# Estimating the Number of Sources in Magnetoencephalography Using Spiked Population Eigenvalues – Supplementary Materials

December 13, 2016

# 1. FORWARD AND INVERSE PROBLEMS IN MAGNETOENCEPHALOGRAPHY (MEG)

## 1.1 The basics of MEG

The MEG signal has been divided into two parts. The *primary* current is the net effect of ionic currents flowing in the dendrites of neurons. Everything else is the *volume* current, which is caused by conduction effects, and flows passively in the medium. The primary current is related to the post-synaptic activity, and it is widely accepted that the net currents, mainly due to the primary currents, can be thought of as current dipoles (Hamalainen et al. 1993). A localized dipole is mathematically characterized as a simple point in the brain, so that the magnetic field generated by this dipole can explain the MEG measurements. Those 'focal' dipoles are presented to activate a patch of cortex in the brain. The focal brain activation can be observed in epilepsy, or it can be induced by a stimulus in neuropsychological experiments; see Brenner et al. (1978) for a famous example of dipolar fields patterns observed in the flow of electric current in the brain in response to the electrical stimulation of the finger.

The current dipoles are usually parametrized by means of their location, orientation and strength, as follows:

$$\mathbf{J}_i^P(\mathbf{r}) = Q_i \delta(\mathbf{r} - \mathbf{r}_i) \tag{1.1}$$

where $\delta(\cdot)$ is the Dirac delta function, and the $Q_i$ is a charged dipole at the point $\mathbf{r}_i$ in the brain volume $\Omega$. Using the quasi-static approximation to Maxwell's equations (that is, ignoring the partial derivatives with respect to time) in Sarvas (1984), the magnetic field $\mathbf{B}$ at $\mathbf{r}$ of a current dipole $\mathbf{J}_i^P$ at $\mathbf{r}_i$ inside a brain volume can be calculated by the Biot-Savart equation,

$$\mathbf{B}_i(\mathbf{r}) = \frac{\mu_0}{4\pi} \int_\Omega \frac{\mathbf{J}_i^P(\mathbf{r}_i) \times (\mathbf{r} - \mathbf{r}_i)}{|\mathbf{r} - \mathbf{r}_i|^3} d\mathbf{r}_i. \tag{1.2}$$

Although a single dipole model has shown great potential in quantifying MEG measurements (Hamalainen et al. 1993) and has been widely used ever since, electric sources in the brain might be too extended to be represented by only one dipole. Common approaches in the literature are to discretize the region of interest (ROI) in the brain into a finite number of volume elements (voxels), and attach mutually orthogonal dipoles with unknown amplitudes at each voxel. To be precise, assume that the electric field $\mathbf{E}$ is generated by the primary current $\mathbf{J}^P$, which comes from the sum

of $N$ localized current dipoles at locations $\mathbf{r}_i$

$$\mathbf{J}^P(\mathbf{r}) = \sum_{i=1}^{N} Q_i \delta(\mathbf{r} - \mathbf{r}_i). \tag{1.3}$$

In this case, where multiple current dipoles are presented, the induced magnetic fields $\mathbf{B}_i(\mathbf{r})$ from $N$ different dipoles would simply add up. This model implicitly states the *forward problem* in MEG; that is, given a particular configuration of current dipoles in the brain, the magnetic field outside the head is calculated as follows,

$$\mathbf{B}(\mathbf{r}) = \sum_{i=1}^{N} g(\mathbf{r}, \mathbf{r}_i) \cdot Q_i \tag{1.4}$$

where

$$g(\mathbf{r}, \mathbf{r}_i) = \frac{\mu_0}{4\pi F^2(\mathbf{r})} \mathbf{r}_i \times [F(\mathbf{r})\mathbf{e} - (\nabla F(\mathbf{r}) \cdot \mathbf{e})\mathbf{r}]$$

is the lead-field vector and

$$F(\mathbf{r}) = |\mathbf{r} - \mathbf{r}_i| \left(|\mathbf{r}| \, |\mathbf{r} - \mathbf{r}_i| + |\mathbf{r}|^2 - \mathbf{r}_i \cdot \mathbf{r}\right).$$

The equation (1.4) essentially means that by assuming the locations of the dipoles to be fixed, the strength and moments of dipoles are related linearly to the magnetic field produced at the sensor location $\mathbf{r}$; and because of this linearity, the calculated magnetic field is a linear combination of the magnetic fields, namely the the lead-field $g(\mathbf{r}, \mathbf{r}_i)$, induced by each individual unit dipole at $\mathbf{r}_i$. In this sense, the $\mathbf{B}(\mathbf{r})$'s, observed at various $\mathbf{r}$'s, are therefore spatial filters encoded with information about dipole locations, and how they are presented at MEG sensors.

