Theory and implementation of Laplacian eigenmaps for probabilistic tractography data

Rajat M. Thomas, Leonardo Cerliani

February 22, 2011

We illustrate the rationale of Laplacian eigenmaps using a classical example from (Tenenbaum et al., 2000) using the "Faces" dataset, available at:
`http://waldron.stanford.edu/~isomap/face_data.mat.Z`

There are N pictures of a face taken from a constant distance, but with varying viewing angles. If the resolution of the images is 64 $\times$ 64 pixels, they can be conceived to represent a dataset of N observations in 4096 dimensions. Intuitively these images should lie on a hypersurface with a much lower dimensionality and in this case a three dimensional structure corresponding to the degrees of freedom of the camera moving in a 3D space at a fixed distance from the face. The task of recovering this 3D structure from the original 4096-dimensional space is an example of a dimensionality reduction problem. Specifically, we ought to find three principal axes that would completely represent the original observations in such a way that the proximity of points in this 3D space reflects the similarity in the viewing angles. In principle, no prior knowledge of the degrees of freedom across the given dataset, which for us is just a matrix of N rows by 4096 columns, need to be assumed. It is the task of the algorithm to reveal, if present, a 3D structure of similarity/variability across all images. Laplacian eigenmaps (Belkin and Niyogi, 2002) implement one such algorithm for nonlinear dimensionality reduction. Other algorithms to perform nonlinear dimensionality reduction are Isomaps (Tenenbaum et al., 2000) and Locally Linear Embedding (Roweis and Saul, 2000). Figure 1 is an example of how the method of Laplacian eigenmaps recover the main trajectories of variability (and therefore of similarity) in the Faces dataset (originally consisting of 698 images in 64 by 64 pixels, varying the pose of the face and the lighting

angle).

In the above example, similarities across different images are used to uncover a global structure underlying all images, and as expected, discover that the main difference across all images is the camera's angle at a fixed distance from the face. Analogoues to the matrix $N \times 4096$ above, in probabilistic tractography data, we input the connectivity maps for each seed voxel, seeking for an anatomically defined trajectory of variation underlying all connectivity maps.

Animal studies on the anatomical connectivity of the insular cortex suggests a main direction of variabilty in the connections along a certain topographically localized axis in the anatomy, in particular along the rostro-caudal axis. If the same holds true in humans, this is the underlying structure which should be recovered from the very high dimensional data derived from probabilistic tractography. In other words, when considering the similarity/dissimilarity of the spatial location of the (ipsilateral) brain targets reached during tractography across insular seed voxels we expect the presence of a trajectory of connection variability which spans the rostro-caudal axis in the anatomy.

Before we dwell on the details and implementation of the algorithm it would be illustrative to reiterate the geometry of the problem in hand. The connectivity matrix has dimensions $N_I \times N_B$, where N_I is the number of voxels considered in the insula and N_B the number of voxels in the ipsilateral hemisphere. We consider each row, with dimension $1 \times N_B$, as a vector in vector-space, $\mathbb{R}^{N_B}$. Thus, voxels in the insula occupy N_I points in this N_B-dimensional space. Given our anatomical hypothesis, the N_I points (each corresponding to one seed voxel's connectivity map) do not occupy random positions in $\mathbb{R}^{N_B}$ but in turn lie on a structure (a manifold, in mathematical terms) whose dimension is much lower than N_B. As an extreme example, let us consider all voxels in the insula connected to a single voxel of the brain along approximately the same connectivity profile. In this case the dimension of the manifold embedded in $\mathbb{R}^{N_B}$ is zero, because all N_I vectors coincide.

The problem described above is a classical example to which techniques like principal component analysis (PCA) and multidimensional scaling (MDS) are applied. We in turn chose to use Laplacian Eigenmaps (Belkin and Niyogi, 2002, LEM). LEMs are relatively insensitive to noise and outliers in the data set and

unlike PCA or MDS, they preserve local neighbourhood information and don't require that the low-dimensional structure of the dataset lies on or near a linear subspace of the high-dimensional input space (Tenenbaum et al., 2000; Belkin and Niyogi, 2002). The technique is similar to the spectral reordering previously used in other connectivity-based parcellation studies (Johansen-Berg et al., 2004; Klein et al., 2007), but additionaly takes advantage of the information about the nearest neighbours of each connectivity feature vector. The final representation resembles those from multidimensional scaling (see for instance (Passingham et al., 2002)), however Laplacian eigenmaps don't need to assume that the low-dimensional structure of the dataset lies on a linear subspace (Belkin and Niyogi, 2002; Tenenbaum et al., 2000)).

