NONLINEAR DIMENSION REDUCTION OF FMRI DATA: THE LAPLACIAN EMBEDDING APPROACH

Bertrand Thirion and Olivier Faugeras

INRIA Sophia Antipolis
Odyssee Laboratory
2004, route des lucioles, 06902 Sophia-Antipolis, France

ABSTRACT

In this paper, we introduce the use of nonlinear dimension reduction for the analysis of functional neuroimaging datasets. Using a Laplacian Embedding approach, we show the power of this method to detect significant structures within the noisy and complex dynamics of fMRI datasets; it outperforms classical linear techniques in the discrimination of structures of interest. Moreover, it can also be used in a more constrained framework, allowing for an exploration of the manifold of the hemodynamic responses of interest. A solution is proposed for the issue of dimension selection, which is not yet completely satisfactory. However, our studies show the power of the method for data exploration, visualization and understanding.

1. INTRODUCTION

Functional MRI (fMRI) is a neuroimaging modality that produces images of a subject's brain while the former undergoes several experimental conditions. A typical fMRI dataset consists of a hundred or more 3D images that represent an indirect (metabolic) measure of neural activity related to brain processes. While the spatial and temporal resolution allow for an effective exploration of brain metabolism, the presence of noise and various confounds (physiological rhythms) makes signal extraction challenging.

A possible way to overcome these difficulties is to consider that the dataset is a mixture of several patterns corrupted by noise. The detection of such patterns is the goal of principal or independent components analysis [1] [4], or more generally of multivariate analyzes [3]. Since these methods rely on linear models which are still not validated, we propose here a nonlinear point of view that is built on the following geometrical perspective: the set of voxel-based time courses that results from the noisy mixture lives in a low dimensional manifold embedded within the signal

space. If we are able to represent the low dimensional manifold, we obtain a concise representation of the main effects present within the dataset.

This still holds if the dataset has been preprocessed by projection onto a subspace of interest [3]: the variability of the different hemodynamic responses related to the variability in activation amplitudes, delay, shape, or simply attributable to the different weights associated with the different experimental conditions require an adequate way to identify the underlying signal structure. Once again, dimension reduction by standard linear methods, e.g. PCA, may be insufficient to uncover the non-Gaussian, nonlinear or non-orthogonal structures of the data.

Recently, several methods have been proposed for nonlinear dimension reduction, namely Isomap [6], Locally Linear Embeddings (LLE)[5] and Laplacian Embeddings [2]. We choose the third method, since it has the advantage of yielding sparse eigenvalue problems -unlike Isomap- and allows for a meaningful global geometrical interpretation -unlike LLE, see [2] for details. We briefly restate the principles of the method when applied to fMRI data analysis in section 2, then we give two possible uses of the method in section 3: one with unprocessed data; one with preprocessed data. We also make a quick comparison with PCA dimension reduction. A discussion of the method and possible applications is provided in section 4.

2. LAPLACIAN EMBEDDINGS OF FMRI DATASETS

Let X be a $N \times T$ matrix that represents an fMRI dataset. N is the number of voxels, while T is the number of scans or time samples. Each row $X(n)$ of the matrix is a voxel-based time course, while each column is an image. We derive a low-dimensional representation of the dataset:
$\forall n \in [1, \cdots, N]\, X(n) \in \mathbb{R}^T \rightarrow M(n) \in \mathbb{R}^d$. d is not known a priori, and should be inferred from the data.

For each voxel $i \in [1, \cdots, N]$, we compute the set $\mathcal{N}(i)$ of its k (usually equal to 10) closest neighbors within the signal space and use them to compute the following weight

function

$$W_{ij} = \begin{cases} \exp{-\frac{\|X(i)-X(j)\|^2}{2\sigma^2}} & \text{if } i \in \mathcal{N}(j) \text{ or } j \in \mathcal{N}(i), \\ 0 & \text{otherwise.} \end{cases}$$
(1)

Note that the final results depend smoothly (see [7]) on the parameter σ which can be chosen infinite - then $W_{ij} \in \{0, 1\}$; this is our choice in the remainder of the paper. A meaningful representation M of the dataset can be derived by minimizing the functional

$$\sum_{i=1}^{N} \sum_{j \in \mathcal{N}(i)} \|M(i) - M(j)\|^2 \exp\left(-\frac{\|X(i)-X(j)\|^2}{2\sigma^2}\right)$$
$$= \sum_{i,j=1}^{N} \|M(i) - M(j)\|^2 W_{ij},$$
(2)