## 1.2 The ill-posedness of the inverse problem

While in general the forward problem described in (1.4) allows one to accurately calculate the magnetic fields produced by a known dipole configuration, or several dipoles, the *inverse problem*, which refers to the calculation of the dipole distribution in the brain given a measured magnetic field, is ill-posed by the definition of J. Hadamard: i.e., the existence, uniqueness and stability of the solutions are not always guaranteed (Tikhonov, Leonov and Yagola 1998). This is due to the fact that (i) one can always construct a large number of nontrivial dipoles in $\Omega$ that have a vanishing magnetic field on the boundary of $\partial\Omega$. (ii) The integral operator in (1.2) is a compact

operator, which has an instability phenomenon in that a small amount of noise in the measurement data can lead to enormous errors in the estimates.

To be precise, consider $\mathcal{A} : L^2(\Omega)^3 \to L^2(\partial\Omega)^3$ to be the operator defined in (1.2) such that

$$(\mathcal{A}\mathbf{J})(\mathbf{r}) = \int_\Omega \mathbf{J} \times \Phi(\mathbf{r}, \mathbf{r}_i) d\mathbf{r}_i, \ \mathbf{r} \in \partial\Omega \tag{1.5}$$

where $\Phi(\mathbf{r}, \mathbf{r}_i) = \frac{\mu_0}{4\pi} \frac{(\mathbf{r}-\mathbf{r}_i)}{|\mathbf{r}-\mathbf{r}_i|^3}$ and $L^2(\Omega)^3 := L^2(\Omega) \times L^2(\Omega) \times L^2(\Omega)$. Note that here we have replaced $\mathbf{J}^P(\mathbf{r}_i)$ with $\mathbf{J}$ just for a better presentation. Theorem 1.1, a generalized version from Kress, Kuhn and Potthast (2002), implies that, theoretically, there is no unique solution for the reconstruction of dipole distributions from their magnetic fields; that is, given that any dipole solution $\mathbf{J}$ in the null space $Null(\mathcal{A}) \subset L^2(\Omega)^3$ is a solution of the inverse problem, a new dipole configuration $\mathbf{J}^*$ satisfying

$$\mathbf{J}^* = \mathbf{J} + \eta \mathbf{J}_\psi, \ \ \forall \eta \in \mathbb{R}$$

where $\mathbf{J}_\psi \in \Psi$ with $\Psi$ being defined shortly after, is also a solution.

**Theorem 1.1** *The null space*

$$Null(\mathcal{A}) = \left\{ \mathbf{J} \in L^2(\Omega)^3 : \mathcal{A}\mathbf{J} = 0 \right\}$$

*of the operator $\mathcal{A} : L^2(\Omega)^3 \to L^2(\partial\Omega)^3$ contains the linear space*

$$\Psi := \left\{ \mathbf{J} \in \Delta\psi : \psi \in \mathring{W}_2^2(\Omega)^3 \right\}.$$

*where $\mathring{W}_2^2(\Omega)^3 = \{\psi \in W_2^2(\Omega)^3 : \gamma\psi = 0\}$. Here, $W_2^2(\Omega)^3$ is the standard Sobolev space and $\gamma : W_2^2(\Omega)^3 \to L^2(\partial\Omega)^3$ is the trace operator (Dautray and Lions 1998).*

Now, consider the instability of the inverse problem in (1.5). Let $\bar{\mathbf{J}}$ be the exact solution. Hence, $\bar{\mathbf{B}} = \mathcal{A}\bar{\mathbf{J}}$ is the corresponding exact measurement. Define $\mathbf{J}_\omega = \bar{\mathbf{J}} + N \sin(\omega|\mathbf{r}_i|)$. Then $\mathbf{B}_\omega = \bar{\mathbf{B}} + N \int_\Omega \Phi(\mathbf{r}, \mathbf{r}_i) \sin(\omega|\mathbf{r}_i|) d\mathbf{r}_i$. By Riemann-Lebesgue lemma we have

$$\|\mathbf{B}_\omega - \bar{\mathbf{B}}\|_{L^2(\partial\Omega)^3} = |N| \sqrt{\int_\Omega d\mathbf{r} \left[ \int_\Omega \Phi(\mathbf{r}, \mathbf{r}_i) \sin(\omega|\mathbf{r}_i|) d\mathbf{r}_i \right]^2} \to 0 \text{ as } \omega \to \infty,$$

which means that the inexact data $\mathbf{B}_\omega$ approximates $\bar{\mathbf{B}}$. However, the corresponding solution $\mathbf{J}_\omega$ by data $\mathbf{B}_\omega$ does not approximate $\bar{\mathbf{J}}$, i.e., $\mathbf{J}_\omega$ will not converge to $\bar{\mathbf{J}}$, either in $C(\Omega)^3$ or in $L^2(\Omega)^3$ since

$$\|\mathbf{J}_\omega - \bar{\mathbf{J}}\|_{C(\Omega)^3} = |N| \text{ and } \|\mathbf{J}_\omega - \bar{\mathbf{J}}\|_{L^2(\Omega)^3} = |N|\sqrt{\frac{|\Omega|}{2} - \frac{D}{2\omega}},$$

where $D = D(\omega, \Omega)$ is a bounded function ($|D(\omega, \Omega)| \leq 1$).

