

Is the best estimate of power equal to the power of the best estimate?

R Hasson

Department of Applied Mathematics, The Open University, Milton Keynes, United Kingdom

Abstract. In an inverse problem, such as the determination of brain activity given magnetic field measurements outside the head, the main quantity of interest is often the power associated with a source. The ‘standard’ way to determine this has been to find the best linear estimate of the source and calculate the power associated with this. This paper proposes an alternative method and then relationship to this previous method of estimation is explored both algebraically and by numerical simulation.

In abstract terms the problem can be stated as follows. Let H be a Hilbert space with inner product $\langle \cdot, \cdot \rangle$. Let L be a linear map: $H \rightarrow \mathbb{R}^n$. Suppose that we are given data $b \in \mathbb{R}^n$ such that $b = Lx + e$ where e is a vector of random variables with zero mean and given covariance matrix which represents measurement errors. The problem that is addressed in this paper is to estimate $\langle x, \hat{X}x \rangle$ where \hat{X} is an operator on H (e.g. the characteristic function of a region of interest).

KEYWORDS: Linear inverse problem, biomagnetic inverse problem, magnetoencephalography (MEG).

AMS classification scheme numbers: 65J20, 92C55, 65R30.

Submitted to: *Inverse Problems*

1. Introduction

This paper solves a problem that arose in the study of the inverse problem in magnetoencephalography (MEG) [1, 2]. The dominant concern in MEG analysis has been to produce source maps of current density in the brain and to co-register these to anatomical data (e.g. [1, 3]). However, this may not be the most appropriate approach when there is a focus on specific source regions in the brain, e.g. the thalamus, fusiform gyrus etc. In these cases it may be more appropriate to generate an activation curve, a graph of the power dissipated in a specified region as a function of time. Several methods of generating activation curves have been proposed (e.g. [4, 5, 6]). This aim of this paper is to derive an algorithm for generating activation curves that is optimal with respect to the L_2 -norm.

Another argument for the use of activation curves is the direct comparison with other functional brain imaging modalities such as positron emission tomography (PET) and functional magnetic resonance imaging (fMRI). These modalities produce images of quantities, e.g. regional cerebral blood flow (rCBF), that are correlated with power dissipated rather than current density. This suggests that in order to compare results across modalities we should use magnetic field data to produce an estimate of the power dissipated, i.e. an activation curve.

In Section 2 a more general problem is solved in the setting of a linear map from a Hilbert space to a finite dimensional Hilbert space. The main result from Section 2 (i.e. Equation 16) can be applied independently to each time instant of the data from a MEG experiment. The method proposed is to find a matrix Y such that $b^T Y b$ approximates $\langle x, \hat{X}x \rangle$ (T denotes matrix transposition). The derivation of the optimal matrix Y (Equation 16) with respect to the L_2 -norm is contained in Section 2. Section 3 goes on to compare the main results of Section 2 with the naïve algorithm which first computes an estimate, x_{reg} , using Tikhonov regularization and then computes $\langle x_{\text{reg}}, \hat{X}x_{\text{reg}} \rangle$. This algorithm was used in [4] to extract measures of brain activity.

In Section 4 we specialize to the study of the MEG inverse problem. Definitions appropriate to this application are introduced and a simulation study is described. In Section 5 an important special case is considered where the region of interest is the whole brain. A simplified equation (Equation 32) for this case is derived and this is compared with the total signal power which is commonly used as an estimate of brain activity. Section 6 is a discussion of the merits of the algorithm together with the issues to be addressed before applying the method in practice.

2. Methods

Let H be a Hilbert space with inner product $\langle \cdot, \cdot \rangle$. Let L be a linear map: $H \rightarrow \mathbb{R}^n$. Suppose that we are given data $b \in \mathbb{R}^n$ such that

$$b = Lx + e \tag{1}$$

where e is an unknown vector of random variables with zero mean and covariance matrix C which represents measurement error. Suppose that the problem of finding an $x \in H$ corresponding to a $b \in \mathbb{R}^n$ is an ill-posed problem. The problem here is to estimate $\langle x, \widehat{X}x \rangle$ where \widehat{X} is an operator on H .

