The Bayesian Power Imaging (BPI) method for magnetic source imaging

Conference or Workshop Item

The Bayesian Power Imaging (BPI) method for magnetic source imaging

R. Hasson[1] and S.J. Swithenby[2]

[1]*Department of Applied Mathematics, The Open University, Milton Keynes, United Kingdom;*
[2]*Department of Physics, The Open University, Milton Keynes, United Kingdom*

1 Introduction

In the biomagnetic inverse problem the main interest is the activation of a region of interest, i.e. the power dissipated in that region. The Bayesian power imaging method (BPI) provides a quantified probability that the activation of a region of interest is above a given threshold. This paper introduces the method and derives the equations used. The method is illustrated in this paper using both experimental and simulated data. Another paper [1] in this volume extends the method to compare task and control experiments.

2 Methods

Before we can begin to describe the method we need a few standard definitions, as used in [2]. Let $m \in \mathbb{R}^N$ be the vector of measurements at a single time instant, i.e. m_i is the measurement of the ith channel, $i = 1, \ldots N$. Let $\vec{L}_i(\vec{r})$ be the lead field of the ith detector.

For simplicity we will use the same Hilbert space of currents as in [2], namely $L_2(Q)$ the square integrable field defined on the brain volume Q with the following inner product between currents

$$\left\langle \vec{j}_1 \, , \, \vec{j}_2 \right\rangle = \int_Q \frac{\vec{j}_1(\vec{r}) \cdot \vec{j}_2(\vec{r})}{\omega(\vec{r})} \, \mathrm{d}\vec{r}$$

where $\omega(\vec{r})$ is a weighting distribution defined on the source space Q. Define the Gramm-Schmidt matrix P, using this inner product, by $P_{ik} = \left\langle \vec{L}_i \, , \, \vec{L}_j \right\rangle$.

This section continues by first deriving expressions for the *a posteriori* probability distribution on the measurements using the results from [2]. In turn this is used to derive an *a posteriori* probability distribution on the activation of brain regions. It is this latter probability distribution that is mapped out in source space.

2.1 Distribution on measurements

The Bayesian method derived in [2] provides the *a posteriori* probability distribution of current distributions $\vec{j}(\vec{r})$. The results of [2] are stated in terms of the projection of this *a posteriori* probability distribution onto one dimensional subspaces of $L_2(Q)$ by using a test current $\vec{t}(\vec{r})$. A special case (namely when the assumed prior current distribution $\vec{j}^{\text{prior}}$ is identically zero) of the results derived in [2] are that the expected mean is $u^{\mathrm{T}}(P + \zeta D)^{-1} m$ and the expected variance is:

$$\alpha^2 \left[u^{\mathrm{T}} \left(P D^{-1} P + \zeta P \right)^{-1} u + \frac{1}{\zeta} \left(\left\langle \vec{t} \, , \, \vec{t} \right\rangle - u^{\mathrm{T}} P^{-1} u \right) \right]$$

where $u_i = \left\langle \vec{t}(\vec{r}) \, , \, \omega(\vec{r}) \vec{L}_i(\vec{r}) \right\rangle$. $\alpha^2 D$ is the covariance matrix of the assumed Gaussian noise and ζ is a regularization parameter. In this Bayesian setting the regularization parameter ζ is equal to α^2/β^2 where β^2 is the variance of the assumed prior probability distribution on currents $\vec{j}(\vec{r})$.

In [2] the test current $\vec{t}$ was scanned across a set of voxels to produce an image. Here the approach is different. By setting $\vec{t}(\vec{r}) = \omega(\vec{r}) \vec{L}_i(\vec{r})$ we can project the *a posteriori* distribution onto the expected measurement in the ith channel. So, the probability distribution represents our best guess of the true value of the measurements given our prior knowledge of the measurement geometry and the data vector m. The expected mean and variance can be calculated using the equations above. The formulae can be simplified by noting that $u_j = \left\langle \omega(\vec{r}) \vec{L}_i(\vec{r}) \, , \, \omega(\vec{r}) \vec{L}_j(\vec{r}) \right\rangle = P_{ij}$, i.e. the vector u is the ith column of P. It follows that the mean and variance for all channels can be calculated in a single matrix equation. In this case the expected mean is $P(P + \zeta D)^{-1} m$ and the covariance matrix is given by

$$\alpha^2 \left[P \left(P D^{-1} P + \zeta P \right)^{-1} P + \frac{1}{\zeta} \left(P - P P^{-1} P \right) \right]$$

which simplifies to $\alpha^2 D (P + \zeta D)^{-1} P$.

