Sparse Methods for Dimensionality Reduction in Medical Imaging

Ben Kandel

Specific Aims

In this proposal, we aim to advance sparse dimensionality reduction techniques in medical imaging to provide more accurate and robust correlations between imaging and cognitive and other clinical data. This work falls under the following three aims.

Specific Aim Ia: Linear Sparse Medical Image Decomposition for Predicting Clinical Data

A variety of sparse matrix decomposition methods have been proposed that aim to reduce the dimensionality of a matrix while retaining non-zero entries on only a fraction of the basis vector entries. Recently, supervised sparse decomposition methods tuned for images and other areas in which both spatial information and image labels are important have been developed. We aim to extend this work to decompositions in which some clinical or other data that consists of a continuous variable is associated with each picture. We also propose modifications of the smoothness constraints that are commonly used in matrix decompositions to both better preserve the reconstruction of the original data and also encourage anatomically interpretable basis vectors.

Specific Aim Ib: Grouped Sparse Medical Image Decomposition

In addition, we will examine the problem of incorporating prior anatomical knowledge into the decomposition, which is not typically a feature of sparse data decomposition. We propose to use a group data decomposition algorithm. This group algorithm will enforce sparsity only on labels (i.e., only a few pre-defined regions of the brain will be allowed to be included in the decomposition), but will not enforce sparsity *within* a given region, so that within a region, all voxels of the region may vary freely. This will encourage sparsity in an anatomically-informed way, so that anatomically meaningful regions as a whole will be allowed to stay in the model. This can be seen as an ROI-based version of a sparse decomposition.

Specific Aim II: Non-Linear Sparse Dimensionality Reduction

We propose to extend nonlinear dimensionality reduction techniques to the sparse setting. Non-linear dimensionality reduction techniques are useful for analysis of data that does not lie on a linear subspace of the original, high-dimensional data. Such nonlinear dimensionality reduction techniques have proven very useful in analysis of image data. In this work, we propose a sparse extension of one of the most popular non-linear dimensionality reduction techniques, Locally Linear Embedding (LLE). We also demonstrate how the ideas presented here can be extended to the supervised decomposition problem.

Specific Aim III: Application to Neurodegenerative Diseases

Although the dimensionality reduction techniques outlined in the previous section are of general interest and can be applied to many types of image data, we are specifically interested in medical applications of the data. Traditionally, medical image analysis is performed on a voxel-wise, mass univariate basis. This technique, however, is

not ideally suited for prediction of disease or clinical data because of the very high dimensionality of the analyzed data. We propose to use the dimensionality reduction procedures outlined above to correlate specific areas with decline in performance on cognitive tests in the setting of Alzheimer's Disease.

Background, Significance, and Innovation

Medical imaging forms an increasingly indispensable role in medical diagnosis and disease characterization because it offers the ability to gain detailed information about the structure and function of patients' bodies. At the same time, the richness and subtleties in medical images present significant methodological challenges for traditional statistical methodologies. Because of the extremely high dimensionality and spatial information contained within medical images, traditional regression and classification techniques are not ideally suited for using medical images to make predictions as to a patient's disease state or clinical data. Voxel-based morphometry (VBM) [2], one of the most widely used methods for analyzing neuroimaging data, looks for voxel-wise differences between groups of images. Because the significance maps that are the output of VBM are still very high-dimensional, the output from VBM is not ideally suited to predictions of disease state or clinical data, and some other dimensionality reduction technique is still required [4]. In addition, because VBM operates in a mass-univariate manner, it can lose the ability to detect more subtle group differences that multivariate statistical methods are better suited to recover [9].

Dimensionality reduction techniques are widely used to convert high-dimensional data into a low-dimensional space that in some way approximates the original high-dimensional data but that is more amenable to classification or other analytical tasks. The methods are widely applied to a variety of data types, including facial images [32, 5], genetic data [14, 38, 1, 42, 33], and medical images [28]. Classic dimensionality reduction techniques, such as Principal Components Analysis (PCA), however, have the drawback that the low-dimensional space is made up of contributions from every component in the high-dimensional space. In the context of genetic analysis, this means that each principal component or eigenvector has non-zero entries that correspond to every gene in a gene array. When using this low-dimensional space for looking for correlations between a given eigenvector and a clinical outcome, non-zero contributions from every gene in the gene array make it difficult to make biologically meaningful conclusions with regard to what any individual gene does. In medical image analysis, a similar problem arises: Non-zero entries in every component in the principal components of a vector make it difficult to interpret which anatomical areas are most important for predicting a given clinical outcome.