It should be noted that no assumptions are made about the noise in the measurement channels other than it has zero mean and a well defined covariance matrix C , i.e. if the measurement noise is denoted by a vector e then the covariance matrix is defined by $C_{ij} = \overline{e_i e_j}$ where $\overline{}$ denotes an expectation value.

Now define the adjoint map L^\dagger by

$$\langle x, L^\dagger b \rangle = (Lx)^T b, \quad \text{for all } x \in H, b \in \mathbb{R}^n. \quad (2)$$

Here we are concerned with the image space \mathcal{I} of L^\dagger . Let $\{\widehat{e}_i : i = 1 \dots n\}$ be the usual basis of \mathbb{R}^n and choose a corresponding basis of \mathcal{I} , $\{\psi_i : i = 1 \dots n\}$, where $\psi_i = L^\dagger \widehat{e}_i$.

The matrix Y will be chosen to minimize the error for points in \mathcal{I} . The starting point in choosing an optimal matrix Y is to derive a suitable cost function to be minimized. We start by expanding $b^T Y b$.

$$b^T Y b = (Lx + e)^T Y (Lx + e) = (Lx)^T Y Lx + e^T Y Lx + (Lx)^T Y e + e^T Y e \quad (3)$$

As mentioned above we focus on points in $\mathcal{I} \subseteq H$, so we express $x \in \mathcal{I}$ in terms of our basis: $x = \sum_{i=1}^n a_i \psi_i$, where $a_i \in \mathbb{R}$ are scalars which will be written collectively as a vector a . Equation 3 can be simplified because the expression Lx appears repeatedly, so start by simplifying this expression:

$$(Lx)^T \widehat{e}_j = \langle x, L^\dagger \widehat{e}_j \rangle = \left\langle \left(\sum_{i=1}^n a_i \psi_i \right), \psi_j \right\rangle = \sum_{i=1}^n a_i \langle \psi_i, \psi_j \rangle. \quad (4)$$

The right hand side of Equation 4 can be written as the j th component of a product Pa where $P_{ij} = \langle \psi_i, \psi_j \rangle$. Note that P is a symmetric positive definite $n \times n$ matrix. Substituting for Lx in Equation 3 gives:

$$b^T Y b = a^T P Y P a + e^T Y P a + a^T P Y e + e^T Y e. \quad (5)$$

The projection of the operator \widehat{X} onto \mathcal{I} has a matrix representation with respect to the basis $\{\psi_i\}$ defined by $X_{ij} = \langle \psi_i, \widehat{X} \psi_j \rangle$ where $i, j = 1, \dots, n$. Hence the target expression can be written in terms of the vector a :

$$\langle x, \widehat{X} x \rangle = a^T X a, \quad \text{where } x = \sum_{i=1}^n a_i \psi_i. \quad (6)$$

For Y to be a good estimator, the right hand sides of Equations 5 and 6 should be ‘close’ for all $a \in \mathbb{R}^n$. One way of achieving this is to minimize the cost function E defined by:

$$E = \|X - P Y P\|_2^2 + \|e^T Y P\|_2^2 + \|P Y e\|_2^2 + \|e^T Y e\|_2^2. \quad (7)$$

where $\|\cdot\|_2$ is the L_2 -norm. Equation 7 can be interpreted in physical terms. The first term is the error in approximating the operator \widehat{X} by Y . The second and third terms give a measure of the overlap, induced by Y , between the measurement error and the

imaging space, \mathcal{I} . Note that these terms are equal for a symmetric Y . The fourth term is a measure of how Y magnifies the measurement error.

