EXTRACTION OF NONLINEAR FEATURES IN MEG AND fMRI DATA OF HUMAN BRAIN

Liya Wang
Department of Biomedical Engineering
University of Wisconsin, Madison WI 53706

Boris Baryshnikov
Department of Medical Physics
University of Wisconsin, Madison WI 53706

Hamid Eghbalnia
Department of Biochemistry
University of Wisconsin, Madison WI 53706

Amir H. Assadi *
Department of Mathematics
University of Wisconsin, Madison WI 53706
assadi@math.wisc.edu

Abstract

We have determined a new method for quantifying the nonlinearities of brain imaging data sets to allow comparison of measures of nonlinearity across comparable experiments. We have an efficient algorithm for finding nonlinear features of high dimensional data from variations in two-dimensional projections and slices. By computing the second order statistics of these features, we are able to construct an analogue of the Riemann Curvature Tensor, thereby describing the nonlinearity of the data set. We demonstrate the usefulness of this technique on real MEG and fMRI data.

Keyword
Principal Components Analysis, multi-channel, data analysis, biomagnetism, auditory perception

Introduction

Working with massive "real world" data is inherently difficult because we often do not know important factors that are significant in any pattern recognition and feature extraction tasks. These factors range from the signal to noise ratio in the data, lack of a statistical distribution for the uncontaminated data, to situations such as not having a robust statistical model for the noise. A simplifying assumption is often helpful, but we must be certain to assess the impact of the assumptions that are made along the way on the final outcome of the analysis. Bayesian methods have the advantage that they make such assumptions explicit in the model. Nonetheless, the Bayesian models will reach their limits in the absence of better estimates on the distribution of data or at least, partial knowledge of such distributions. We propose a method for obtaining

* Li Ya Wang, Hamid Eghbalnia and Amir Assadi were partially supported through NSF-KDI-LIS and other grants to Amir Assadi. Boris Baryshnikov is partially supported through grants to Professor Ron Wakai. The authors are grateful to Professor Ron Wakai of Dept of Medical Physics (UW-Madison) for providing the MEG facilities and scientific supervision of experimental protocols.

relevant probabilistic distributions by extracting information about nonlinearity in those representations of data that lend themselves to certain operations that we call for short “local-to-global integration” and “local linearization”. We apply this technique to a variety of data-types that could fit under the general category of “multi-channel data”. By multi-channel data registration we mean any methodology that extracts the time-series signals from the brain or any other biological source of electromagnetic emission, including EEG, MEG, MCG, fMRI, and multi-electrode recording.

In the application of dynamic PCA to MEG data, we propose a new method for quantifying the nonlinearities of the human brain response during auditory perception and with focus on auditory attention. The results permit a rigorous quantitative comparison of the brain’s processing of auditory information across subjects, and potentially, it provides further insight into general principles underlying processing mechanisms of the human brain. We apply our method to fMRI data and show that for a set of 8 randomly selected voxels, a small temporal window of eigenvectors corresponding to the first three eigenvales can “describe” the data. Our preliminary analysis here illustrates that dynamics PCA can be used as a powerful detection and classification algorithm.

Method

The theoretical underpinning of our algorithms is traced back to viewing data as a geometric entity. A data set is represented by the vectors that register observable features of data points, and forms a geometric subset of the Euclidean space of possibly high dimension. Other representations of the data set could be obtained from geometric properties of this subset, provided that we account for key geometric entities such as metric and topological properties required from the physical properties, as well as biological/behavioral constraints. The intuition behind our method for capturing nonlinear features in data could be briefly outlined as follows. From differential geometry, we know that the Riemannian curvature tensor (RCT) gives a comprehensive description of the nonlinearity in the underlying spaces endowed with certain abstractions of the infinitesimal calculus. A possible approach to description of nonlinearity in data is to look for a concept similar to this marvelous discovery of Riemann. In particular, we confine our study to data sets that can be represented as a collection of points in the Euclidean space. In this context, we define a new geometric structure that lends itself to associate to the data at various points something that is analogous to the Riemann Curvature Tensor that we refer to as “data Curvature Tensor” (dCT). If the data is sampled sufficiently finely from the points on a Riemannian manifold in the standard sense, then the Riemann Curvature Tensor and the dCT are within the approximation bounds defining the sampling of data from the manifold. Thus, the Riemann curvature tensor, and hence nonlinearities of the manifold can be recovered by computing the dCT from available data points. We prove a result that under reasonable circumstances, data a collection of two-dimensional subspaces that linearly approximate local statistics of data suffices for estimation of dCT. Therefore, in the dCT approach, we must determine the nonlinearity of features in the prescribed family of two-dimensional planes in order to have the basic estimates from which we calculate estimates for other nonlinear features of the data. We consider, therefore, estimation of nonlinearities of data in the plane.