It is well known that to obtain a stable approximate solution for an ill-posed problem, a regularization procedure should be employed. Hence, solving the inverse problem

$$(\mathcal{A}\mathbf{J})(\mathbf{r}) = \mathbf{B}(\mathbf{r}), \ \mathbf{r} \in \partial\Omega \tag{1.6}$$

requires certain regularization, often known as Tikhonov regularization. Obviously, the regularizing operator for the linear operator (in our case, it is an integral operator) is not always linear. Suppose that in practice, instead of the exact data $\{\mathcal{A}, \mathbf{B}\}$, we are only given the approximate "admissible" data $\{\mathcal{A}_h, \mathbf{B}_\delta\}$ and a vector of error levels $\eta = (h, \delta)$ such that $\|\mathcal{A}_h - \mathcal{A}\|_{L^2(\Omega)^3 \to L^2(\partial\Omega)^3} \leq h$ and $\|\mathbf{B}_\delta - \mathbf{B}\|_{L^2(\partial\Omega)^3} \leq \delta$. Within the framework of the Tikhonov regularization, equation (1.6) is substituted by

$$\inf \|\mathcal{A}_h\mathbf{J} - \mathbf{B}_\delta\|_{L^2(\partial\Omega)^3} + \alpha\Gamma[\mathbf{J}], \tag{1.7}$$

where functional $\Gamma$ is known as a Tikhonov stabilizer, which contains additional (*a priori*) information on the solution, such as the smoothness of the solution function. For the traditional Tikhonov method, $\Gamma[\mathbf{J}] = \|\mathbf{J}\|^2_{W_1^2}$, where $W_1^2$ is a standard Sobolev space. Here, $\alpha > 0$ is the so-called regularization parameter, and the regularizing procedure is to select a parameter $\alpha = \alpha(W_h, \mathbf{B}_\delta, \eta)$ such that it ensures the convergence of the obtained approximate solution (minimizer of (1.7)) to the exact one $\bar{\mathbf{J}}$ when $\eta \to 0$.

There are various regularization parameter selection methods, such as *a priori* methods (Tikhonov et al. 1995) or *a posteriori* methods (Tikhonov et al. 1998). However, all of them need the error information $\eta$, since a regularizing algorithm without using error level $\eta$ exists only for well-posed problems. Therefore, if the error level is not available, one might consider reformulating the original problem into a well-posed one by adding some additional information on the solution, such as statistical assumptions.

## 2. A COMPLETE ALGORITHM DETERMINING THE INTRINSIC DIMENSIONALITY (ID) (THE NUMBER OF SOURCES)

**Algorithm:** Estimate the spiked population eigenvalues and determine the ID (the number of sources)

**Input**: Set $\gamma_T = K/T$, $l = 1$, $k_1 = 1$. Compute the sample covariance matrix $\widehat{\mathbf{R}}$ of the data.

1. If the data has a negligible noise, set $\widehat{\mathbf{R}}_{adj} = \widehat{\mathbf{R}}$, and go to Step 3; otherwise, go to the next step.

2. Given any estimated $\widehat{\mathbf{R}}_n$, calculate the quasi-optimal transformation $\widehat{\mathbf{X}}$ and the corresponding quasi-optimal associated covariance matrix $\widehat{\mathbf{R}}_{adj}$.

3. Calculate all eigenvalues of $\widehat{\mathbf{R}}_{adj}$: $\{\hat{\lambda}_{T,k}\}_{k=1}^{K}$ and sort them in descending order.

4. Set $\epsilon_0 = \hat{\lambda}_{T,K}$.

5. Choose an appropriate threshold $\delta$:

   (a) Set index $i = K$ and $\delta = \hat{\lambda}_{T,K}$.

   (b) If condition $\hat{\lambda}_{T,i-1} - \hat{\lambda}_{T,i} < \epsilon_0/2$ holds, go to the next step; otherwise, output $\delta$ as the appropriate threshold.

   (c) Set $i = i-1$ and $\delta = \hat{\lambda}_{T,i}$. If $i < 2$, output the error information "The spiked eigenvalues model cannot be employed"; otherwise, go back to Step (b).

6. Obtain the spiked sample eigenvalues of $\widehat{\mathbf{R}}_{adj}$: $\{\hat{\lambda}_{T,k} : \hat{\lambda}_{T,k} > \delta, k = 1, ..., K\}$.

7. Set the iteration index $j = 1$ and an initial threshold $\epsilon = \epsilon_0$. Set index $l = 1$.

8. Calculate $J_l$, $k_l$, $\hat{\lambda}_T^{(l)}$ and $\underline{s}_T^{(l)}$ through (11)- (13) in Section 2.2 of Yao et al. (2016) with the current threshold $\epsilon$. Set $k_l = k_{l-1} + |J_l|$. If $k_l < K$, set $l = l + 1$ and go back to the beginning of this step; otherwise, go to the next step.