To minimize E , $\partial E/\partial Y_{ik}$ is derived for each element of the matrix Y . This gives N^2 equations to solve for the N^2 unknowns Y_{ik} . These may be written as a single matrix equation. In order to illustrate the manipulations involved, the method will be elaborated for the fourth term in Equation 7. The fourth term is expanded using the definition of the L_2 -norm:

$$\|e^T Y e\|_2^2 = \left(\sum_{\alpha, \beta} e_\alpha Y_{\alpha\beta} e_\beta \right)^2. \quad (8)$$

This is differentiated to obtain:

$$\frac{\partial \|e^T Y e\|_2^2}{\partial Y_{ik}} = 2 \left(\sum_{\alpha, \beta} e_\alpha Y_{\alpha\beta} e_\beta \right) e_i e_k = 2 \sum_{\alpha, \beta} e_i e_\alpha Y_{\alpha\beta} e_\beta e_k. \quad (9)$$

We proceed by replacing the products of random variables with their expectation values, i.e. $\overline{e_i e_\alpha} = C_{i\alpha}$ and $\overline{e_\beta e_k} = C_{\beta k}$:

$$\frac{\partial \|e^T Y e\|_2^2}{\partial Y_{ik}} = 2 \sum_{\alpha, \beta} C_{i\alpha} Y_{\alpha\beta} C_{\beta k}. \quad (10)$$

This is the ik th term of the matrix product CYC . Similarly, all of the other terms in Equation 7, when differentiated, give terms that can be written as the ik th elements of a product. So, the equations can be collected as:

$$-2PXP + 2P^2YP^2 + 2P^2YC + 2CYP^2 + 2CYC = 0. \quad (11)$$

This may be written in the form:

$$(P^2 + C)Y(P^2 + C) = PXP. \quad (12)$$

This equation can be solved in many ways, for example by defining $Z = Y(P^2 + C)$ and solving for Z first and then for Y . This easily implemented procedure was rejected as it computes a non-symmetric Y when starting with a symmetric matrix X , because of the numerical problems associated with ill-conditioned matrices. So an alternative scheme which preserves symmetry was devised. Let λ_i be the eigenvalue of the matrix P with eigenvector ϕ_i . Then the matrices X and C can be represented with respect to the basis $\{\phi_i\}$ as new matrices X' and C' , i.e.

$$X = \sum_{ik} \phi_i X'_{ik} \phi_k^T, \quad \text{where } X'_{ik} = \phi_i^T X \phi_k, \quad (13)$$

$$C = \sum_{ik} \phi_i C'_{ik} \phi_k^T, \quad \text{where } C'_{ik} = \phi_i^T C \phi_k. \quad (14)$$

With these definitions, the matrix Y can be finally expressed as:

$$Y = (P^2 + C)^{-1} P \left(\sum_{ik} \phi_i X'_{ik} \phi_k^T \right) P (P^2 + C)^{-1} \quad (15)$$

$$= \sum_{ik} \lambda_i \lambda_k \phi_i (C' + \lambda_i^2 I)^{-1} X'_{ik} (C' + \lambda_k^2 I)^{-1} \phi_k^T \quad (16)$$

The matrix Y computed using the above formula is always symmetric for a given input symmetric matrix X .

Frequently the covariance matrix C is not known and the assumption is made that the random variables e_i are independent Gaussian random variables with a variance ζ that is considered to be a parameter of the method. With this assumption $C = \zeta I$ and Equation 16 becomes:

$$Y = \sum_{ik} \frac{\lambda_i}{\lambda_i^2 + \zeta} \frac{\lambda_k}{\lambda_k^2 + \zeta} \phi_i X'_{ik} \phi_k^T \quad (17)$$

3. Comparison with naïve method

We now compare Equation 17 with the corresponding equation derived by the naïve method mentioned in the introduction. The naïve method for computing $\langle x, \widehat{X}x \rangle$ is to compute a minimum norm estimate using Tikhonov regularization to get x_{reg} and then compute the inner product.

To compute a x_{reg} the first step is to choose a finite dimensional subspace $R \subseteq H$ that has an orthonormal basis $\{r_\alpha : \alpha = 1, \dots, m\}$. The subspace R will be called the representation space and the regularized solution x_{reg} will lie in this space. The linear map $L : H \rightarrow \mathbb{R}^n$ defines a linear map from R to \mathbb{R}^n by restriction that we will also call L .