To alleviate this problem, one would try to find approximations to standard PCA decompositions that have a relatively small number of non-zero components. Constraints on the number of non-zero components of a solution are referred to as *sparsity* constraints, and the solutions of problems with sparsity constraints are referred to as sparse solutions. Sparsity has been employed as a constraint in matrix decompositions in a wide variety of fields [18, 8, 42, 27, 41, 33, 42, 17, 16, 15, 4, 3]. Because of the close relationship between PCA and multivariate regression, techniques originally developed for regularizing linear regression, such as the Lasso and its variants [30, 31] can be profitably adapted to constraining PCA and other dimensionality reduction techniques, such as canonical correlation analysis (CCA) [34].

Within this methodological context, we seek to tune existing strategies for dimensionality reduction techniques to the application of medical image analysis. Although medical image matrix decompositions that take into account image labels have been proposed [4], none have been proposed that incorporate clinical data, such as a cognitive test, that consist of continuous variables. Incorporating such continuous data into the decomposition will enable more robust detection of structural changes that are correlated with cognitive data. Because cognitive tests are often designed to focus on one specific aspect of cognition, incorporating these data into the decomposition has the potential to shed light on which anatomical areas are correlated with which cognitive abilities in the context of neurodegenerative diseases. At the same time, reducing the dimensionality of the input data will help ameliorate multiple comparisons issues and make linear regression models more powerful.

A second aspect in which we will advance current methods in matrix decomposition involves the penalty terms used to ensure that the basis vectors obtained from the decomposition are anatomically meaningful. Most current matrix decomposition techniques in the field of medical imaging use a second-order smoothing term. In many cases, however, first-order smoothing terms, such as those present in total variation denoising, have been found to outperform second-order smoothing terms when the data is known to contain large jumps [23]. Because of the sparsity constraints on our basis vectors, we in fact do expect large jumps in the basis vectors, so first-order, or ℓ_1, penalties on the vector derivatives are more appropriate. We demonstrate how penalties of this sort have already been proposed for regularizing linear regression and show how they can be incorporated into a matrix decomposition framework.

Existing sparse matrix decomposition methods for medical imaging also do not admit the possibility of incorporating prior knowledge, such as labels or region of interests (ROI's), into the decomposition. We show how prior knowledge of anatomically meaningful areas can be incorporated into the decomposition to determine which anatomical regions show the most correlation with cognitive data.

A second major class of dimensionality reduction techniques, in addition to the linear techniques discussed above, is nonlinear dimensionality reduction techniques. Many forms of high-dimensional data, such as images, are not necessarily adequately explained by linear statistical techniques and are instead more aptly described as living on a non-linear manifold. A wide variety of techniques have been proposed to deal with dimensionality reduction in this context, including kernelized versions of PCA [24] and manifold-based techniques [29, 22, 6, 19]. Manifold-based techniques have been used effectively in analysis of neuroimaging data [35, 36], but to date no sparse formulations have been proposed. We propose a sparse extension to Locally Linear Embedding (LLE), one of the most widely used non-linear manifold-based dimensionality reduction techniques, and show how similar extensions may be made to other non-linear dimensionality reduction techniques. Because manifold-based techniques do not involve mapping the input data to a higher-dimensional space, as in kernel methods, it is more straightforward to include spatially-informed priors on the basis vectors.

Following the theoretical formulation of the dimensionality reduction techniques, we aim to implement the techniques and use them to analyze existing neuroimaging databases for connections between neuroanatomical variations and cognitive deficits. Although a large body of research has focused on using anatomical or functional imaging for predicting the presence or absence of Alzheimer's Disease based on MRI scans, relatively few studies have used sparse multivariate techniques similar to those proposed here to predict cognitive scores from neuroimaging scans. Our methods have the potential to find areas of the brain that are maximally effective for predicting cognitive scores, and extensive testing, such as that proposed here, will both offer the opportunity to find novel neurobiological results and also demonstrate the power of the proposed methods to other researchers interested in using these methods.

Notation

To avoid confusion, we will describe the notation we use. Matrices are denoted by bold capitalized letters ($\mathbf{X}$). If indexed, matrices are still bolded, so that the i'th matrix is $\mathbf{X}_i$. Vectors are capitalized, but not bolded (V). The i-th column vector of a matrix is denoted by an capitalized letter with a subscript (X_i), and the entry corresponding to row i and column j is x_{ij}. Scalars are lower-case (x), and sets in script ($\mathcal{N}$). In keeping with standard usage, Greek letters used to denote vectors are not capitalized, but the intent should be clear from the context. Estimated values are given hats ($\hat{v}$). The ℓ_p norms $\|x\|_p$ are defined as $(\sum_{i=1}^{n} |x_i|^p)^{1/p}$, with the ℓ_0 pseudo-norm returning the number of non-zero entries in a vector.