In this paper, we propose an information-theoretic estimate for describing the nonlinear dependencies of data in the family of two-dimensional linearization. With this in mind, we must remark that extracting the feature from a general data set is an ill-posed problem in the mathematical sense, and we make certain assumptions to transform it to a well-defined estimate. Thus, it is important to recover such assumptions when we have implicit knowledge from the system that gives rise to the data, here, the MRI instrumentation and basic facts about neurophysiology of the brain underlying the acquired signals. The technical aspect outside of mathematics is to justify the realistic nature of such estimates from fMRI data. We propose to a large-scale data analysis experiment that will either validate our hypotheses, or provide alternative estimates that could be used for analysis of nonlinearity in fMRI in the dCT framework.

We are given time series data from N sources (such as MEG channels) arranged spatially. Each x is the time series from one source and t is sampled to generate L data points. They can be written

$$X(t) = \{x_1(t), x_2(t), \ldots, x_N(t)\}$$
$$t = (t_1, t_2, \ldots, t_L)$$

as,

$$x_{ij} = x_i(t_j)$$

So we have a matrix, where x_{ij} is the element in both i-th row and j-th column. So each row of this matrix is the time course of one recorded source and each column represents the data recorded for all sources at that instant of time.

Consider this matrix as a large window. We choose a small window that has the same height but smaller width. Next, we move this small window across the matrix and also change the window width every time, while at the same time, we do Principal Component Analysis (PCA) to get the sum of the largest two eigenvalues. This process yields a surface such as the following (Figure 1.)

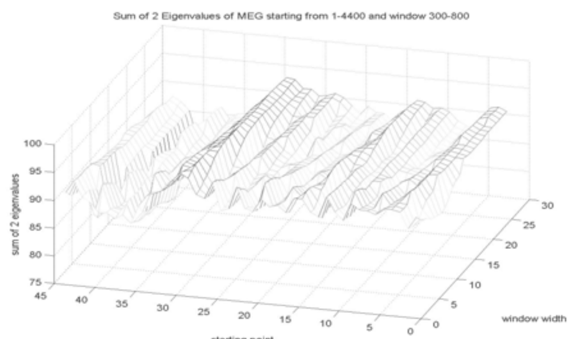

Figure 1

The original matrix for this picture is constructed from fMCG (fetal MCG) signals. The surface depicts that the information captured by two eigenvalues ranges from 79% to 96%. One reason for improvement of data analysis results by dynamic PCA is that, data represented by X is always noisy and often non-stationary. However, we may assume that X is at least quasi-stationary, that is, for each signal from a single source, the local covariance structure varies sufficiently slowly.

Results

In our paper we demonstrate these techniques when applied to real data from fMRI, MEG of human brain and MCG of fetal heart signals. The outcome describes the nonlinearity of the temporal relationship between regions of the human brain during a motor activation task, and in turn, we show what the results reveal about nonlinear dynamics in fMRI.

A 37-channel SQUID magnetometer (Magnes II, Biomagnetic Technologies) in a magnetically shielded room was used for MEG recordings of the brain activity. Five young adults participated in the study. In our protocol subjects were presented with two types of tones. In the active phase of the experiment base tones (1000 Hz, 200 ms) and target tones (1050 Hz, 200 ms) were randomly intermixed in each run. Subjects were asked to count target tones and report the number at the end of the experiment. The brain activity was recorded over auditory cortex area for the corresponding hemisphere. At least 150 epochs were collected per single run with average inter-stimulus interval of 1.5 sec; sampled at 520.8 Hz.