9. Set $L = l - 1$ and $\tilde{\lambda}_i^{(j)} := -1/\underline{s}_T^{(i)}$, $i = \overline{1, L}$.

10. Generate $V$ random samples $\{\mathbf{Z}_i\}_{i=1}^{V}$ distributed with the mean vector $\bar{\mathbf{Z}}$ and covariance matrix $\widehat{\mathbf{V}}_K$ by the estimated spiked population eigenvalues $\{\tilde{\lambda}_l^{(j)}\}_{l=1}^{L}$, with multiplicity $J_l$.

11. Compute the relative discrepancy $\Delta_j = \|\frac{1}{V}\sum_{i=1}^{V}\mathbf{Z}_i(\delta,\epsilon) - \bar{\mathbf{Z}}\|_F / \|\bar{\mathbf{Z}}\|_F$.

12. If $\Delta_j < \Delta_{j-1}$, set $\tilde{L} = L$ and $\{\tilde{\lambda}_l\} = \{\hat{\lambda}_l^{(j)}\}$.

13. If $\epsilon > \hat{\lambda}_{T,1} - \hat{\lambda}_{T,K}$, output results; otherwise, update $\epsilon = \epsilon + \epsilon_0$, $j = j + 1$, set $l = 1$, and go back to Step 8.

**Output**: the ID (the number of sources) equal $\tilde{L}$ and $\{\tilde{\lambda}_i\}_{i=1}^{\tilde{L}}$ are the estimated population spiked eigenvalues.

## 3. COMPUTER SIMULATION

### 3.1 More experiments for the demonstration of Algorithms 1 – 3

In Section 2 of Yao et al. (2016), we demonstrate the Algorithms 1, 2 and 3 on a generated dataset. Here, more numerical experiments are presented to show the accuracy of the estimation. We generate four more datasets using the same setup in Section 2 of Yao et al. (2016). Figure 2 and Table 1 present the results on the four datasets. We can see that due to the randomness of the simulated data, the estimated spiked population eigenvalues are slightly different. However, the number of the estimated sources strictly equals four. Therefore, the number of the estimated spiked population eigenvalues can be considered invariant, and it is reasonable to use it as an estimate of the number of sources.

[Figure 1 about here.]

[Table 1 about here.]

### 3.2 A demonstration of Algorithms 1, 2 and 3 with 8 dipoles

We also consider an synthetic problem in Section 2 of Yao et al. (2016), with a larger number of dipoles. We first generate a random (Gaussian) sample matrix ($1000 \times 10000$, i.e., $K = 1000$ and $T = 10000$) with a mean zero and a covariance matrix

$$\mathbf{V}_{1000} = \mathrm{diag}(\lambda_1 I_{10}, \lambda_2 I_{10}, \lambda_3 I_{10}, \lambda_4 I_{10}, \lambda_5 I_{10}, \lambda_6 I_{10}, \lambda_7 I_{10}, \lambda_8 I_{30}, I_{900}), \tag{3.1}$$

where the $\lambda_1 = 20, \lambda_2 = 18, \lambda_3 = 16, \lambda_4 = 14, \lambda_5 = 12, \lambda_6 = 10, \lambda_7 = 8, \lambda_8 = 6$ are the eight true spiked population eigenvalues. Then, the eigenvalues of the sample covariance matrix $\mathbf{R}$ (in "$\circ$") are displayed in Figure 2, wherein the same figure, the exact spiked population eigenvalues (in "$\square$") and the estimated spiked population eigenvalues (in "$*$"), are also highlighted.

[Figure 2 about here.]

### 3.3 More experiments for the computer simulation in Section 3.1 of Yao et al. (2016)

In this subsection, we present more numerical results for the computer simulation, described in Section 3.1 of Yao et al. (2016). The same experiment has been repeated four times. The setup of the parameters is the same as that in Section 3.1. The results are summarized in Table 2,

where PCA, AIC, MDL and EIF give slightly different results for different experiments due to the randomness of the noise, whereas the SPE has been shown to be a more stable method.

[Table 2 about here.]

## 4. NOISE ESTIMATION

### 4.1 Noise Estimation by the Fourier Transform

The utility of the Fourier transform lies in its ability to analyze a signal in the time domain by its frequency content. In terms of our previous definition for $Y_j(t)$ (a sequence indexed by $t$, $j$) as a component-wise of $\mathbf{Y}(t)$ in (1) of Section 2, denote $\mathbf{Y}_j = \{Y_j(t), 1 \leq t \leq T\}$. The discrete Fourier transform (DFT) relates two finite sequences, $\mathbf{Y}_j$ and $\hat{C}_r$ for $r = 0, 1, \ldots, T-1$, defined by

$$\hat{C}_r = \frac{1}{T} \sum_{t=1}^{T} Y_j(t) e^{-i2\pi rt/T}.$$

If we further assume the noise is additive and independent, the variance of the all complex Fourier coefficients $C_r$ (based on the complex exponential basis functions) relates to the variance of noise $\sigma_j^2$ at sensor $j$

$$\mathrm{Var}(\hat{C}_r) = \frac{\sigma_j^2}{T},$$

where $1 \leq j \leq K$. Notice that this relation essentially allows for the noise variance estimation on the frequency domain by calculating the sample variance of the calculated Fourier coefficients.