Specific Aim I: Regression-Based and Grouped Sparse Linear Medical Image Decomposition

Regression-Based Image Decomposition

In order to fully appreciate how clinical data can be used to guide medical image decomposition, we will begin by briefly reviewing methods for sparse PCA decompositions. Once we have shown how unsupervised sparse PCA decompositions are formulated, we will extend the decomposition to situations in which we have outcome data of a continuous variable.

We represent a set of n mean-centered medical images, each containing p voxels, as an $n \times p$ matrix $\mathbf{X} \in \mathbb{R}^{n \times p}$. We seek a decomposition of this matrix into a loading matrix $\mathbf{W} \in \mathbb{R}^{n \times r}$ and a basis matrix $\mathbf{V} \in \mathbb{R}^{r \times p}$ so that

$$\mathbf{X} \approx \mathbf{WV}. \tag{1}$$

Two common ways of formulating the PCA objective are finding the projection matrix$\mathbf{V}$ that maximizes the variance of the data matrix,

$$\hat{\mathbf{V}} = \arg\max_{\mathbf{V}} \|\mathbf{XV}\|_2^2, \tag{2}$$

and finding the projection matrix that minimizes the reconstruction error,

$$\hat{\mathbf{V}} = \arg\min_{\mathbf{V}} \|\mathbf{X} - \mathbf{VV}^T\mathbf{X}\|_2^2, \tag{3}$$

subject to the orthogonality constraint $\mathbf{V}^T\mathbf{V} = I$ [12]. The loading matrix is given by the projections of the data matrix onto the principal components

$$\mathbf{W} = \mathbf{XV}^T, \tag{4}$$

with matrix dimensions as before. An alternative derivation of the principal components is by minimizing the reconstruction error :

$$\hat{\mathbf{V}} = \arg\min_{V} \|\mathbf{X} - \mathbf{VV}^T\mathbf{X}\|, \tag{5}$$

To enforce sparsity on this decomposition, Zou [42] introduces the ℓ_1 penalty into Equation 3. Introducing a penalty on the ℓ^1 norm of a solution is a standard convex relaxation to a penalty on the ℓ^0 "norm," because optimizing over an ℓ^0 penalty is an NP-hard problem. (Donoho–citation for l1 here.) To do this, he uses two matrices, α and β, for use in the reconstruction, and then imposes orthogonality on α and sparsity on β:

$$(\hat{\alpha}, \hat{\beta}) = \arg\min_{\alpha,\,\beta} \sum_{i=i}^{n} \|\mathbf{X}_i - \alpha\beta^T\mathbf{X}_i\|^2 + \lambda \sum_{j=1}^{k} |\beta_j|, \tag{6}$$

subject to $\alpha^T\alpha = I$. The $\hat{\beta}$ obtained from this equation will then satisfy $\hat{\beta} \propto \mathbf{V}_i$. In addition to the sparsity constraint, some sort of clustering or smoothing penalty is required to avoid returning scattered non-zero components, as in [4]. This gives us an objective function of the form

$$(\hat{\alpha}, \hat{\beta}) = \arg\min_{\alpha,\,\beta} \sum_{i=i}^{n} \|\mathbf{X}_i - \alpha\beta^T\mathbf{X}_i\|^2 + \lambda_1 \sum_{j=1}^{k} |\beta_j| + \lambda_2 p(\beta), \tag{7}$$

where $p(x)$ is some penalty that discourages spatially incoherent structures and and λ_1 and λ_2 are parameters.

The simplest type of constraint that discourages scattered non-zero pixels is some sort of smoothing constraint. This can take the form of a linear difference operator, such as the fused Lasso penalty [31], which has the form $\lambda_2 \sum_{j \in \mathcal{N}_i} |x_i - x_j|$, when $\mathcal{N}_i$ is the neighborhood of x_i. Other options include a quadratic smoothing

term, as proposed in [40] and adopted in [4], or an ℓ^2 penalty on the gradient of the vector, as in [13]. Although these options do decrease the likelihood of finding isolated non-zero voxels, they may also force smoothing of the decomposition to an undesirable degree. I propose to investigate other types of penalties that will encourage clustering. The first penalty I will examine is one of the form $\lambda_2 \sum_{j \in \mathcal{N}_i} ||x_i|^\alpha - |x_j|^\alpha|$, where $0 < \alpha \leq 1$. Although this penalty is non-convex, it comes closer to the desired property of penalizing scattered non-zero voxels, while not enforcing smoothness on the non-zero entries. Another possibility is to include a regularization based on the curvature of the non-zero components, as is often done in level set-based methods [26]. Optimization of such a constraint would also not be possible using standard convex optimization techniques, would likely require a coordinate descent-type approach. Experimental studies will be necessary to determine whether the improvement in results from these alternative methods is significant as compared to existing smoothness constraints.

