), 1–22 (2010)
9. Cherkassky, V., Ma, Y.: Another look at statistical learning theory and regularization. *Neural Netw.* 22(7), 958–969 (2009)
10. Friedman, J., Hastie, T., Tibshirani, R.: Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* 9(3), 432–441 (2008)
11. Yushkevich, P.A., Avants, B.B., Das, S.R., Pluta, J., Altinay, M., Craige, C., A.D.N.I.: Bias in estimation of hippocampal atrophy using deformation-based morphometry arises from asymmetric global normalization: an illustration in ADNI 3 t MRI data. *Neuroimage* 50(2), 434–445 (2010)
12. Holland, D., Dale, A.M., A.D.N.I.: Nonlinear registration of longitudinal images and measurement of change in regions of interest. *Med. Image Anal.* 15(4), 489–497 (2011)
13. Avants, B.B., Yushkevich, P., Pluta, J., Minkoff, D., Korczykowski, M., Detre, J., Gee, J.C.: The optimal template effect in hippocampus studies of diseased populations. *Neuroimage* 49(3), 2457–2466 (2010)
14. Ashburner, J., Friston, K.J.: Unified segmentation. *Neuroimage* 26(3), 839–851 (2005)
15. Irwin, D.J., McMillan, C.T., Toledo, J.B., Arnold, S.A., Shaw, L.M., Wang, L.S., Trojanowski, J.Q., Lee, V.M.Y., Grossman, M.: Comparison of Cerebrospinal Fluid Levels of Tau and ABeta1-42 in Alzheimer’s Disease and Frontotemporal Degeneration Using Two Analytical Platforms. *Archives of neurology* (2011) (submitted)
16. Witten, D.M., Tibshirani, R., Hastie, T.: A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics* 10(3), 515–534 (2009)
17. Kim, H., Park, H.: Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares for microarray data analysis. *Bioinformatics* 23, 1495–1502 (2007)
18. Heiler, M., Schnörr, C.: Learning sparse representations by non-negative matrix factorization and sequential cone programming. *Journal of Machine Learning Research* 7, 1385–1407 (2006)
19. d’Aspremont, A., El Ghaoui, L., Jordan, M.I., Lanckriet, G.R.G.: A direct formulation for sparse pca using semidefinite programming. *SIAM Rev.* 49, 434–448 (2007)
20. Lee, K., Tak, S., Ye, J.C.: A data-driven sparse glm for fMRI analysis using sparse dictionary learning with mdl criterion. *IEEE Trans. Med. Imaging* 30(5), 1076–1089 (2011)
21. Sill, M., Kaiser, S., Benner, A., Kopp-Schneider, A.: Robust biclustering by sparse singular value decomposition incorporating stability selection. *Bioinformatics* 27(15), 2089–2097 (2011)