Although the goal of the Fourier transform is to suppress the effect of signal while preserving the noise, the effect of signal on the calculated Fourier coefficients (part of the $\hat{C}_r$) is not always negligible. This is particularly true for MEG, as the signal tends to be stochastic and, thus, deviates from the common additive model. To avoid using Fourier coefficients that are largely affected by the signal, the median of the modulus of the Fourier coefficients is used instead of the sample variance of the coefficients. This is closely related to the median absolute operation, but is defined on the Fourier domain. Because the coefficients are Hermitian, we need only use the first half (or first quarter) of the Fourier coefficients (not including coefficients for $r = 0$) in the noise estimation. For each sensor $j$, if the complex coefficients $\hat{C}_r$ are used, the standard deviation of the noise is

estimated by

$$\hat{\sigma}_j \approx \nu \text{Med}(|\hat{C}_r|, 1 \leq r \leq T/2)/\Phi^{-1}(3/4) \tag{4.1}$$

where $|\hat{C}_r|$ is the modulus of $C_r$ and $\Phi^{-1}$ is the inverse of the Gaussian quartile function. We set the scale term $\nu = \sqrt{T/1.5247}$ and its approximation is guided by Gaussian heuristics (see the end of this subsection for details). Finally, we have the estimated noise covariance matrix $\widehat{\mathbf{R}}_n = \text{diag}((1/\hat{\sigma}_1)^2, (1/\hat{\sigma}_2)^2, ..., (1/\hat{\sigma}_K)^2)$.

**Determining the scale term in the noise estimation by using the Fourier basis**

In equation (4.1), when the complex coefficients $\hat{C}_r$ are used for the noise estimation, the scale term is chosen by $\nu = \sqrt{T/1.5247}$. To find an exact scale term in equation (4.1) is difficult, due to the median operation taken on $|\hat{C}_r|, 1 \leq r \leq T/2$. Since we prefer to use complex coefficients, the empirical scale term $\nu = \sqrt{T/1.5247}$ is derived by the following procedure:

1. Draw $10^5$ samples from $N(0, 1)$ to represent the data (only noise).

2. Set the standard deviation for the real part coefficients ($\sigma_r = \sqrt{0.5/10^5} = 0.0022$) and imaginary part coefficients ($\sigma_i = \sqrt{0.5/10^5} = 0.0022$).

3. Draw $10^5$ samples from $N(0, \sigma_r^2)$ to represent the real part coefficients and draw $10^5$ samples from $N(0, \sigma_i^2)$ to represent the imaginary part coefficients (the coefficients from the Fourier transform of $N(0, 1)$ are still normal but not $N(0, 1)$).

4. Calculate the modulus of the simulated coefficients, compute the median (0.0026) of them, and then divide by 0.6745 (median/0.6745 = 0.0039).

5. Compute the scale term from the known standard deviation of the data, that is $10^5/(1/0.0039)^2 \approx 1.5247$.

We used this scale term on all the simulated datasets in Section 3 when we knew the true noise variance. In all of these cases, the scale term worked very well. We also used this scale term on our real data in Section 4 of Yao et al. (2016).

### 4.2 Noise Estimation by Residual Analysis

We will improve the noise estimation method developed in Roger and Arnold (1996) based on residual analysis. The goal is to accommodate possible correlation between sensors. The decom-

position of the data covariance matrix $\mathbf{R}$ can be expressed as $\mathbf{R} = \mathbf{D}_K\mathbf{P}_K\mathbf{D}_K$, where $\mathbf{D}_K = \text{diag}(\sigma_1, \sigma_2, ..., \sigma_K)$ with $\{\sigma_j^2\}_{j=1}^K$ being diagonal elements of $\mathbf{R}$, and

$$\mathbf{P}_K = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} & \cdots & \rho_{1K} \\ \rho_{21} & 1 & \rho_{23} & \cdots & \rho_{2K} \\ \rho_{31} & \rho_{32} & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \rho_{(K-1)K} \\ \rho_{K1} & \rho_{K2} & \cdots & \rho_{K(K-1)} & 1 \end{bmatrix}$$

with $\rho_{mn}$ being the correlation coefficient at the $(m, n)$-th entry of $\mathbf{R}$. Similarly, the decomposition of the inverse, $\mathbf{R}^{-1}$, is $\mathbf{R}^{-1} = \mathbf{D}_{K^{-1}}\mathbf{P}_{K^{-1}}\mathbf{D}_{K^{-1}}$, where $\mathbf{D}_{K^{-1}} = \text{diag}(\varsigma_1, \varsigma_2, ..., \varsigma_L)$ with $\{\varsigma_j^2\}_{j=1}^K$ being diagonal elements of $\mathbf{R}^{-1}$, and

$$\mathbf{P}_{K^{-1}} = \begin{bmatrix} 1 & \xi_{12} & \xi_{13} & \cdots & \xi_{1K} \\ \xi_{21} & 1 & \xi_{23} & \cdots & \xi_{2K} \\ \xi_{31} & \xi_{32} & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \xi_{(K-1)K} \\ \xi_{K1} & \xi_{K2} & \cdots & \xi_{K(K-1)} & 1 \end{bmatrix}$$