During this experiment evoked (time-locked to the stimulus onset) and induced (not time-locked) oscillatory gamma activity was observed in all subjects. Preliminary analysis showed differences in topography of gamma oscillations and slow auditory evoked response.

Based on the concept of dynamic PCA outlined above, we parameterize spatially and temporally the search space intended for feature extraction from the MEG data. Spatial parameterization determines two independent parameters that specify the subset of channels to be considered, namely, spatial radius and channel location on the sensor. In order to avoid very large search spaces, we use several fixed space configurations for the data analysis below: 1) all channels (maximal radius, central position); 2) the subset of channels that exhibit maximum power in a particular frequency range (slow auditory evoked response or gamma-band response). The latter decision is made based on a preliminary time-frequency analysis of the data. Time parameterization includes two parameters: the size of the window and its origin relative to the beginning of the epoch.

It is believed that gamma oscillations (high frequency oscillatory EEG or MEG activity in the range of approximately 25 to 60 Hz) represent an important mechanism by which the brain binds or integrates spatially distributed processing related to the same percept. Throughout the 1990s in studies in humans and in animals it was shown that oscillatory synchronization in the gamma range occurs between neurons responding to the same percept. This synchronization of neural activity, which was mediated by gamma-band oscillations, was shown to be associated with various stimulus properties such as continuity, vicinity, and common motion [1-3]. The next figures demonstrate the difference in characteristic surfaces for gamma band response in base and target tones presentations. There is a clearly identifiable difference in base vs. target surface.

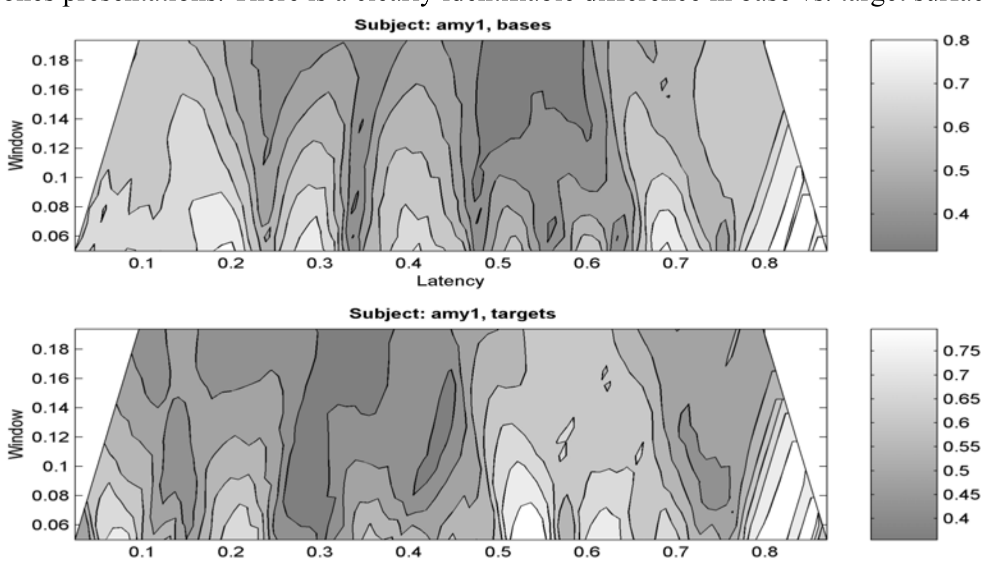

Figure 2. Characteristic surfaces of an averaged gamma band response. In the active listening phase (targets) there is a clear region of complex brain activity (likely to be a high-dimensional space) around 0.3 sec. Both latency and window are in seconds. Stimulus onset is at 0.1 sec on these graphs.