A second extension I propose is to incorporate regression of the projections onto the available clinical data in the objective function. This will give me a penalty term of the form $\ell(y, \mathbf{X}V^T)$, where y is the available clinical data. The clinical data can take the form of cognitive scores, disease state (in logistic regression), or other outcomes one is trying to predict. This can be considered a generalization of the classification loss term considered by [4]. Denoting the regression coefficients γ, so that the standard regression equation becomes $Y = X\gamma$, we can then finally write

$$(\hat{\alpha}, \hat{\beta}, \hat{\gamma}) = \arg\min_{\alpha, \beta} \sum_{i=i}^{n} \|\mathbf{X}_i - \alpha\beta^T\mathbf{X}_i\|^2 + \lambda_1 \sum_{j=1}^{k} |\beta_j| + \lambda_2 p(\beta) + \lambda_3 \|Y - \mathbf{X}V^T\gamma\|^2. \tag{8}$$

Incorporating a loss term based on the prediction of the clinical data will encourage the decomposition to be correlated with the clinical outcomes, and changes the unsupervised decomposition to a supervised decomposition reminiscent of partial least squares regression.

Group Image Decomposition: NEEDS RECONSTRUCTION ERROR TERM!!!

One drawback to the proposed image decomposition methods is that they do not offer a possibility of incorporating prior anatomical knowledge into the decomposition. Because there is no prior anatomical knowledge, there is no guarantee that the decompositions will correspond to anatomically meaningful regions. Although the penalties proposed do encourage the decomposition to correspond to regions that *can* be anatomically meaningful (coherent, smooth, etc.), the projections may still spread over distinct anatomical regions. I propose to introduce a modified version of the decomposition that will incorporate prior anatomical knowledge in the form of a labeled atlas. At the same time, we are still interested in decompositions that do not spread over the entire brain. We are interested in having a limited number of predefined regions in the brain, but within those regions do not have further sparsity constraints. In other words, we are interested in region-wise sparsity, but not voxel-wise sparsity. Denoting each brain region as $\mathbf{V}_j$, with all the brain region vectors incorporated into a matrix $\mathbf{V}$ and the projection of the images onto all the brain regions as $\mathbf{X}\mathbf{V}^T$, we then have an objective of the form

$$\|y - \mathbf{X}\mathbf{V}^T\gamma\| + \lambda\|\gamma\|^2 + p(k), \tag{9}$$

when k is the number of regions incorporated into the model and $p(k)$ is a penalty on the number of regions incorporated into the model. The ℓ^2 regularization on γ is the standard ridge (or Tikhonov) regularization necessary to avoid degenerate solutions. This objective turns in to a modified model selection problem, and a wide variety of methods are available to find (local) minima.

Perhaps the simplest method for solving this problem involves a greedy selection algorithm. A greedy algorithm would first choose the region V_i, corresponding to one column in $\mathbf{V}$, with the highest correlation with the outcome variable:

$$\left(\hat{V}_i, \gamma_i\right) = \arg\min_{V_i, \gamma_i} \|y - \mathbf{X}\mathrm{diag}(V_i)\gamma_i\|^2, \tag{10}$$

where $\mathrm{diag}(x) : x \in \mathbb{R}^m \rightarrow X \in \mathbb{R}^{m \times m}$ takes a vector $x \in \mathbb{R}^m$ and creates an $m \times m$ matrix with diagonal entries equal to entries in x. Next, it would find the region with the highest correlation with the residual (i.e., the response not accounted for by the region already included in the model):

$$\left(\hat{V}_{i+1}, \gamma_{i+1}\right) = \underset{V_{i+1}, \gamma_{i+1}}{\arg\min} \|y - \mathbf{X}\mathrm{diag}(V_i)\gamma_i - \mathbf{X}\mathrm{diag}(V_{i+1})\gamma_{i+1}\|^2 \tag{11}$$

We would then test if incorporating the extra region significantly improves the model, as determined by whether or not the BIC score [25] decreases with the added region.