But it can be shown that $\frac{1}{2}\sum_{i,j=1}^{N} \|M(i) - M(j)\|^2 W_{ij} = tr(M'LM)$, where L is defined from the weight matrix W as follows: letting D be the diagonal matrix $D_{ii} = \sum_j W_{ji}$, then $L = D - W$. This shows that the solutions of (2) are the lowest eigenvectors of L. There remains an arbitrary scale factor in the embedding, that can be removed by setting the constraint $tr(M'DM) = d$. The minimization of $tr(M'LM)$ together with the constraint $tr(M'DM) = d$ is performed by solving the following generalized eigenvalues problem:

$$Lm = \lambda Dm$$
(3)

We sort the solutions of equation (3) by increasing λ. Note that $\lambda_0 = 0$, since by construction $Lm_0 = 0$ for the constant map $m_0 \equiv 1$. The mapping $i \to (m_1(i),.., m_d(i))$ is a low dimensional representation of the dataset. If d is chosen adequately, the geometrical interpretation is that $X(i) \to M(i) = (m_1(i),.., m_d(i))$ is the desired map of the data manifold.

The eigenvalues of the problem determine the amount of distortion related to each of the mappings $X(i) \to m(i)$, and the intrinsic dimension of the data has to be inferred from the sequence of the eigenvalues. This works well on synthetic datasets (see e.g. [7]), but is ambiguous on real datasets. For visualization purposes, we prefer in general two or three-dimensional embeddings.

Before turning to applications, let us outline the computational feasibility of the method: though the matrices W, D and L are of size $N \times N$ -with N ranging typically from 10^4 to 10^5-, all of them are sparse with no more than $O(k)$ nonzero elements per row or column. This allows for sparse coding and quick computation of the first eigenvectors. The computationally heaviest part remains the derivation of the k nearest neighbors of each voxels within the signal space.

3. APPLICATION

3.1. On raw data

We give here a first application of the method in raw fMRI data analysis. By this, we mean that each voxel time series has only been demeaned, detrended and normalized to variance 1. The dataset is a motion localizer on the macaque monkey which is used in [9]; the monkey, which has been injected a contrast agent, passively views static and moving textures in a block design. We analyze only one session of the dataset, consisting of $T = 120$ images, and $N = 12320$ voxels. The total computation time does not exceed 8 minutes on a standard PC. The resulting optimal 2-dimensional map, as well as the time courses associated with each direction, are shown in figure 1. For comparison, the optimal 2D representation obtained by PCA on this same dataset is presented on the same figure. Let us emphasize that the eigenvalues of equation (3) did not give a clear-cut indication of the intrinsic dimension of the dataset. We use a 2D representation for the sake of readability. Moreover, our main point here is the comparison with PCA.
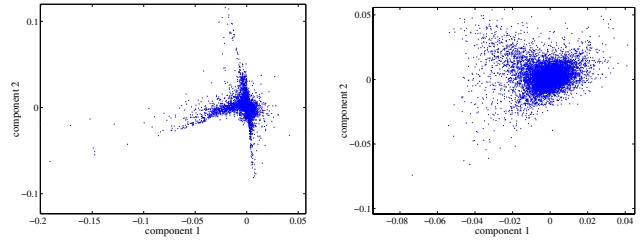
(a) Laplacian 2D representation (b) PCA 2D representation
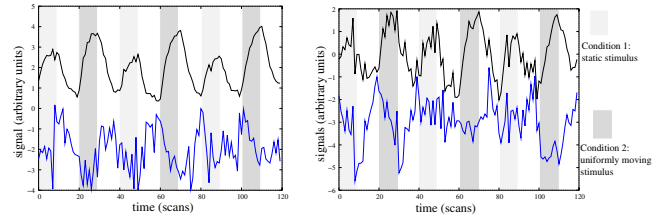
(c) Laplacian time courses   (d) PCA time courses

Fig. 1. (a)-(b): 2D representation of the real dataset, obtained through the Laplacian Embedding (left), and PCA (right). (c)-(d) Main time courses obtained with both methods, i.e. (left) with Laplacian embedding, and (right) PCA. These time courses are weighted average of the voxel time courses, the weights being the first and second coordinate in (a) and (b) respectively.

The comparison of the 2D representations obtained with each method shows the ability of the Laplacian embedding

approach to put in evidence structures of interest, while the PCA yields fuzzy clusters with little or no structure. While the time courses associated with the first component of either decomposition (figure 1, (c) and (d)) are similar, the Laplacian method yields a less noisy curve. This component is typically task-related and represents the average response to the stimuli within the whole dataset. The second time course shows a different oscillating phenomenon, once again more localized (see [7]) and visible through the Laplacian embedding approach.