For analysis of fMRI data we use a public data set available via www.fmridc.dartmouth.edu. The data provided in this data set is not highly resolved (64x64x19x91) but is sufficient to illustrate the application of our method. We selected the time series of 8 voxels and used the standard Bartlett test for significance of the number of eigenvectors for the entire time series to describe the data to within $p < 0.05$. The significance results indicated that 4 eigenvectors were needed. We then used our method and tested for significance of one to four eigenvectors. These results indicated that three eigenvectors with the window size of approximately 18 were sufficient to achieve the same significance results. The results are illustrated in the Figure 3 below.

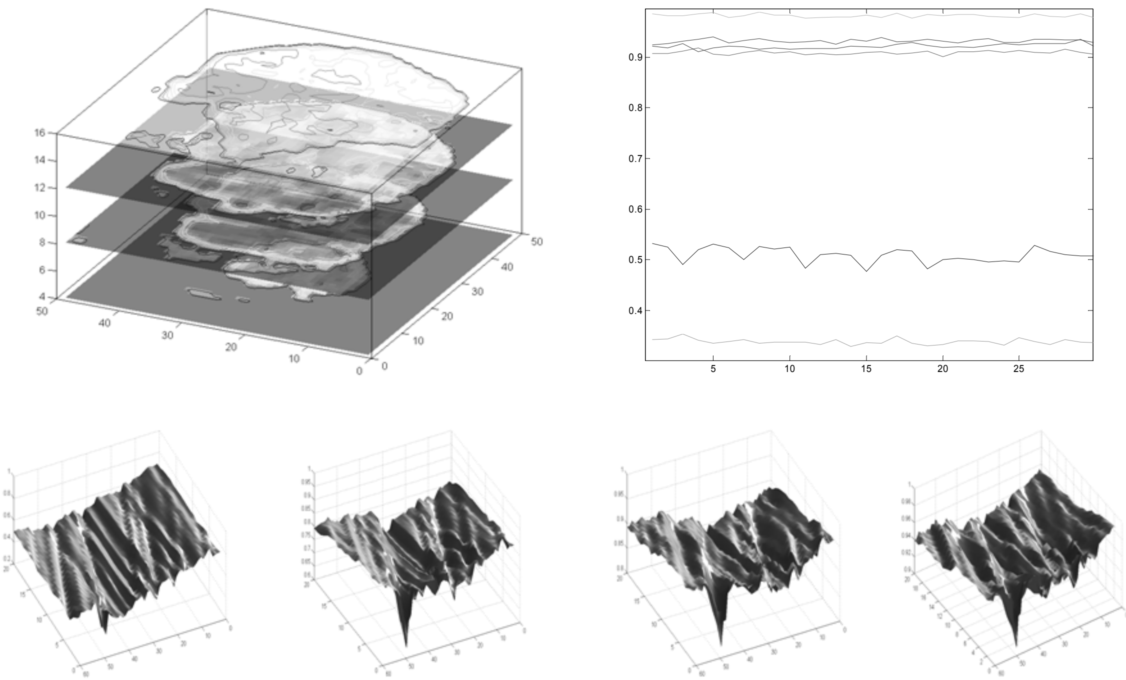

Figure 3. – The top row shows a sample of coronal slices from the dataset at time index 12 on the left and a sample of 6 Voxel time series chosen for analysis on the right for an initial segment of 30 points (out of a total of 91). Voxel values are normalized to one and shown on the y-axis. The bottom row illustrates the variation of total information measured by the percentage of total variance as a function of window size and window location for 8 Voxles (partially shown above). The leftmost window is the first eigenvalue followed by the sum of first two eigenvalues, sum of three, etc. The height in the figures shows the percentage of total variance. The axis labeled with 0-60 indicates the position of the window and the axis labeled with 0-20 indicates the increase in size of the window beyond a nominal size of 10.

The following figures (Figure 4.) show more examples of such surfaces from real data. Our analyses visualize the surfaces such that we can distinguish two key properties: first, the information content exhibits sensitivity to the choice of the initial temporal point; (b) the possible durations could be selected from a range of intervals that proves robustness of the results with respect to the duration of the signal. Depending on applications, other features could be also distinguished.