Now compute a singular value decomposition of $L : R \rightarrow \mathbb{R}^n$ as $L = U\Sigma V^T$, where Σ is a diagonal matrix with non-negative entries $\sigma_1, \sigma_2, \dots, \sigma_n$ and U and V are matrices with orthonormal columns, i.e. $U^T U = V^T V = I$. Applying Tikhonov regularization [7] to the inverse problem gives $x_{\text{reg}} = VDU^T b$, where D is a diagonal matrix given by $D = (\Sigma^2 + \zeta I)^{-1} \Sigma$. So the power dissipated by this source can be computed by:

$$\langle x_{\text{reg}}, \widehat{X}x_{\text{reg}} \rangle = (b^T U D V^T) \mathcal{X} (V D U^T b), \quad (18)$$

where \mathcal{X} is the matrix representation of the operator \widehat{X} on R , i.e. $\mathcal{X}_{\alpha\beta} = \langle r_\alpha, \widehat{X}r_\beta \rangle$. The right hand side of Equation 18 is of the form $b^T \widetilde{Y} b$ where \widetilde{Y} is defined to be:

$$\widetilde{Y} = U D V^T \mathcal{X} V D U^T \quad (19)$$

The comparison with the method in the previous section relies on the relationship between the linear operator L and the Gram-Schmidt matrix P that we will now derive. Suppose for a moment that the representation space R was the whole of H and that the basis $\{r_\alpha\}$ is a complete orthonormal basis for $R = H$. In this case:

$$P_{ij} = \langle \psi_i, \psi_j \rangle = \sum_{\alpha} \langle \psi_i, r_\alpha \rangle \langle r_\alpha, \psi_j \rangle, \quad \text{by completeness,} \quad (20)$$

$$= \sum_{\alpha} \langle L^\dagger \widehat{e}_i, r_\alpha \rangle \langle r_\alpha, L^\dagger \widehat{e}_j \rangle, \quad \text{using the definition of } \psi \quad (21)$$

$$= \sum_{\alpha} \widehat{e}_i^T (L r_\alpha) (L r_\alpha)^T \widehat{e}_j, \quad \text{using the definition of } L^\dagger \quad (22)$$

$$= \widehat{e}_i^T L \left(\sum_{\alpha} r_\alpha r_\alpha^T \right) L^T \widehat{e}_j, \quad \text{by linearity,} \quad (23)$$

$$= \hat{e}_j^T LL^T \hat{e}_i. \quad (24)$$

The right hand side of this equation is the ij th component of the matrix product LL^T . So under the assumption that $\{r_\alpha\}$ is a complete orthonormal basis for H then $P = LL^T$.

Now returning to the case when $R \subset H$ we can see that for a good choice of R the matrix \tilde{P} defined to be LL^T will be approximately equal to P . This is not surprising since to compute the Gram-Schmidt matrix P on a computer one usually takes a suitable representation space R and computes LL^T . The singular value decomposition of L immediately gives an eigenvalue decomposition of \tilde{P} since

$$\tilde{P} = LL^T = U\Sigma V^T V\Sigma U^T = U\Sigma^2 U^T, \quad (25)$$

where the last equality follows from the fact that the columns of V are orthonormal. So the matrix \tilde{P} has eigenvalues σ_i^2 with eigenvectors, $\tilde{\phi}_i$ given by the columns of U .

By a similar argument to the above it can be seen that the matrix \tilde{X}' defined to be $V^T \mathcal{X} V$ approximates the matrix X' so we have:

$$\tilde{Y} = U D \tilde{\mathcal{X}} D U^T = \sum_{ik} \frac{\sigma_i}{\sigma_i^2 + \zeta} \frac{\sigma_k}{\sigma_k^2 + \zeta} \tilde{\phi}_i \tilde{X}'_{ik} \tilde{\phi}_k^T, \quad (26)$$

Now we can compare Equation 26 with Equation 17. For a good representation space R we have $\tilde{\phi}_i \simeq \phi_i$, $\tilde{X}' \simeq X'$ and so the major difference between the two approaches is that $\lambda_i \simeq \sigma_i^2$. The effect of this change can be seen by plotting out the graphs of the functions on the interval $[0, 1]$ (this is the only range of interest since we could dividing by the largest singular value restrict to this interval). These graphs are shown in Figure 1 where it can be seen that Equation 17 attenuates the contribution from the small singular values and has a sharper cut-off than is the case for Equation 26. The effect of this is that Equation 17 should attenuate the noise component, which is usually associated with the small singular values.