More sophisticated group model selection algorithms have also been proposed. Grouped versions of the Lasso and Least Angle Regression (LARS) [10] have been proposed to deal with this type of problem [39]. The group version of LARS performs the following steps: First, it picks the group of variables that is most correlated with the outcome variable. It then moves the solution in this direction until the projection of the residual of the outcome (i.e., the outcome not explained by the current model) onto another group has as small an angle with the outcome as the original direction has with the outcome. Subsequent groups are added in a similar way until some stopping criterion is met. The stopping criterion can be some threshold on prediction error or some other metric, such as the BIC, that takes into account model size.

Now that we have laid out some of the specific mathematical forms that we will be optimizing, we will reiterate the significance of the proposed methods for medical image analysis. The fundamental issue that confronts analysis of medical image data is the extremely high dimensionality of the input data. Because of this high dimensionality, some dimensionality reduction method is necessary to make meaningful correlations with clinical or other data. First, we proposed an unsupervised dimensionality reduction technique that takes into account anatomically meaningful constraints to ensure that the output vectors can be meaningfully interpreted. Next, we extended this model to encourage the decomposition to be correlated with available clinical data. Finally, we outlined a method for incorporating prior anatomical knowledge into the decomposition. This method has the practical effect of enabling a researcher to determine which anatomical regions, out of a set of predefined regions, are most important for predicting some clinical outcome. In all these methods, sparsity is incorporated as a constraint to make sure that the model corresponds to only certain portions of the anatomy, and does not spread over the entire image. Until now, though, we have assumed that the clinical data can be modeled as a multivariate linear function of the input images. In the next section, we will see how it may be possible to relax that assumption and still obtain sparse decompositions.

Specific Aim II: Non-Linear Sparse Dimensionality Reduction

Although the multivariate linear dimensionality reduction techniques proposed above will enable recovery of linear correlations between imaging data and clinical or other data, they will not be able to perform non-linear dimensionality reduction (Figure 1). As an example technique, we will propose a sparse extension to Locally Linear Embedding (LLE), one of the most widely used non-linear dimensionality reduction techniques. Building on recent extensions of LLE to the supervised learning context, we will propose both an unsupervised and a supervised dimensionality reduction technique. The practical performance of the various manifold and other non-linear dimensionality reduction techniques is heavily application-dependent, but the widespread use and straightforward implementation of LLE makes it a reasonable place to start for proposing a sparse extension.

LLE works by computing the following three steps:

1. For each data point X_i, find a neighborhood $\mathcal{N}_i$ of other data points that are most similar to it. Assignments can be made by k-nearest neighbors or by incorporating all neighbors such that the distance between X_i and its neighbors is less than some ϵ.

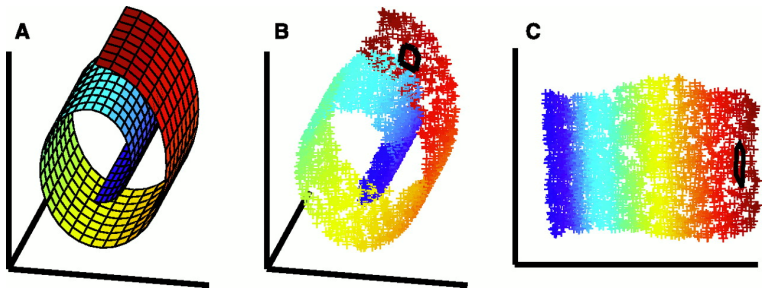

Figure 1: Demonstration of "Swiss roll" data that demonstrates the utility of non-linear dimensionality reduction techniques. The data cannot be satisfactorily explained by a linear projection, but the non-linear manifold dimensionality reduction technique Locally Linear Embedding (LLE) successfully "unwraps" the curve and presents a reasonable two-dimensional decomposition of the data. Figure taken from [22].

2. Determine optimal reconstruction weights to reconstruct each data point X_i from its neighbors by minimizing the reconstruction error
$$\varepsilon(W) = \arg\min_W \sum_i \|X_i - \sum_{j \in \mathcal{N}_i} W_{ij} X_j\|^2. \tag{12}$$

3. With weights fixed, find the low-dimensional embedding that most closely approximates the high-dimensional weight structure: $\Phi(Y) = \sum_i \|Y_i - \sum_{j \in \mathcal{N}_i} W_{ij} Y_j\|^2$.