3.2. On preprocessed data

To study in more details the signal space of interest, it may be worthwhile to preprocess the data, i.e. to project each voxel time course onto the space of assumed task-related responses. This has been done e.g. in [3] for linear multivariate analysis of fMRI data. Here we present a slightly different approach: we fit to each voxel-based time series a finite impulse response model as in [8], thus obtaining a collection of estimates of local task-related responses. Though the fitting procedure considerably reduces the dimension of the dataset, the data still lives in a sub-manifold of this space, which needs to be charted. We thus repeat the Laplacian Embedding experiment, but on the fitted task-related patterns rather than the raw data. We use the same dataset, on which task-related patterns are estimated from 11 sessions instead of 1. Moreover, we only use a subset of $N' = 1320$ voxels on which the estimated task-related patterns passes a minimum complexity test as in [8].

The results are presented in figure 2. The eigenvalues obtained from equation (3) are displayed in (a). We further study a 2D embedding $M = (m_1, m_2)$ of the data. The embedding itself is shown in (b) and reveals a boomerang-shaped manifold. Since the proper visualization of the resulting time courses and spatial maps requires color, we only describe some of the most important features. To simplify matters, we study different segments of the manifold. The latter are color-defined in (b) and obtained by thresholding m_1 and m_2. The average time courses of the voxels within each segment are shown in (c), and the associated spatial maps in (d)-(f). Note that one of the time courses is likely to be an outlier. The spatial layout of the different parts is displayed in figure 2. More complete and color results are available in [7].

The analysis on preprocessed data makes more precise some informations provided by the exploratory analysis. One interesting point is the possibility of visualizing different hemodynamic responses, i.e. responses that differ by the relative weights associated with the different experimental conditions, shape, delay etc. In particular, the 2D embedding obtained here shows that a fraction of the fitted time courses correspond to irrelevant patterns (left wing of the *boomerang* in figure 2(a), light grey time course in figure

2(b)). On the other hand, the three other clusters correspond to consistently task-related patterns that differ by the relative weight of two experimental conditions: while the grey pattern associated with the upper wing of the boomerang shows an activation for both conditions, as well as the darker one (middle part of the boomerang), the other time course (black), is clearly more selectively related to the second experimental condition. These patterns are likely to correspond to functionally different regions, i.e. they are related to different motion sensitivities across different regions of the monkey visual cortex (see figures (d)-(f)).

4. DISCUSSION

The Laplacian Embedding method is a novel tool for fMRI data analysis, that has the following features:

- It reduces the dimension for the dataset. This is of the greatest interest for issues of visualization.

- It is completely data-driven, in the sense that it aims at recovering the manifold from which the different signals are sampled, given only these signals.

- It is specifically designed for accounting for the variability of the activation patterns across the cortex.

- It can be used on raw or preprocessed data.

- It proposes a well-founded way to estimate the dimension of the embeddings, though this procedure may fail on real datasets for which the definition of a dimension is theoretically and practically difficult.

One can also consider Laplacian Embeddings as a low dimensional representation of the data that preserves functional connectivity, the latter being defined as the similarity of the fMRI time courses -the method can indeed be used with any metric that measures functional connectivity.

There are some technical issues that have to be considered with such a method: the choice of the parameters k, σ and the possible existence of multiple connected components. Though there is no place here for long discussions, our own experiments show that:

- Parameter σ can be set to ∞ without creating inconsistencies. The only requirement is that it should be greater or equal to the expected noise variance.

- Parameter k should also be large enough to account for the intrinsic topology of the dataset. In [7], we have shown on a real dataset that $k = 5$ yields a spurious solution, while $k = 10$ or $k = 20$ yield very similar embeddings.

- The presence of multiple connected components in the $k-$neighboring system does not have more serious consequences than splitting the problem into smaller

problems, since the Laplacian matrix is defined for each connected component.

Nonlinear dimension reduction is a novel tool for the exploration of neuroimaging datasets. We have shown the ability of the Laplacian Embedding approach to reveal intrinsic data structures, where standard linear method yield more confuse representations of the data. Moreover, it is also a powerful way to explore the variability of task-related responses, when a preprocessing of the data is performed prior to dimension reduction. This is an important benefit for data visualization. Further work may include more complete assessment of the data dimension and use of more sophisticated measures of functional connectivity.