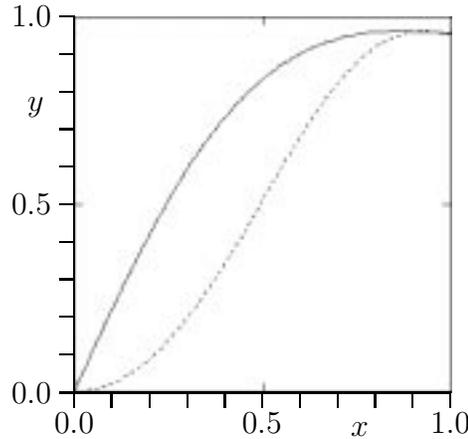


Figure 1. Graphs of the functions $x/(x^2 + \zeta)$ (solid curve) and $x^2/(x^4 + \zeta)$ (dashed curve) for $\zeta = 0.5$.

4. Application

Now we apply our results to the MEG inverse problem, i.e. the problem of recovering information about source current density inside the brain given measurements of the magnetic field outside the brain. Let Ω denote the brain volume. The Hilbert space of interest to us is, $L_2(\Omega)$, the space of square integrable vector fields defined on the brain volume Ω together with the inner product:

$$\langle \vec{j}_1, \vec{j}_2 \rangle = \int_{\Omega} \frac{\vec{j}_1(\vec{r}) \cdot \vec{j}_2(\vec{r})}{\omega(\vec{r})} d\vec{r}, \quad \text{for all } \vec{j}_1, \vec{j}_2 \in L_2(\Omega). \quad (27)$$

The factor $\omega(\vec{r})$ is a weighting factor that allows some flexibility in the procedure. The only restriction imposed on $\omega(\vec{r})$ is that the integral over each voxel is finite. In other papers the factor $\omega(\vec{r})$ has been interpreted as a probability weight [8].

It is interesting in this context to look at the the spatial selectivity implicit in the use of the matrix Y as it varies in source space. Then the sensitivity profile of Y at a point in source space, \vec{r}_0 , is defined to be

$$I(\vec{r}_0) = \sum_{i=1}^3 (L\vec{d}_{\vec{r}_0}^i)^T Y (L\vec{d}_{\vec{r}_0}^i), \quad (28)$$

where $\vec{d}_{\vec{r}_0}^i$ is the current dipole distribution, i.e. $\vec{d}_{\vec{r}_0}^i(\vec{r}) = \delta(\vec{r} - \vec{r}_0) \hat{e}_i$ where $\{\hat{e}_i : i = 1, 2, 3\}$ is an orthogonal set of unit vectors and $\delta(\cdot)$ denotes the Dirac delta function.

The spatial selectivity, $I(\vec{r}_0)$, may be thought of as an instrumental generalization of the lead field of a single measurement channel. The definition is designed so that in the case when $Y_{ik} = 1$ when $i = k = n_0$ and 0 otherwise then the sensitivity $I(\vec{r}_0)$ is the square of the magnitude of the lead field of channel n_0 . Note that the above definition of $I(\vec{r}_0)$ is different from the original definition proposed in [9].