To see how we can extend this method to incorporate sparsity, we will briefly review linear sparse PCA. The reconstruction errorformulation of PCA seeks to find a vector which will minimize the reconstruction error of the matrix $\mathbf{X}$:
$$\hat{v} = \arg\min_v \|\mathbf{X} - vv^T\mathbf{X}\| \tag{13}$$
This equation is similar to the reconstruction weight equation (Equation 12), except that Equation 12 only utilizes the data points in the neighborhood of the target image. From a sparse version of LLE, we would hope to recover both the reconstruction weights (so that we can compute a low-dimensional embedding) and the vector that specifies a section of the original matrix that is most critical for the LLE reconstruction. In the simplest case, this would correspond to a Boolean vector v whose entries belong to the set $\{0, 1\}$, which would give us a reconstruction error of the form
$$\varepsilon_{sparse}(W, v) = \arg\min_{W,v} \sum_i \|X_i - \sum_{j \in \mathcal{N}_i} W_{ij} (v \odot X_j)\|^2, \tag{14}$$
where $\odot$ stands for element-wise multiplication. However, optimizing over Boolean vectors is generally very difficult, so we will relax this constraint and reformulate the problem in terms of a change of basis that is more similar to the sparse PCA formulation discussed above.

Modifying Equation 12 in this way gives us
$$\varepsilon_{sparse}(W, v) = \arg\min_{W,v} \sum_i \|X_i - \sum_{j \in \mathcal{N}_i} W_{ij} vv^T X_j\|^2. \tag{15}$$
When v is the identity matrix, Equation 15 is clearly equivalent to Equation 12, but we are interested in incorporating sparsity in the form of an ℓ^1 penalty on the solution:
$$\varepsilon_{sparse}(W, v) = \arg\min_{W,v} \sum_i \|X_i - \sum_{j \in \mathcal{N}_i} W_{ij} vv^T X_j\|^2 + \|v\|_1, \tag{16}$$

where as before, v is orthogonal. Note that the v matrix is not indexed. This objective will return a PC-like sparse matrix that is best suited to reconstructing each individual data point from a weighted combination of its neighbors. The sparse matrix will then indicate which parts of the image, or in our case, which anatomical component, is most important for reconstruction of the original image. The connections between this decomposition and the linear sparse decomposition methods described above are clear; the difference is that in this case, we only consider the reconstruction of each data point based on the data points (or images) that are most similar to it.

To extend this decomposition to the supervised domain, we must take into account some clinical data by incorporating the available labels into the decomposition. A variety of supervised versions of manifold-based non-linear dimensionality reduction techniques have been proposed for greater discrimination power between members of different groups [21, 37, 11, 20] . The basic idea is to define the distance (in step 1) between data points from within the same class differently from distances between data points in different classes. By incorporating class information into the decomposition, a greater degree of discrimination is possible between data points from different classes. In its simplest form, it consists of simply adding a constant to distances between data points between different classes [21], but more sophisticated versions are possible that give greater control over the distance between images from different classes. Incorporating class labels may improve the ability of the low-dimensional embedding to preserve distinctions between images of different classes, which will improve the ability to classify unseen images.

Once a low-dimensional embedding is computed, it is possible to use the low-dimensional embedding to predict clinical data in either a regression-type or classification-type framework. We leave the possibility of incorporating continuous data into the decomposition to future research.

Specific Aim III: Validation and Clinical Application

Although the theoretical formulations above are necessary to be able to perform the decompositions, the ultimate utility of the proposed decomposition methods depends on their performance on actual data sets. We will begin the practical evaluation of the methods proposed by implementing them in Matlab and evaluating their performance on synthetic datasets. In the first stage, we will generate synthetic datasets by forming images with simple geometric shapes, such as circles. We will then introduce a perturbation, such as an area of increased brightness, in some of the images. Synthetic "clinical" data may consist of a value that scales linearly with the intensity of perturbation in the test images. The sparse PCA algorithm should be able to retrieve the area of perturbation, and this will serve as the first stage of validation that the algorithm has been correctly implemented. After this initial test of the linear dimensionality reduction techniques, we can extend the validation by changing the shape of the perturbation and adding scattered points of perturbation to test the effect of the various weighting parameters that encourage the PCA algorithm to find smooth eigenvectors.

Following validation of the linear dimensionality reduction algorithms, we will implement the non-linear sparse dimensionality reduction algorithm. As a set of test images for algorithm, we propose to use existing image datasets. The image sets generated for testing Isomap [29] are freely available `http://isomap.stanford.edu/datasets.html` and will be a good place to perform preliminary testing.