with $\xi_{mn}$ being the correlation coefficient at the $(m, n)$-th entry of $\mathbf{R}^{-1}$. This method estimates the noise covariance matrix by $\widehat{\mathbf{R}}_n = \text{diag}((1/\varsigma_1)^2, (1/\varsigma_2)^2, ..., (1/\varsigma_K)^2)$, which is a diagonal matrix, and

$$\varsigma_j = \frac{1}{\sqrt{(\sigma_j)^2(1 - r_{K\backslash j}^2)}} \tag{4.2}$$

where $r_{K\backslash j}^2$ is the multiple correlation coefficients of sensor $\mathbf{Y}_j$ on the other $K-1$ sensors $\{\mathbf{Y}_k\}_{1\le k\le K, k\neq j}^L$ from the multiple regression theory. The advantage of using $\varsigma_j$ is that $\varsigma_j$ removes its correlation on the other $\varsigma_j$'s while $\sigma_j$ does not. However, one should use (4.2) only when there is dense correlation among sensors. In cases of sparse correlation, we modify (4.2) accordingly by

$$\tilde{\varsigma}_j = \frac{1}{\sqrt{(\sigma_j)^2(1 - \tilde{r}_{K^*\backslash j}^2)}},$$

where $\tilde{r}_{K^*\backslash j}^2$ is the multiple correlation coefficients of channel $\mathbf{Y}_j$ only on the other significant (not all) $K^*$ sensors.

### 4.3 Noise Estimation by thresholding method

A final option to be considered is the thresholding method. Covariance thresholding is used here to remove the effect of the signal before noise estimation. In contrast with the common entry-wise thresholding, whose goal is to remove most of the noise, the thresholded covariance matrix here would be obtained as an estimate of the noise covariance. We now describe the algorithm of the covariance thresholding: threshold the off-diagonal of $\mathbf{R}$ by

$$\eta(\mathbf{R})_{i,j} = \mathbf{R}_{i,j} I(\mathbf{R}_{i,j} \geq \tau_{K,T}), \quad 1 \leq i, j \leq T, i \neq j$$

where $I(\cdot)$ is the indicator function, $\tau_{K,T} = C\sqrt{T/\log K}$ and $C$ is a tuning parameter that varies case by case. This thresholding is to obtain possible correlated noise among sensors. Mainly in order to accommodate the low SNR ratio situation in our setting, we remove the high value of off-diagonals that mainly comes from the signal. With low SNR, it is reasonable to threshold 'at the noise level'. We follow the idea of Deshpande and Montanari (2016) and choose the $\tau_{K,T}$ inversely proportional to the SNR. For $i = j$, we remain the diagonal of $\mathbf{R}$ or estimate the diagonal of $\eta(\mathbf{R})$ by $\widehat{\mathbf{R}}_n$ from (4.1). The estimate of $\mathbf{R}_n$ is then given

$$\widehat{\mathbf{R}}_n = \eta(\mathbf{R}) + \mathrm{diag}((1/\hat{\sigma}_1)^2, (1/\hat{\sigma}_2)^2, ..., (1/\hat{\sigma}_K)^2).$$

In the simulation, we use $\tau_{K,T} = \lambda_{max}(\mathbf{R}) 10^{2/(1.1-\log(SNR))}$ for $SNR \leq 10$ (for $SNR > 10$ we do not need any error estimation). In practice, when the SNR is unknown, we set

$$\tau_{K,T} = \min\left\{\lambda_{max}(\mathbf{R})/10, \tau_s\right\},$$

where threshold $\tau_s$ is obtained by the following splitting technique in Bickel and Levina (2008). Split the sample randomly into two pieces, of size $T_1$ and $T_2$, where a choice that can be "justified" theoretically is $T_1 = T/\log(T)$, $T_2 = T - T_1$, and repeat this $V$ times ($V = \sqrt{T}$ in our experiments). Let $\mathbf{R}_1^v$ and $\mathbf{R}_2^v$ be the empirical covariance matrices based on the $T_1$ and $T_2$ observations, respectively, from the $v$-th split. Then

$$\tau_s = \underset{\tau}{\mathrm{argmin}} \frac{1}{V} \sum_{v=1}^{V} \|\eta_\tau(\mathbf{R}_1^v) - \mathbf{R}_2^v\|_2^2 .$$

REFERENCES


Bickel, P., and Levina, E. (2008), "Covariance regularization by thresholding," *Ann. Stat.*, 36, 2577–2604.

Brenner, D., Lipton, J., Kaufman, L., and Williamson, S. J. (1978), "Somatically evoked magnetic fields of the human brain," *Science*, 199, 81–83.

Dautray, R., and Lions, J.-L. (1998), *Mathematical Analysis and Numerical Methods for Science and Technology, Vol. 2, Functional and Variational Methods* Springer-Verlag , Berlin.