To illustrate the method a simple simulated experimental system (Figure 2) has been investigated. The head is modelled as a homogeneous conducting sphere of radius 8.9 cm with its centre at $(0, 0, -0.07 \text{ cm})$. The source space is a $9 \text{ cm} \times 9 \text{ cm}$ square thin lamina consisting of 33×33 voxels in the plane $z = -0.01 \text{ cm}$ with centre $(0, 0, -0.01 \text{ cm})$. 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 $z = 4 \text{ cm}$. Now consider, in the context of the simulated system, the simplest possible region of interest operator $\hat{X} = \delta(\vec{r} - \vec{r}_c)$ where $\vec{r}_c = (0, 0, -0.01 \text{ cm})$ is the centre of source space. This type of operator might be adopted if one simply wished to focus on a small volume of source space. The matrix Y used as an estimator from this operator is calculated using Equation 17. The sensitivity profile for this Y matrix is shown in Figure 2.

The reconstruction of an activation curve has been tested on simulated data using this region of interest operator and simulated data from a time varying target dipole at $(0, 0, 0 \text{ cm})$, i.e. 1 cm from the region of interest. The moment of the dipole varies sinusoidally at 10 Hz, with an envelope that rises linearly from zero at 200 ms to a maximum at 300 ms after which it remains constant. To show the insensitivity to dipole orientation the dipole moment was made to rotate smoothly in a tangential plane —

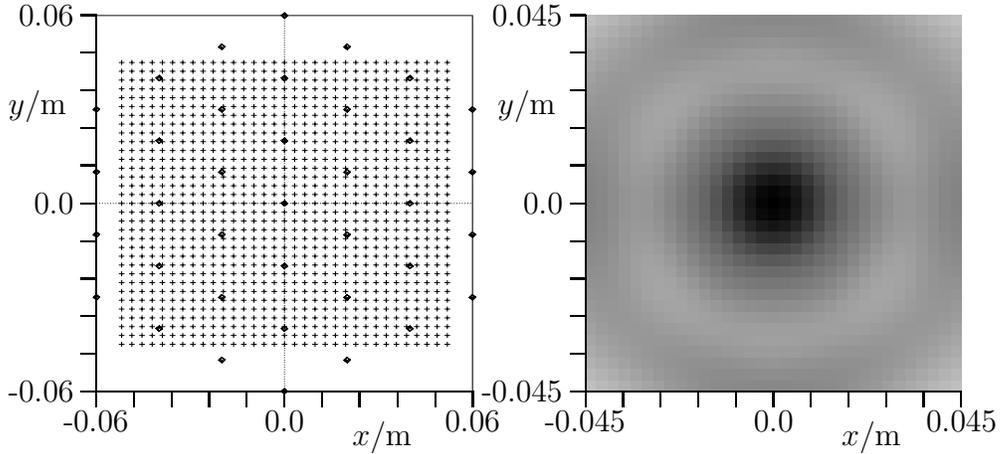


Figure 2. (left) A plan view of the experiment geometry. Crosses denote source space voxels and diamonds denote the projections of the centres of the detector coils. (right) The sensitivity profile in source space of the Y matrix that is derived from the operator $\hat{X} = \delta(\vec{r} - \vec{r}_c)$.

this rotation is not discernible in the activation curve. In addition to the target dipole there is distractor dipole at $(0, 0.02 \text{ cm}, 0)$, which is active from 0 to 100 ms (triangular envelope) and again from 400 ms (square envelope).

In the period from 200 ms to 400 ms when only the target dipole is active, the calculated (power) activation curve matches closely that of the target. However, the existence of the distractor dipole within the sensitive region (see Figure 2) gives rise to apparent activity between 0 ms and 100 ms and inaccuracy in the calculated activation curve for the period after 400 ms. The distractor dipole adds to the estimated power dissipated when it is parallel to the target and subtracts when the target dipole has rotated to be anti-parallel.

Error bars for the activation curve can be estimated using the last term in Equation 7 to give the amount of measurement noise reflected in the activation curve. The estimate is given by $\sum_{\alpha,\beta} C_{\alpha\beta} Y_{\alpha\beta}$.