After the initial validation and debugging of the implementations are complete, we aim to move towards analysis of neuroimaging data. To work with neuroimaging data, it is usually necessary to use a lower-level language than Matlab to avoid limits on memory usage and to improve speed. We aim to implement the algorithms in the `sccan` program, which is part of the ANTs software developed by PICSL. Matrix manipulations are performed using the VNL matrix operation library, which is integrated into ITK for processing medical images.

For test data, we propose to analyze two independent sets of Alzheimer's Disease patients. First, we will use the publicly available Alzheimer's Disease Neuroimaging Initiative, which has a large dataset of both imaging and clinical data. We will also perform the analysis on a large dataset of AD and control patients available from the University of Pennsylvania Memory Center, which has a more ethnically and demographically varied

subject group than the ADNI dataset. Performing the analysis on two independently collected sets of data will augment our confidence that any neurobiological results we find constitute meaningful and robust results. Using the modified linear regression techniques proposed above will enable probing the connection between individual brain structures and clinical scores. We will examine the connection between cortical thickness and scores on the Boston Naming Test (BNT), a test of language and recall function; the Trail-Making test, a test of cognitive processing speed (part A) and executive function (part B) ability [7]; and the Mini-Mental State Exam (MMSE). Both the dataset from the Penn Memory Center and the ADNI data have these tests available.

Literature Cited

[1] O. Alter, P.O. Brown, and D. Botstein. Singular value decomposition for genome-Wide expression data processing and modeling. *Proceedings of the National Academy of Sciences of the United States of America*, 97(18):10101–10106, 2000.

[2] J Ashburner and K J Friston. Voxel-based morphometry–the methods. *NeuroImage*, 11(6 Pt 1):805–821, June 2000. PMID: 10860804.

[3] N. Batmanghelich, A. Dong, B. Taskar, and C. Davatzikos. Regularized tensor factorization for multi-modality medical image classification. *Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention*, 14(Pt 3):17–24, 2011.

[4] N. Batmanghelich, B. Taskar, and C. Davatzikos. A general and unifying framework for feature construction, in image-based pattern classification. *Information processing in medical imaging : proceedings of the ... conference*, 21:423–434, 2009.

[5] P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711–720, 1997.

[6] M. Belkin and P. Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15(6):1373–1396, 2003.

[7] Christopher R Bowie and Philip D Harvey. Administration and interpretation of the trail making test. *Nature Protocols*, 1(5):2277–2281, 2006.

[8] E.J. Cands, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2):489–509, 2006.

[9] Christos Davatzikos. Why voxel-based morphometric analysis should be used with great caution when characterizing group differences. *NeuroImage*, 23(1):17–20, September 2004.

[10] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani. Least angle regression. *The Annals of statistics*, 32(2):407—499, 2004.

[11] X. Geng, D.-C. Zhan, and Z.-H. Zhou. Supervised nonlinear dimensionality reduction for visualization and classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 35(6):1098–1107, 2005.

[12] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition*. Springer, 2nd ed. 2009. corr. 3rd printing 5th printing. edition, February 2009.

[13] M. Hebiri and S. van de Geer. The smooth-lasso and other 1 + 2-penalized methods. *Electronic Journal of Statistics*, 5:1184–1226, 2011.

[14] N.S. Holter, M. Mitra, A. Maritan, M. Cieplak, J.R. Banavar, and N.V. Fedoroff. Fundamental patterns underlying gene expression profiles: Simplicity from complexity. *Proceedings of the National Academy of Sciences of the United States of America*, 97(15):8409–8414, 2000.

[15] P. O Hoyer. Non-negative matrix factorization with sparseness constraints. *The Journal of Machine Learning Research*, 5:1457–1469, 2004.

[16] I.T. Jolliffe, N.T. Trendafilov, and M. Uddin. A modified principal component technique based on the LASSO. *Journal of Computational and Graphical Statistics*, 12(3):531547, 2003.

[17] I.T. Jolliffe and M. Uddin. The simplified component technique: An alternative to rotated principal components. *Journal of Computational and Graphical Statistics*, 9(4):689–710, 2000.

[18] S. Levy and P.K. Fullagar. Reconstruction of a sparse spike train from a portion of its spectrum and application to high-resolution deconvolution. *Geophysics*, 46(9):1235–1243, 1981.

[19] L.J.P. van der Maaten, E. O. Postma, and H. J. van den Herik. *Dimensionality Reduction: A Comparative Review*. 2008.