Laplacian eigenmaps are a local information preserving non-linear dimensionality reduction technique where the basis vectors in $\mathbb{R}^{N_B}$ are transformed to a new set of coordinates $q_0, q_1 \ldots q_{N_B-1}$, such that the direction of maximum information/variation is q_0 and the variation decreases with increasing i in q_i. The number of these basis vectors q_i we select are therefore a measure of how well we retain the information on the correlation and coherence (neighbourhood information) of the voxels in the insula. Thus, if neigbouring voxels are highly correlated, only a small number of basis vectors need to be retained to recover most of the information.

It is important to note that in our implementation of the LEM, we slighly modify the geometric picture. Instead of the connectivity matrix as the starting point in the algorithm, we first calculate the cross correlation matrix of the N_I vectors, giving a matrix of size $N_I \times N_I$, which is provided as input to the algorithm. The reason for inputing the cross-correlation matrix is two fold; one it reduces the number dimensions of the space from N_B to N_I and two, the errors of estimating the connectivity path from probabilistic tractography are different for different seed voxels and the process of cross-correlating would mitigate the effects of this noise.

Our implementation of the LEM technique is detailed below.

Let $\mathbf{x_0}, \mathbf{x_1}, \ldots, \mathbf{x_{N_I-1}}$ be the N_I points represented by the rows of the correlation matrix of the connectivity maps in $\mathbb{R}^{N_I}$. The algorithm consists of formulating and solving an eigenvalue problem. The eigenvectors and their associated eigenvalues then determine the optimal basis to represent the data.

1. Calculate the **Adjacency matrix**. The adjacency matrix of a simple graph

is a matrix with rows and columns labeled by graph vertices, with a 1 or 0 at positions (i,j) according to whether $\mathbf{x_i}$ and $\mathbf{x_j}$ are adjacent or not. Two nodes are said to be adjacent or connected if $\|\mathbf{x_i} - \mathbf{x_j}\|^{\mathbf{2}} < \epsilon$, where $\epsilon \in \mathbb{R}$, is usually an arbritary parameter chosen judiciously. In our implementation, we choose ϵ as the minimum distance which yielded a connected graph (see Belkin and Niyogi, 2002).

2. Set up a **Weight matrix** as;

$$W_{ij} = \begin{cases} 1 & \text{if connected} \\ 0 & \text{if unconnected}, \end{cases}$$

where W_{ij} is the weight on the vertex connecting nodes i and j.

3. The **Eigenvalue problem** to be solved is,

$$L\mathbf{f} = \lambda D\mathbf{f}. \tag{1}$$

Where,

$$D_{ii} = \sum_j W_{ij}, \tag{2}$$

and

$$L = D - W, \tag{3}$$

is the Laplacian matrix.

Let the eigenvalues of eq(1) be $\lambda_0, \lambda_1, \ldots, \lambda_{N_I-1}$ and their corresponding eigenvectors $\mathbf{f_0}, \mathbf{f_1}, \ldots, \mathbf{f_{N_I-1}}$, such that,

$$\lambda_0 \leq \lambda_1 \leq \ldots \leq \lambda_{N_I-1}. \tag{4}$$

4. Choose the first $\mathbf{m}$ non-zero eigenvalues and associated eigenvectors. The number of components $\mathbf{m}$ is chosen depending on how well these points in the $\mathbf{m}(< N_i)$-dimensional representation of the data set reflect the positional configuration in $\mathbb{R}^{N_I}$-dimensions. To quantify positional similarity on an $\mathbf{m}$-dimensional manifold, and an $\mathbf{N_I}$-dimensional euclidean space, we need to construct a distance matrix for the $\mathbf{m}$-dimensional manifold. The non-euclidean nature of the manifold poses a problem of setting up a metric

in this space.

In our study we hypothesized that the insular connectivity patterns would show a variation consistent with the evidence from tracer-injection studies in macaques, that is, a one-dimensional variation in the insular cortex along the rostro-caudal anatomical axis. A visual inspection of the Laplacian eigenmaps using the first two smallest nonzero eigenvectors (that is, a two-dimensional representation on a plane) revealed the presence of a linear or curvilinear main trajectory of variation across all connectivity maps. Laplacian eigenmaps produced using 3 and 4 components did not alter the topology of the structure, therefore we choose to use the two main components for further processing.

5. We remove points on the plane lying very distant from the low-dimensional structure, which could represent potential outliers, or points at the transition between relatively segregated substructures. Specifically, this is performed by calculating the distribution of the euclidean distances between each point and each other in the Laplacian eigenmaps, and discarding all the points whose median distance from their respective 50 nearest neighbours lies above the 95% of the distribution (almost identical results were obtained for other choices of the number of nearest neighbours in the range 10..100). This 'cleaning' of the Laplacian eigenmap is performed in order to increase the precision if a subsequent curve fitting is required, in case Laplacian eigenmaps recovers a curvilinear low-dimensional structure of variability, and to evidence the presence of potential clusters, which could be masked by few points lying at the transition between clusters.