A few checks on the reasonableness of the above equations are possible by looking at extreme cases of the regularization parameter ζ.
As $\zeta \to 0$, the mean tends to m and the covariance tends to $\alpha^2 D$. This is reasonable because, in this Bayesian setting, $\zeta \to 0$ corresponds to having either perfect data (i.e. $\alpha \to 0$) or no prior knowledge (i.e. $\beta \to \infty$). In this case, the formulae show that the best estimate of the measurements is given by the data with covariance equal to the assumed covariance of the noise.
As $\zeta \to \infty$, both the mean and the covariance matrix tend to zero. This is also reasonable since $\zeta \to \infty$ corresponds to either worthless data (i.e. $\alpha \to \infty$) or perfect prior knowledge (i.e. $\beta \to 0$). The formulae above indicate that the best estimate of the measurements is equal to our prior prejudices (in this case zero current density) with complete certainty.

2.2 Distribution on activation estimates

The above formulae can be used in the direct power estimation algorithm derived previously [3]. In this method an optimal estimation matrix Y_θ for the θth voxel is derived and the expression $m^{\mathrm{T}} Y_\theta m$ is used to estimate the activation of the θth voxel. The expression derived in [3] for Y_θ is

$$(P^2 + D) Y_\theta (P^2 + D) = P X_\theta P \qquad (1)$$

where X_θ is a matrix which is related to the choice of region of interest. In this paper the region of interest is a single voxel in a source space so the expression for X_θ is simply $L_\theta^k L_\theta^{k\,\mathrm{T}}$ where L_θ^k is a column of the gain matrix for the problem, i.e. L_θ^k is the measurement vector that would result from a dipole at position $\vec{r}_\theta$ oriented in direction $\widehat{e}_k$.
The matrix Y in Equation 1 is then given by;

$$Y_\theta = \sum_{k=0}^{2} \Psi_\theta^k \Psi_\theta^{k\,\mathrm{T}} \qquad (2)$$

where $\Psi_\theta^k = (P^2 + D)^{-1} P L_\theta^k$. For this special case it is more efficient not to form the matrix Y but to calculate the activation directly using Ψ_θ^k;

$$A_\theta(m) = m^{\mathrm{T}} Y_\theta m = \sum_{k=0}^{2} \left(\Psi_\theta^{k\,\mathrm{T}} m \right)^2 \qquad (3)$$

Here the method is put in a Bayesian context because we now know the *a posteriori* distribution of measurement channels. Unfortunately the probability distribution of the quadratic form $m^{\mathrm{T}} Y_\theta m$ is not known in a closed form. Analytic approximations have been derived to calculate the probability distribution of quadratic forms in Gaussian random variables (e.g. [4]) but these methods are computationally expensive. In practice the quickest way of achieving a given level of accuracy for the probability is by Monte Carlo integration. First the measurements are normalized by factoring out the known covariance matrix:

$$m^\star = \frac{1}{\alpha} (P + \zeta D)^{-1/2} P^{1/2} m \qquad (4)$$

Then the probability can be estimated by picking M Gaussian random vectors of length N with zero mean and unit covariance matrix, say z_i, $i = 1, \ldots M$, and computing the probability:

$$P_\theta(x) = \frac{1}{M} \sum_{i=1}^{M} \Phi(x - A_\theta(z_i)) \qquad (5)$$

where $P_\theta(x)$ is the one-sided probability of the estimate x and Φ is the Heaviside step function.
Maps of probability can then be computed by scanning θ across a defined source space.

3 Results

The effectiveness of the algorithms has been tested in two simulated experiments and on real data.
The first simulated experiment has a simple geometry (Figure 1). The head is modelled as a homogeneous conducting sphere of radius 8.9 cm with its centre at $(0, 0, -0.07\,\mathrm{cm})$. The source space is a $6\,\mathrm{cm} \times 6\,\mathrm{cm}$ square thin lamina consisting of 33×33 voxels in the plane $z = -0.01$ cm with centre (0,0,-0.01 cm). The measurement instrument is a hexagonal array of 37 second order axial gradiometers with baseline 5 cm with the lowest 'sensing' coils in the plane $z = 4$ cm.