[20] B. Raducanu and F. Dornaika. A supervised non-linear dimensionality reduction approach for manifold learning. *Pattern Recognition*, 45(6):2432–2444, 2012.

[21] Dick Ridder, Olga Kouropteva, Oleg Okun, Matti Pietikinen, and Robert P. W. Duin. Supervised locally linear embedding. In Okyay Kaynak, Ethem Alpaydin, Erkki Oja, and Lei Xu, editors, *Artificial Neural Networks and Neural Information Processing ICANN/ICONIP 2003*, volume 2714, pages 333–341. Springer Berlin Heidelberg, Berlin, Heidelberg.

[22] Sam T Roweis and Lawrence K Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, December 2000.

[23] L. I Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1-4):259268, 1992.

[24] Bernhard Schlkopf, Alexander Smola, and Klaus-Robert Mller. Kernel principal component analysis. In Wulfram Gerstner, Alain Germond, Martin Hasler, and Jean-Daniel Nicoud, editors, *Artificial Neural Networks ICANN'97*, volume 1327 of *Lecture Notes in Computer Science*, pages 583–588. Springer Berlin / Heidelberg, 1997.

[25] G. Schwarz. Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464, 1978.

[26] J A Sethian. A fast marching level set method for monotonically advancing fronts. *Proceedings of the National Academy of Sciences*, 93(4):1591–1595, February 1996.

[27] K. Sjostrand, E. Rostrup, C. Ryberg, R. Larsen, C. Studholme, H. Baezner, J. Ferro, F. Fazekas, L. Pantoni, D. Inzitari, and G. Waldemar. Sparse decomposition and modeling of anatomical shape variation. *Medical Imaging, IEEE Transactions on*, 26(12):1625 –1635, December 2007.

[28] S.J. Teipel, C. Born, M. Ewers, A.L.W. Bokde, M.F. Reiser, H.-J. Mller, and H. Hampel. Multivariate deformation-based analysis of brain atrophy to predict alzheimer's disease in mild cognitive impairment. *NeuroImage*, 38(1):13–24, 2007.

[29] Joshua B Tenenbaum, Vin De Silva, and John C Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, December 2000.

[30] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–288, January 1996. ArticleType: research-article / Full publication date: 1996 / Copyright 1996 Royal Statistical Society.

[31] Robert Tibshirani, Michael Saunders, Saharon Rosset, Ji Zhu, and Keith Knight. Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(1):91108, 2005.

[32] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.

[33] Daniela M Witten, Robert Tibshirani, and Trevor Hastie. A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics (Oxford, England)*, 10(3):515–534, July 2009. PMID: 19377034.

[34] D.M. Witten and R.J. Tibshirani. Extensions of sparse canonical correlation analysis with applications to genomic data. *Statistical Applications in Genetics and Molecular Biology*, 8(1), 2009.

[35] R. Wolz, P. Aljabar, J.V. Hajnal, A. Hammers, and D. Rueckert. LEAP: learning embeddings for atlas propagation. *NeuroImage*, 49(2):1316–1325, 2010.

[36] R. Wolz, P. Aljabar, J.V. Hajnal, J. Ltjnen, and D. Rueckert. Nonlinear dimensionality reduction combining MR imaging with non-imaging information. *Medical Image Analysis*, 16(4):819–830, 2012.

[37] M.-H. Yang. Extended isomap for classification. In *Proceedings - International Conference on Pattern Recognition*, volume 16, pages 615–618, 2002.

[38] K.Y. Yeung and W.L. Ruzzo. Principal component analysis for clustering gene expression data. *Bioinformatics*, 17(9):763–774, 2001.

[39] M. Yuan and Y. Lin. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 68(1):49–67, 2006.

[40] Rafal Zdunek and Andrzej Cichocki. Blind image separation using nonnegative matrix factorization with gibbs smoothing. In Masumi Ishikawa, Kenji Doya, Hiroyuki Miyamoto, and Takeshi Yamakawa, editors, *Neural Information Processing*, volume 4985 of *Lecture Notes in Computer Science*, pages 519–528. Springer Berlin / Heidelberg, 2008.

[41] S. Zhang, Y. Zhan, M. Dewan, J. Huang, D.N. Metaxas, and X.S. Zhou. Deformable segmentation via sparse shape representation. *Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention*, 14(Pt 2):451–458, 2011.

[42] H. Zou, T. Hastie, and R. Tibshirani. Sparse principal component analysis. *Journal of computational and graphical statistics*, 15(2):265—286, 2006.