6. Finally, we choose as a metric of the recovered one-dimensional structure of variability the distance along the structure from one of its extremes. To maintain consistency across subjects, the extreme on the structure corresponding to the most anterior voxel in the brain was always used as a reference. In case Laplacian eigenmaps produced a linear structure, we simply rotate it to align it to the x axis and calculate the distance of each point on the structure from the point with the minimum abscissa. In case Laplacian eigenmaps produced a curvilinear structure (like a horseshoe), we fit the points with a 4th

degree polynomial curve. A goodness-of-fit threshold was set to $R^2 \geq 0.95$. The measured distances were normalized in the range 0..1 and overlayed as a color-coded map onto the anatomical image.

References

Belkin, M. and Niyogi, P. (2002). Laplacian eigenmaps and spectral techniques for embedding and clustering.

Johansen-Berg, H., Behrens, T. E., Robson, M. D., Drobnjak, I., Rushworth, M. F., Brady, J. M., Smith, S. M., Higham, D. J., and Matthews, P. M. (2004). Changes in connectivity profiles define functionally distinct regions in human medial frontal cortex. *Proc. Natl. Acad. Sci. U.S.A.*, 101:13335–13340.

Klein, J. C., Behrens, T. E., Robson, M. D., Mackay, C. E., Higham, D. J., and Johansen-Berg, H. (2007). Connectivity-based parcellation of human cortex using diffusion MRI: Establishing reproducibility, validity and observer independence in BA 44/45 and SMA/pre-SMA. *Neuroimage*, 34:204–211.

Passingham, R. E., Stephan, K. E., and Kotter, R. (2002). The anatomical basis of functional localization in the cortex. *Nat. Rev. Neurosci.*, 3:606–616.

Roweis, S. T. and Saul, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–6.

Tenenbaum, J. B., de Silva, V., and Langford, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–23.

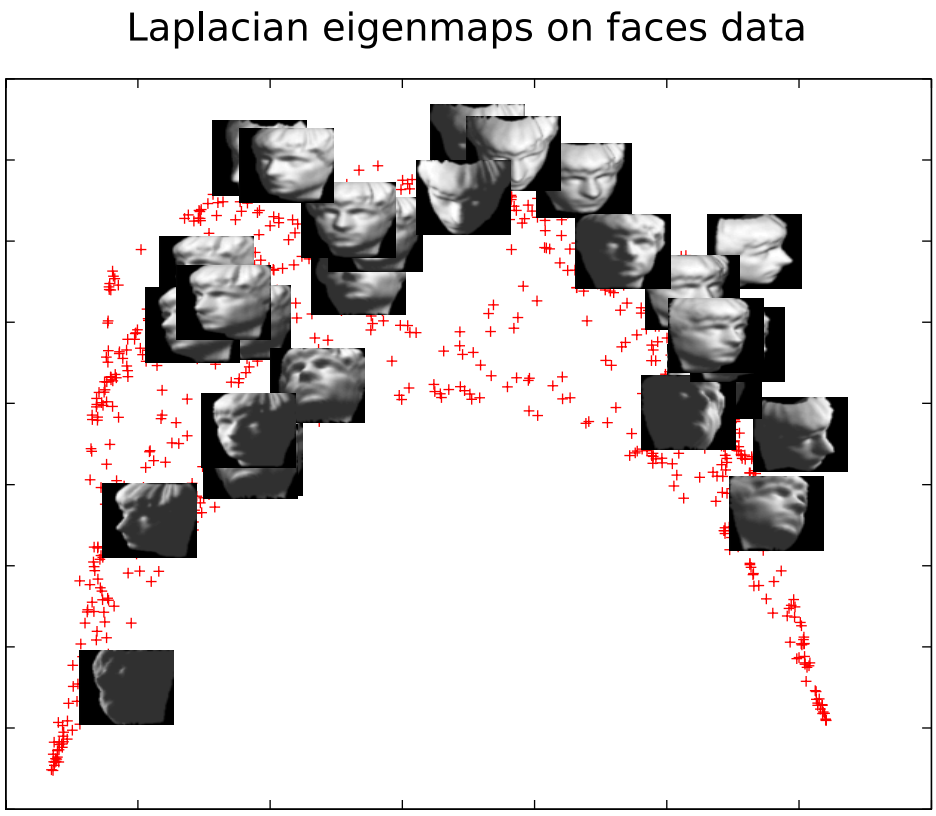

Figure 1: Laplacian Eigenmaps of the faces dataset using a k = 7 nearest neighbours. The two smallest nonzero eigenvectors have been used to remap the original 4096-dimensional dataset on a plane. A sample of the original images are also presented in the location of their corresponding datapoint, to show the trajectory of variability recovered by the algorithm. This image was generated with the code available on the web page of Ali Ghodsi at http://www.math.uwaterloo.ca/˜aghodsib/courses/f06stat890/Asg/