Deshpande, Y., and Montanari, A. (2016), "Sparse PCA via Covariance Thresholding," *J. Mach. Learn. Res.*, 17, 1–41.

Hamalainen, M. S., Hari, R., Ilmoniemi, R. J., Knuutila, J., and Lounasmaa, O. V. (1993), "Magnetoencephalography–theory, instrumentation, and applications to noninvasive studies of signal processing in the human brain," *Rev. Mod. Phys.*, 65, 413–497.

Kress, R., Kuhn, L., and Potthast, R. (2002), "Reconstruction of a current distribution from its magnetic field," *Inverse Probl.*, 18, 1127–1146.

Roger, R. E., and Arnold, J. F. (1996), "Reliably estimating the noise in AVIRIS hyperspectral images," *Int. J. Remote Sens.*, 17, 1951–1962.

Sarvas, J. (1984), "Basic mathematical and electromagnetic concepts of the biomagnetic inverse problem," *Phys. Med. Biol.*, 32, 11–22.

Tikhonov, A., Goncharsky, A., Stepanov, V., and Yagola, A. (1995), *Numerical Methods for the Solution of Ill-posed Problems* Dordrecht: Kluwer.

Tikhonov, A., Leonov, A., and Yagola, A. (1998), *Nonlinear Ill-posed Problems* London: Chapman and Hall.

Yao, Z., Zhang, Y., Bai, Z., and Eddy, W. F. (2016), "Estimating the number of sources in Magnetoencephalography using spiked population eigenvalues," *Manuscript*, Unpublished.

List of Figures

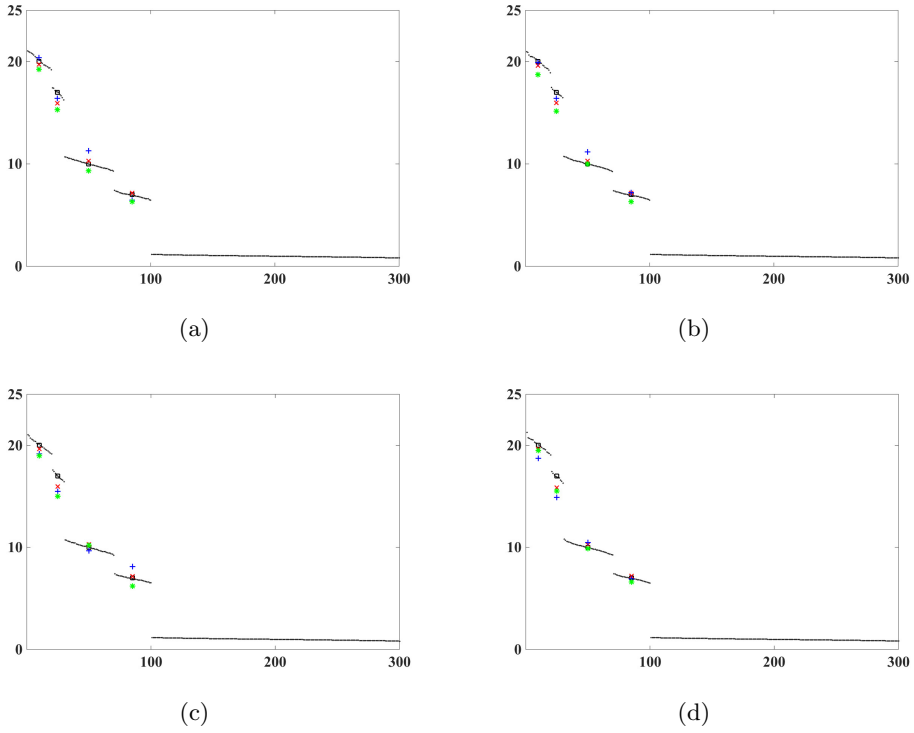
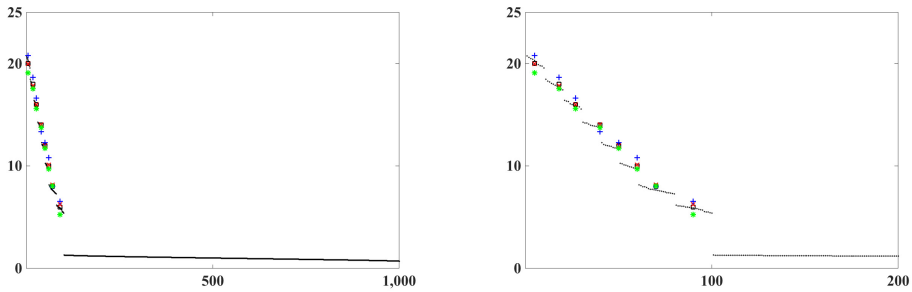



Figure 1: Four repetitions of experiments for the demonstration of Algorithms 1 – 3 with different distributions in Step 2 of Algorithm 2. "$\circ$" presents the sample eigenvalues and "$\square$" represents the true spiked population eigenvalues. "$\times$", "$+$" and "$*$" represent the estimated spiked population eigenvalues with normal distribution, uniform distribution and $t$-distribution, respectively.