5. Total brain activity

As a special case of Equation 17 the task of finding an estimate of the total activity in the source space is considered. In this case the operator \hat{X} is the identity and so

$$X_{ij} = \langle \psi_i, \hat{X} \psi_j \rangle = \langle \psi_i, \psi_j \rangle = P_{ij} \quad (29)$$

So the matrix X' can be calculated as follows

$$X'_{ij} = \phi_i^T X \phi_j = \phi_i^T P \phi_j = \lambda_j \phi_i^T \phi_j = \lambda_j \delta_{ij} \quad (30)$$

where δ_{ij} is the Kronecker delta. So, in this case, Y is given by the simplified formula:

$$Y = \sum_{ij} \frac{\lambda_i}{\lambda_i^2 + \zeta} \frac{\lambda_j}{\lambda_j^2 + \zeta} \phi_i \lambda_j \delta_{ij} \phi_j^T = \sum_i \frac{\lambda_i^3}{(\lambda_i^2 + \zeta)^2} \phi_i \phi_i^T \quad (31)$$

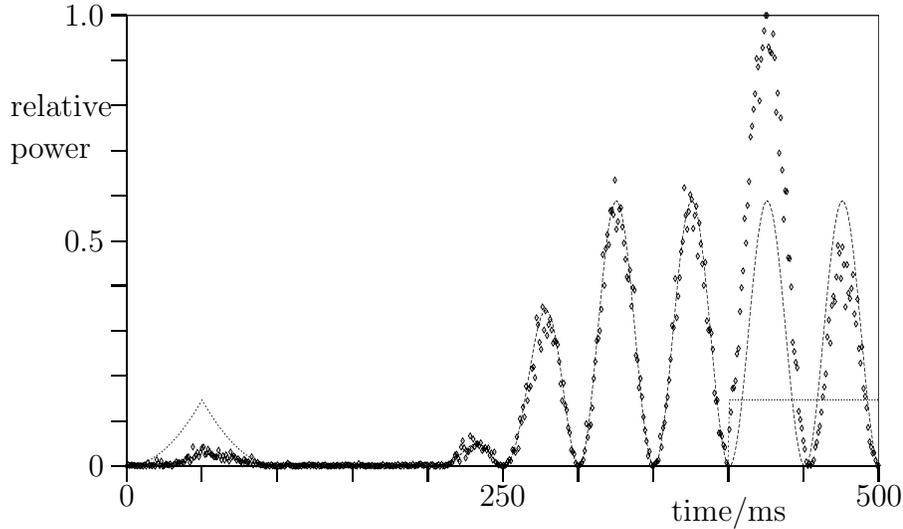


Figure 3. Activation curves for a simulated experiment. The solid line and the dotted lines are the activation curves of the target and distractor dipoles. The diamonds are the calculated activation curve from the Y matrix whose sensitivity profile is shown in Figure 2. The error bars, omitted for clarity, would be approximately twice the height of the diamonds.

This gives the following formula for computing the total activity.

$$\text{Total activity, } A(t) = \sum_i \frac{\lambda_i^3}{(\lambda_i^2 + \zeta)^2} (\phi_i^T b(t))^2 \quad (32)$$

where $b(t)$ is the vector of measurements collected at time t .

Previously when an estimate of the total brain activity was needed the power in the signals was used, i.e.

$$\text{Total signal power, } B(t) = b(t)^T b(t) \quad (33)$$

These two methods have been compared for the simulated data described above as shown in Figure 4. In Figure 4 it can be seen that the estimate $A(t)$ (shown as the solid line on the left) more closely approximates the true activation of the dipoles (dashed curve) than the estimate $B(t)$. In fact, if the error in the estimate is measured by the integral of the squared discrepancies between the curves then the error for $A(t)$ is 2.6×10^{-4} whilst the error for $B(t)$ is 6.6×10^{-4} .

6. Discussion

We have shown that it is possible to directly compute the ‘power’ associated with a source without computing the source first. The method seems robust to noise and is not dependent on the noise having a Gaussian profile. Correlations between measurement channels are fully taken into account. In particular it was shown that activation curves of brain regions can be obtained from magnetic field data. The method provides an easily computable way of tracking the power dissipated in a specific region of the brain.

