NeuroImage 102 (2014) 945–954

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Technical Note

Optimising beamformer regions of interest analysis

Ashwini Oswal [a,b,*], Vladimir Litvak [a], Peter Brown [b], Mark Woolrich [a,b,c], Gareth Barnes [a]

[a] Wellcome Trust Centre for Neuroimaging, UCL Institute of Neurology, 12 Queen Square, London, UK
[b] Nuffield Department of Clinical Neurosciences, John Radcliffe Hospital, Oxford, UK
[c] Oxford Centre for Human Brain Activity (OHBA), Oxford, UK

ARTICLE INFO

Article history:
Accepted 9 August 2014
Available online 16 August 2014

Keywords:
Beamforming
Regions of interest
Bayesian PCA

ABSTRACT

Beamforming is a spatial filtering based source reconstruction method for EEG and MEG that allows the estimation of neuronal activity at a particular location within the brain. The computation of the location specific filter depends solely on an estimate of the data covariance matrix and on the forward model. Increasing the number of M/EEG sensors, increases the quantity of data required for accurate covariance matrix estimation. Often however we have a prior hypothesis about the site of, or the signal of interest. Here we show how this prior specification, in combination with optimal estimations of data dimensionality, can give enhanced beamformer performance for relatively short data segments. Specifically we show how temporal (Bayesian Principal Component Analysis) and spatial (lead field projection) methods can be combined to produce improvements in source estimation over and above employing the approaches individually.

Introduction

Beamforming is an adaptive spatial filter based method of estimating electrical activity in the human brain based on signals from an M/EEG sensor array. Typically per-location summary statistics of electrical change are used to provide three-dimensional images of brain function. The spatial filter corresponding to a particular brain region is determined based on knowledge of the lead field matrix and from an estimate of the data covariance matrix (Van Veen et al., 1997; Gross et al., 2001; Hillebrand et al., 2005; Brookes et al., 2008). Precise estimation of both the lead fields and the data covariance is therefore essential for accurate beamformer solutions.

This paper focuses around the accuracy of covariance matrix estimation which, perhaps counterintuitively, is inversely proportional to the number of channels (Brookes et al., 2008). The logic being that one needs more data to make an accurate estimate of the covariance between more channels. In fact it can be shown that doubling the number of M/EEG sensors necessitates that the number of data samples (alternatively the time-bandwidth product) is increased four-fold in order to maintain the same covariance matrix estimation error (see Brookes et al., 2008 for further details). This can become a problem when one is interested in relatively short duration or narrow band phenomena (for example the 0.5–1 sec beta rebound, see Pfurtscheller and Lopes da Silva, 1999). In this paper we consider the case in which we do not require whole-brain coverage from the MEG system, but rather have a specific region of interest in mind. This allows us to decrease the effective number of channels and thereby to make either more accurate estimates or estimates of the same accuracy but with less data.

A well tested channel reduction approach involves projection to a sub-space designed to optimally represent sources within a region of interest (ROI) (see Taulu et al., 2004; Ozkurt et al., 2006; Rodríguez-Rivera et al., 2006 for an overview). Often the ROI may be selected a priori based on the experimenter's prior knowledge about areas of task related activity. For example, in Rodríguez-Rivera et al. (2006), the projection is based on the eigenvectors of the source leadfields within an anatomical ROI, and the number of orthogonal components for the projection must be specified by the user. Importantly, this approach for channel reduction incorporates information from all channels and has been shown to produce more precise source estimates than approaches, involving sub-selecting channels based on either power or location (see Rodríguez-Rivera et al., 2006 for more details).

Given a reduced set of sensors (or linear sensor combinations) there remains however the question of whether there is sufficient data to make a reliable covariance matrix estimate. Recent work has shown how using Bayesian PCA one can make an estimate of the latent dimensionality (effective useful number of channels) (Woolrich et al., 2011). Projecting the data into this space and hence ending up with a reduced covariance matrix based on fewer channels is equivalent to optimally regularizing (or diagonally loading) the full covariance matrix.

In this work we propose a two-step procedure that unifies the approaches described above. Firstly, an ROI projection is used to reduce the effective number of channel components a-priori. This step is only

* Corresponding author at: The Wellcome Trust Centre for Neuroimaging, Institute of Neurology, UCL, Queen Square, London, WC1N 3BG, UK.
E-mail address: ashwinioswal@gmail.com (A. Oswal).

http://dx.doi.org/10.1016/j.neuroimage.2014.08.019

based on the forward model. Secondly, Bayesian PCA is applied to further refine the dimensionality estimate based on covariance of the ROI-projected data. As both steps have the effect of reducing the effective number of channels the covariance estimate becomes more robust for the same amount of data.

We proceed by outlining and demonstrating the use of ROI projection and bPCA separately. And then go on to show how the combination of these steps improves the accuracy and resolution of beamforming estimates.

Methods

Spatial dimensionality reduction (ROI projection)

Details of this method can be found in Rodríguez-Rivera et al. (2006). In what follows however we will provide a brief overview of the general principles.