(a) (b)

Figure 2: A demonstration of Algorithms 1 – 3 with 8 dipoles. "$\circ$"presents the sample eigenvalues and "$\square$" (black) represents the true spiked population eigenvalues. "$\times$", "$+$" and "$*$" are the estimated spiked population eigenvalues with normal distribution, uniform distribution and $t$-distribution in Step 2 of Algorithm 2. (a) The original plot. (b) The zoom-in plot.

List of Tables

Table 1: Estimated spiked population eigenvalues for the four repetitions of experiments for the demonstration of Algorithms 1 – 3 with different distributions in Step 2 of Algorithm 2. The true spiked population eigenvalues are $(20, 17, 10, 7)$.

| Trail | Gaussian | Uniform | $t$-distribution |
|---|---|---|---|
| 1 | $(19.73, 15.92, 10.28, 7.16)$ | $(20.39, 16.41, 11.27, 6.46)$ | $(19.25, 15.28, 9.34, 6.32)$ |
| 2 | $(19.60, 15.95, 10.29, 7.17)$ | $(19.88, 16.39, 11.18, 7.19)$ | $(18.71, 15.16, 10.00, 6.31)$ |
| 3 | $(19.64, 15.96, 10.28, 7.16)$ | $(19.16, 15.48, 9.66, 8.11)$ | $(18.98, 15.00, 10.22, 6.19)$ |
| 4 | $(19.63, 15.86, 10.26, 7.18)$ | $(18.73, 14.90, 10.48, 6.89)$ | $(19.48, 15.51, 9.87, 6.61)$ |

Table 2: Comparison of results from PCA, AIC, MDL and EIF and spiked population eigenvalues (SPE) with artificial data. The first column shows the labels of experiments with different signal-to-noise ratios. The second to sixth columns are the estimated number of dipoles from the simulation data where the true number of dipoles is four.

| Experiments (SNR=1) | PCA (0.9/0.8/0.7) | AIC | MDL | EIF | SPE (FFT) | SPE (RS) | SPE (TH) |
|---|---|---|---|---|---|---|---|
| 1 | 21/3/2 | 120-127 | 120-127 | 3-19 | 4 | 4 | 4 |
| 2 | 21/3/2 | 120-127 | 120-127 | 3-19 | 4 | 4 | 4 |
| 3 | 21/3/2 | 120-127 | 120-127 | 3-19 | 4 | 4 | 4 |
| 4 | 21/3/2 | 120-127 | 120-127 | 3-19 | 4 | 4 | 4 |
| Experiments (SNR=.1) | PCA (0.9/0.8/0.7) | AIC | MDL | EIF | SPE (FFT) | SPE (RS) | SPE (TH) |
| 1 | 69/48/34 | 122-127 | 120-127 | 2-11 | 4 | 4 | 4 |
| 2 | 69/48/34 | 122-127 | 121-127 | 1-11 | 4 | 4 | 4 |
| 3 | 68/47/34 | 122-127 | 120-127 | 2-11 | 4 | 4 | 4 |
| 4 | 69/48/34 | 122-127 | 120-127 | 2-11 | 4 | 4 | 4 |
| Experiments (SNR=.01) | PCA (0.9/0.8/0.7) | AIC | MDL | EIF | SPE (FFT) | SPE (RS) | SPE (TH) |
| 1 | 76/57/44 | 121-127 | 120-127 | 1-11 | 4 | 4 | 4 |
| 2 | 76/57/43 | 120-127 | 120-127 | 2-11 | 4 | 4 | 4 |
| 3 | 77/56/43 | 121-127 | 120-127 | 1-12 | 4 | 4 | 4 |
| 4 | 76/57/44 | 121-127 | 121-127 | 1-12 | 4 | 4 | 4 |
| Experiments (SNR=.001) | PCA (0.9/0.8/0.7) | AIC | MDL | EIF | SPE (FFT) | SPE (RS) | SPE (TH) |
| 1 | 78/59/46 | 120-127 | 120-127 | 1-11 | 4 | 4 | 4 |
| 2 | 79/56/47 | 120-127 | 120-127 | 1-10 | 4 | 4 | 4 |
| 3 | 78/59/46 | 120-127 | 120-127 | 1-11 | 4 | 4 | 4 |
| 4 | 78/59/46 | 121-127 | 120-127 | 1-11 | 4 | 4 | 4 |
| Experiments (SNR=.0001) | PCA (0.9/0.8/0.7) | AIC | MDL | EIF | SPE (FFT) | SPE (RS) | SPE (TH) |
| 1 | 78/58/46 | 120-127 | 120-127 | 1-12 | 4 | 4 | 4 |
| 2 | 78/59/46 | 120-127 | 121-127 | 1-12 | 4 | 4 | 4 |
| 3 | 77/58/46 | 122-127 | 120-127 | 1-13 | 4 | 4 | 4 |
| 4 | 78/59/46 | 121-127 | 120-127 | 1-12 | 4 | 4 | 4 |