The sources for the first numerical study were two current dipoles with positions $(0\,\mathrm{cm}, 0\,\mathrm{cm})$ and $(-2.25\,\mathrm{cm}, -2.25\,\mathrm{cm})$ and oriented along the x and y axes respectively. Uncorrelated Gaussian noise is added to the pure signal so that the resulting signal to noise ratio was 9.7.
In analyzing the experimental data there are many possible methods of determining noise levels etc. In this paper we have tried to keep the analysis choices simple in order to concentrate on the fundamentals of the method. With this in mind, all the examples

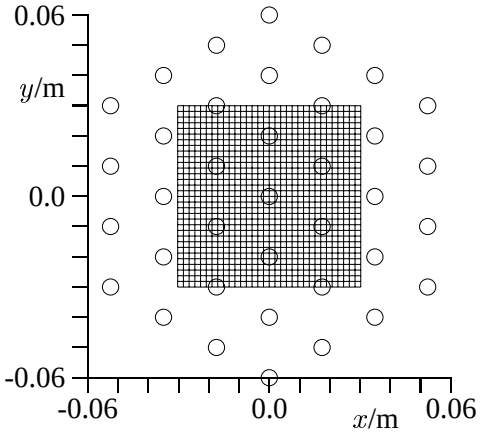

Figure 1: A plan view of the experiment geometry. The grid depicts the source space voxels and the circles represent the position of the measurement coils of the second order gradiometers.

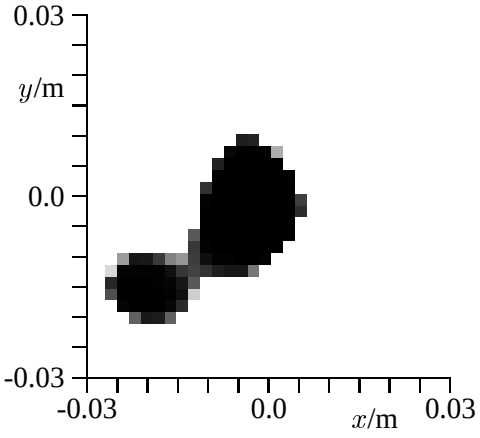

Figure 2: A probability map across the source space shown in Figure 1. The scale goes from white (representing a probability of less than 0.1) to black (probability greater than 0.9).

have the same choices for the parameters: the noise is assumed to be uncorrelated Gaussian noise of variance α^2, the regularization parameter is fixed to be $0.1 \times \text{trace}(P)/N$ and the number of trials, M, for the Monte-Carlo probability estimation method was 1000.

If the Monte-Carlo probability estimation method is applied directly to the raw power estimates, large areas of the source space are found to be significant. This is because it is very unlikely for there to be zero power dissipated in each voxel. An easy calculation shows that the expected power for the θth voxel is given by $\text{trace}(Y_\theta)$. The sensitivity to the (unknown) variance of the noise in the measurements can be minimized by using the mean and standard deviation of the power estimates across the source space as a measure of expected power. This is equivalent to the assumption that the desired signal produces significant power estimates on only a small proportion of the source space. In this paper the examples show probability maps of the probability that the activation is more than three standard deviations above the mean activation. The resulting thresholded image is shown in Figure 2. The two sources are located accurately and with a very high significance.

In the second example, a current dipole source was embedded in the innermost shell of a 3-shell model conducting sphere. The shell radii are 73 mm, 80 mm and 86 mm and the corresponding conductivities are 0.33 Sm^{-1}, 0.0042 Sm^{-1} and 0.33 Sm^{-1}. A detector system consisting of 65 EEG electrodes was used to measure the scalp potential. The electrode positions and a representation of the potentials measured by the electrodes are shown in Figure 3. Also shown in Figure 3 is the conducting shell used as the source space for the reconstruction.

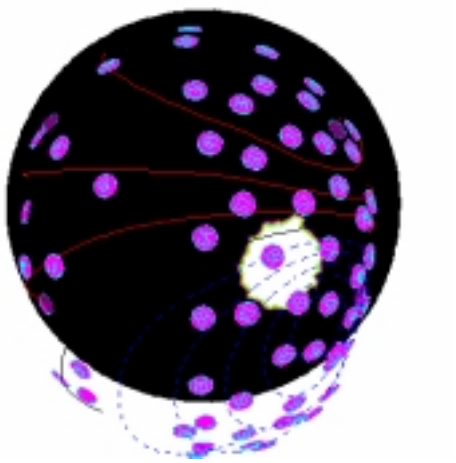

Figure 3: An EEG simulated dipolar dataset, together with the corresponding BPI map. The white blob that appears on the shell is the region where the activation is greater than the threshold (at the 99.9% significance level).