We formulate sensor level MEG activity, x, measured at N channels and T time points as follows:

$$\mathrm{x} = \sum_{l=1}^{L} H(\theta_l)m(\theta_l) + q \tag{1}$$

$H(\theta_l)$ is an N × 3 lead field matrix representing the scaling of the projection of a unit amplitude dipole at location θ_l, to N channels, in the x, y and z directions respectively. Additionally $m(\theta_l)$ represents a 3 × T matrix of time courses in the x, y and z directions (in this paper we will use the MNI coordinate system) for a dipole located at θ_l, where $l = 1\ldots L$. Activity is summed over all sources before adding isotropic Gaussian white noise, q to the sensors.

The goal is to find a transformation, U_r that minimises the error between the representation of the activity of sources, selected from a ROI, in the original data and in the projected data. Assuming that U_r is an N × M matrix with orthonormal columns, where M < N, the projected data takes the form.

$$x_r = {U_r}^t x \tag{2}$$

The N × T matrix x has been transformed to an M × T matrix, x_r corresponding to a reduction in the number of channels from N to M. Rodríguez-Rivera et al. (2006) show that U_r can be computed from the singular value decomposition of the following symmetric matrix:

$$USU^t = \sum_{r=1}^{R} H(\theta_r)H(\theta_r)^t. \tag{3}$$

Accordingly, U_r is set to the M columns of U corresponding to the M largest eigenvalues of B (the eigenvalues may be determined from the diagonal of S). This last formulation simply reduces to the approach for dimensionality reduction used in Friston et al. (2008) (in which case the ROI was defined by the space of lead fields on the cortical surface). In addition to dimensionality reduction of the data, a new leadfield set is computed for each brain location θ_l (see Eq. (10)). The above formulation can be applied when no prior information is available about the dipole moment, or when the dipole moment is known a priori e.g. in the case of surface constrained orientations. As an example of this if we consider a source with known orientation along the x-axis, the projection matrix U_r would be computed only from the first column of $H(\theta_r)$.

An important issue with this approach is selecting the dimensionality M. This dimensionality determines the trade-off between the accuracy of the representation of the ROI and the spatial resolution of the resulting projection. In other words, increasing M leads to a more accurate representation of sources in the ROI, but this comes at the cost of also representing sources outside the ROI. Further insights into this trade off can be gained by considering the mean squared error of the linear transformation, which is represented as the sum of the N–M smallest eigenvalues (given by the diagonal elements in S in Eq. (3)), normalised by the sum of all eigenvalues.

$$e(M)^2 = \frac{\sum_M^N \lambda_i}{\sum_{i=1}^N \lambda_i} \tag{4}$$

Lower values of this error are obtained by minimising the difference between N and M and are associated with more accurate representations of sources in the ROI. A local measure of the ability of the transformation, U_r to represent sources is gained by considering the ratio of the projected source and the original source energies at each spatial location, which mathematically corresponds to the following.

$$F_M(\theta_l) = \frac{tr\left({U_r}^t H(\theta_l)H(\theta_l)^t U_r\right)}{tr\left(H(\theta_l)H(\theta_l)^t\right)} \tag{5}$$

An ideal value for $F_M(\theta_l)$ is 1 for sources within the ROI and 0 for sources outside the ROI. From this it is evident that increasing M will increase the values of the numerator term for sources within and also outside the ROI (see Fig. 1). Additionally it is also evident that this term will depend on both the size of the ROI and on the sampling resolution of the leadfields within the ROI.

Bayesian PCA (bPCA)

The underlying principle of Bayesian PCA is to estimate the true dimensionality of the data based on a generative model and appropriately selected priors. Within the context of beamforming, this estimated dimensionality is then used to act as a surrogate for finding the optimal amount of regularisation required to estimate the data covariance matrix. Expressed more formally, the generative model for bPCA is as follows:

$$x = Gv + q. \tag{6}$$

Here the temporally demeaned data with dimensions N × T (see Eq. (2)) is represented by x. G is of dimensions N × P, where P corresponds to the principal component sensor maps. Finally, v is a P × T matrix of Gaussian latent (or hidden) variables which when multiplied by the principal component sensor maps with additive zero mean isotropic white noise, $q \sim N(0, \sigma^2 I)$, result in the projected data. Woolrich et al. (2011) use a Variational Bayes (VB) approach (Bishop, 1999) based on Automatic Relevance Determination (ARD) hyperparameter thresholding in order to estimate the optimal number of components P, and hence the dimensionality of the data. An alternative approach based on Bayesian Model Selection (BMS) (Minka, 2008) has been shown to be both more accurate and also more computationally efficient, by virtue of avoiding an iterative VB updating routine. This is the approach we use in the present analysis.

The BMS approach involves computing the evidence for differing latent dimensionality models (or values of P from Eq. (6)) of the data. The model with the greatest evidence is then used to infer the true data dimensionality. In essence, a Gaussian likelihood function of the data given the PCA parameters is defined. Combining this likelihood with the required priors gives a complex integral for the model evidence that is efficiently and accurately approximated, using either Laplace's method or the Bayesian Information Criterion (BIC). In practice, the Laplace approximation tends to be more accurate and is for that reason used in the present paper. A detailed mathematical description and derivation of the BMS method can be found in Minka (2008), and a MATLAB implementation of the code is provided in the SPM12 distribution (see spm_pca_order.m).