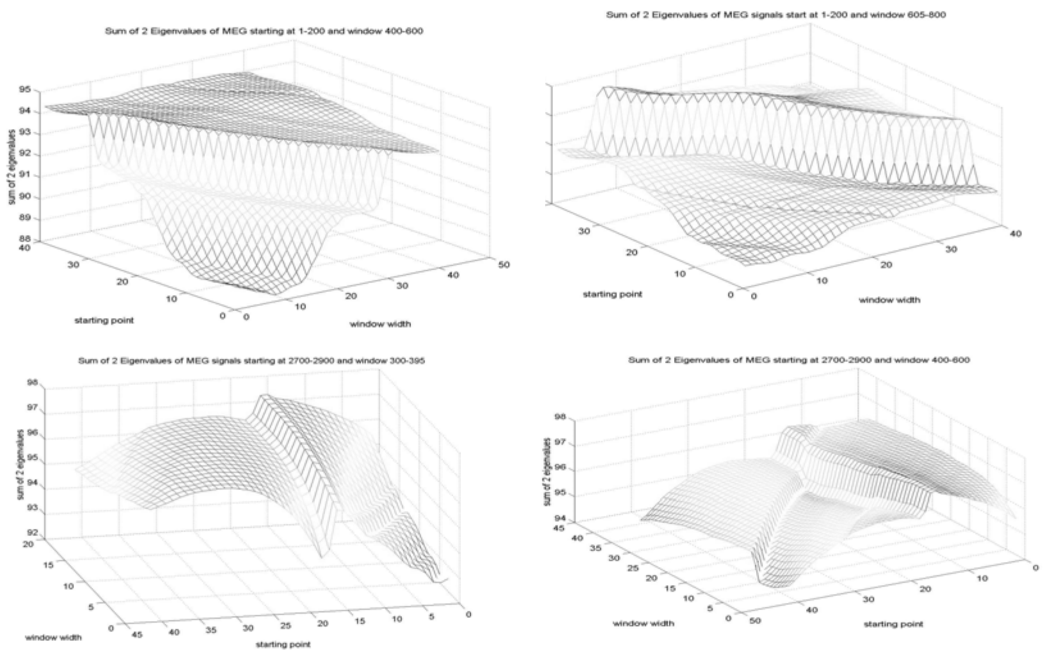

Figure 4

Conclusion

We have proposed a novel approach to identification of the high-dimensional nonlinearity of a data set that embraces ideas from both differential geometry, probability (information geometry). Our algorithm recovers high-dimensional nonlinearity from information-theoretic features in a family of two-dimensional surfaces. This general method is bound to have many applications in multi-channel data analysis. Nonlinearity of information processing in brain could be estimated in terms of the dynamics of information content (e.g. by generalizing Shannon's mathematical theory of communication) from a multi-channel communication system. The encoding part of the multi-channel is in terms of spiking neurons and their secondary measurements as in the BOLD effect in fMRI, electromagnetic fields in MEG or EEG. The decoding side (receiver) guesses the coarse patterns of messages by optimizing extraction of information content in a sparse set of virtual channels (i.e. the algebraically weighted superposition of the output of the physical channels). The paper has given a summary of the progress by our research group to extract biologically significant features from multi-channel data using the above-mentioned framework.

Bibliography

1. D. Fallahati, H. Eghbalnia, M. Backonja, A. Assadi. Dynamic PCA For Network Feature Extraction In Multi-electrode Recording Of Neurophysiologic Data In Cortical Substrates Of Pain. (Accepted for publication in J. Neurocomputing, to appear in 2002)
2. Semmi Pasha, Hamid Eghbalnia, A. Assadi. Dynamical Models for Nonlinear Feature Extraction and Applications to Face Recognition. (Accepted for publication in J. Neurocomputing, to appear in 2002).
3. Singer W., Neuronal synchrony: a versatile code for the definition of relation?, *Neuron*, **24**: 49-65, 1999
4. von der Malsburg C., The what and why of binding: the modeler's perspective, *Neuron*, **24**: 95-104, 1999
5. Engel A.K., Fries P., Singer W., Dynamic predictions: oscillations and synchrony in top-down processing, *Nature Reviews Neuroscience*, **2**(10): 704-716, 2001.