The third and last example involves real MEG data from an evoked response study of face-processing [6]. The experimental instrument used is the Neuromag-122^{TM} [5]. Human subjects were presented briefly with photographs of human faces. It is known that the early response to face images involves widespread activity in the posterior brain but there is limited ev-

idence for the precise distribution and timecourse of the neuronal sources. One suggestion is that there are three major areas of activity; in occipital cortex and both right and left ventral occipito-temporal cortex [7]. Strong occipital activity (starting about 100 ms after the stimulus) is expected to lead to concurrent activity in the two other regions with a stronger response in the right hemisphere [6].

The result shown in Figure 4 is derived from the response in an individual subject at a latency of 160 ms after the stimulus. At this time, there is a global peak in the total field power for this experiment. The source space was restricted to the surface of the cortex, as revealed by MRI and also shown in Figure 4. The findings of the BPI algorithm are consistent with earlier studies in that there is a significant response in both right and left cortices, with a stronger response in right hemisphere. Further work is needed to refine these predictions as the source space has a complex folded structure and we have not yet disentangled the contributions of superficial and deeper source regions

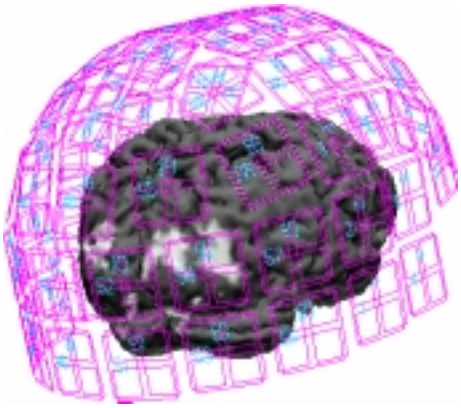

Figure 4: The BPI map on the cortical source space used. The 122-channel helmet detector system is also shown.

4 Discussion

The main impetus behind the development of quadratic algorithms to estimate power rather than current density is that other modalities (e.g. fMRI, PET) naturally produce images of power dissipated. So if multi-modal comparisons are to be made then it is sensible to produce power images from MEG data. The BPI method has the additional advantage of being able to produce spatial images of the variation of probability that are easy to interpret.

In this paper the BPI method was shown to work effectively for simulated MEG and EEG data. It was also demonstrated for experimental MEG data and produced results consistent with earlier analyses.

Acknowledgements

We would like to thank Dr. Sven Bräutigam for many helpful conversations and for the preparation of the face processing data.

References

[1] R. Hasson, J.A. James, and S.J. Swithenby. The Bayesian Power Imaging (BPI) test for task/control experiments. Submitted to Biomag2000, Helsinki, August 2000.

[2] R. Hasson and S.J. Swithenby. A Bayesian test for the appropriateness of a model in the biomagnetic inverse problem. *Inverse Problems*, 15(6):1439–1454, December 1999.

[3] R. Hasson and S.J. Swithenby. Activation curves from optimally shaped regions. In T. Yoshimoto, M. Kotani, S. Kuriki, H. Karibe, and N. Nakasato, editors, *Recent Advances in Biomagnetism*, pages 205–208. Tohoku University Press, Sendai, 1999. ISBN 4-925085-19-0 C3047.

[4] J.P. Imhof. Computing the distribution of quadratic forms in normal variables. *Biometrika*, 48(3/4):419–426, December 1961.

[5] J.E.T. Knuutila, A.I. Ahonen, M.S. Hämäläinen, M.J. Kajola, P.P. Laine, O.V. Lounasmaa, L.T. Parkkonen, J.T.A Simola, and C.D. Tesche. A 122-channel whole cortex squid system for measuring the brains magnetic-fields. *IEEE Trans. Mag.*, 29 (6):3315–3320, 1993.

[6] S.J. Swithenby, A.J. Bailey, S. Bräutigam, O.E. Josephs, V. Jousmäki, and C.D.Tesche. Neural processing of human faces: a magnetoencephalographic study. *Expt. Brain Res.*, 118 (4):501–510, 1998.

[7] S.T. Lu, M.S. Hämäläinen, R. Hari, R.J. Ilmoniemi, O.V. Lounasmaa, M. Sams, and V. Vilkman. Seeing faces activates 3 separate areas outside the occipital visual cortex in man. *Neuroscience*, 43 (2/3):287–290, 1991.