5. REFERENCES

[1] A. H. Andersen, D.M. Gash, and M.J. Avison. Principal Components Analysis of the Dynamic Response Measured by fMRI: a Generalized Linear Systems Framework. *Magnetic Resonance Imaging*, 17(6):795–815, 1999.

[2] Mikhail Belkin and Partha Niyogi. Laplacian Eigenmaps for Dimensionality Reduction and Data Representation. *Neural Computation*, 15:1373-1396, 2003.

[3] Ferath Kherif, Jean-Baptiste Poline, Guillaume Flandin, Habib Benali, Olivier Simon, Stanislas Dehaene, and Keith J. Worsley. Multivariate model specification for fMRI data. *NeuroImage*, 16(4):1068–1083, August 2002.

[4] Martin J. McKeown, S. Makeig, et al. Analysis of fMRI data by blind separation into independant spatial components. *Human Brain Mapping*, 6:160–188, 1998.

[5] Sam. T. Roweis and Lawrence K. Saul. Nonlinear Dimensionality Reduction by locally Linear Embedding. *Science*, 290:2323–2326, 2000.

[6] Joshua Tenenbaum, Vin De Silva, and C. Langford. A Global Geometric Framework for Nonlinear Dimensionality Reduction. *Science*, 290:2319-2323, 2000.

[7] Bertrand Thirion fMRI data analysis: statistics, information and dynamics. *PhD Thesis*, 2003.

[8] Bertrand Thirion and Olivier Faugeras. Dynamical components analysis of fMRI data through kernel PCA. *NeuroImage*, 20(1):34–49, 2003.

[9] Wim Vanduffel, Denis Fize, Joseph B. Mandeville, Koen Nelissen, Paul Van Hecke, Bruce R. Rosen, Roger B.H. Tootell, and G. Orban. Visual motion processing investigated using contrast-enhanced fMRI in awake behaving monkeys. *Neuron*, 2001.
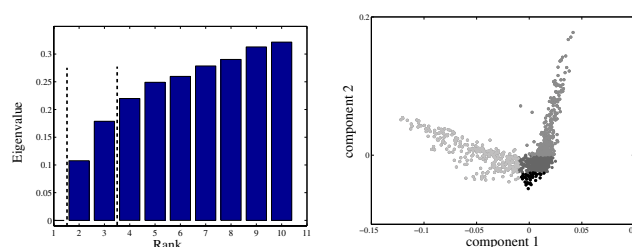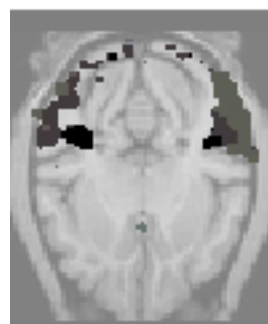
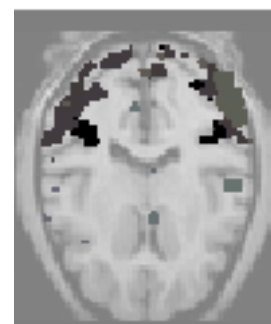(a) Laplacian eigenvalues   (b) Laplacian 2D representation
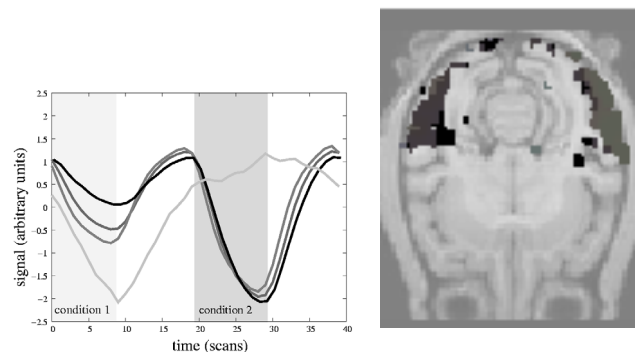
(c) Time courses

(d) Spatial distribution of the clusters

(e) Spatial distribution of the clusters

(f) Spatial distribution of the clusters

Fig. 2. Laplacian embedding of preprocessed data. (a) Sequence of Laplacian eigenvalues; note that there is no clear-cut definition of an intrinsic dimension. (b) 2D embedding of the preprocessed data. This results in a boomerang-shaped manifold, which we divide in four pieces for further study. (c) Time courses obtained for the different clusters. Note that one of them is not a consistent activation pattern. (d-f) Three axial slices of the spatial maps associated with the four clusters defined in (b).