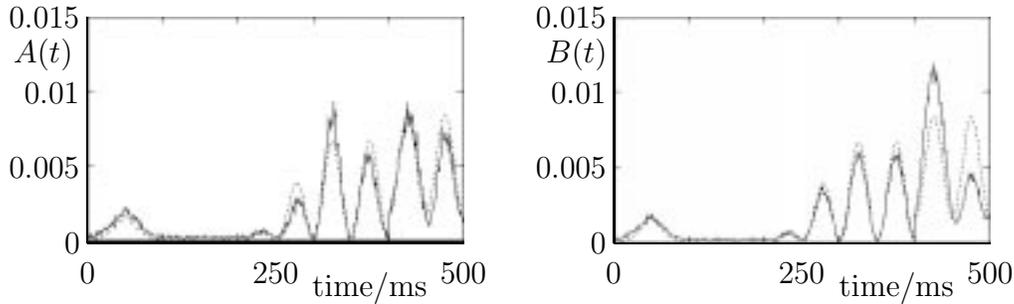


Figure 4. (left) A comparison of the total brain activity, $A(t)$, (solid line) with a plot of the power of the dipolar sources that generated the simulated data (dashed line). In order to compare with the right-hand diagram both curves are normalized to enclose a unit area. (right) A comparison of the total signal power, $B(t)$, (solid line) with a plot of the power of the dipolar sources that generated the simulated data (dashed line). In order to compare with the left-hand diagram both curves are normalized to enclose a unit area.

To use the method effectively the practical problem is to effectively estimate the covariance matrix. For evoked response experiments the covariance matrix, C , can be estimated from the prestimulus period. For other experiments it might be more suitable to make the *a priori* assumption that the noise is uncorrelated Gaussian noise with variance a α^2 that could be considered as a parameter. As α increases, the more closely the Y matrix sensitivity pattern matches the region of interest, but the larger the error bars on the resulting activation curve.

Finally, to answer the question in the title, I would say that if best is interpreted in a least L_2 -norm sense then the answer is no. The best way to estimate the power associated with a source is to compute it directly.

References

- [1] Jukka Sarvas. Basic mathematical and electromagnetic concepts of the biomagnetic inverse problem. *Phys. Med. Biol.*, 32(1):11–22, 1987.
- [2] M. Hämäläinen, R. Hari, R.J. Ilmoniemi, J. Knuutila, and O.V. Lounasmaa. Magnetoencephalography - theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of modern physics*, 65(2):413–497, 1993.
- [3] D. Schwartz, D. Lemoine, E. Poisot, and C. Barillot. Registration of MEG/EEG data with 3D MRI: methodology and precision issues. *Brain Topography*, 9(2):101–116, 1996.
- [4] K.D. Singh, A.A. Ioannides, R. Hasson, U. Ribary, F. Lado, and R. Llinas. Extraction of dynamic patterns from distributed current solutions of brain activity. In M. Hoke, S.N. Ern e, Y.C. Okada, and G.L. Romani, editors, *BIOMAGNETISM: Clinical Aspects*, pages 767–773, Amsterdam, August 1992. Elsevier.
- [5] C.D. Tesche, M.A. Uusitalo, R.J. Ilmoniemi, M. Huotilainen, M. Kajola, and O. Salonen. Signal space projections of MEG data characterise both distributed and well-localised neuronal sources. *Electroenceph. Clin. Neurophysiol.*, 95:189–200, 1995.
- [6] S. E. Robinson and D. F. Rose. Current source image estimation by spatially filtered MEG. In M. Hoke, S. N. Ern e, Y. C. Okada, and G. L. Romani, editors, *Biomagnetism: Clinical Aspects*, pages 761–765, Amsterdam, 1992. Elsevier.
- [7] P.C. Hansen. Regularization tools. *Numerical Algorithms*, 6:1–35, 1994.

- [8] R. Hasson and S.J. Swithenby. The theoretical basis of iterative distributed solutions to the biomagnetic inverse problem. In *Advances in Biomagnetism Research: BIOMAG96*, (Eds: C. Aine et al) Springer-Verlag, New York, In press, 1999.
- [9] 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.