On Nonparametric Feature Filters in Electromagnetic Imaging

Jian Zhang[1]

School of Mathematics, Statistics and Actuarial Science, University of Kent, Canterbury, Kent CT2 7NF, UK.

Abstract

Estimation of sparse time-varying coefficients on the basis of time-dependent observations is one of the most challenging problems in statistics. Our study was mainly motivated from magnetoencephalographic neuroimaging, where we want to identify neural activities using the magnetoencephalographic sensor measurements outside the brain. The problem is ill-posed since the observed magnetic field could result from an infinite number of possible neuronal sources. The so-called minimum-variance beamformer is one of data-adaptive nonparametric feature filters to address the above problem in the literature. In this paper, we propose a method of sure feature filtering for a high-dimensional time-varying coefficient model. The new method assumes that the correlation structure of the sensor measurements can be well represented by a set of non-orthogonal variance-covariance components. We develop a theory on the sure screening property of the proposed filters and on when the beamformer-based location estimators are consistent or inconsistent with the true ones. We also derive the lower and upper bounds for the mean filtering errors of the proposed method. The new theory is further supported by simulations and a real data analysis.

Keywords: MEG neuroimaging, beamforming, nonparametric feature selection, source screening and reconstruction

[1]Email: jz79@kent.ac.uk

1. Introduction

Suppose that we observe an n-dimensional time-course $\mathbf{Y}(t)$ from the model

$$\mathbf{Y}(t) = \int_{\Xi} \mathbf{x}(r, \eta(r))\beta(r, t)dr + \varepsilon(t),$$

where Ξ is a bounded subset of R^3, $\beta(r, t)$ is a latent univariate time source of interest at location r, $\mathbf{x}(r, \eta(r))$ is a design vector with nuisance parameter $\eta(r)$, and $\varepsilon(t)$ is a noise. Assume that $\beta(r, t)$ is sparse, i.e., the temporal variability (called power or the marginal variance) $\text{var}(\beta(r, .)) = 0$ for all $r \in B$ except a few locations (i.e., non-null sources). We want to localize these non-null sources among an infinite number of candidates. Given the limited number of time-courses we observed, the problem is ill-posed and high-dimensional. To simplify it, we discretize the integration, obtaining

$$\mathbf{Y}(t) = \sum_{k=1}^{p} \mathbf{x}(r_k, \eta(r_k))\beta(r_k, t) + \varepsilon(t), \tag{1}$$

where $\Omega = \{r_1, ..., r_p\}$ is a sieve (or grid) approximation to the source space. The problem becomes a large-p-small-n problem. Several new methodologies have been developed for addressing large-p-small-n problems in regression settings, including least absolute shrinkage and selection operator (LASSO) (Tibshirani,1996), smoothly clipped absolute deviation (SCAD) (Fan and Li, 2001), and correlation screening (SIS) (Fan and Lv, 2008). Many important theoretical results have also been established recently in selection consistency (e.g.,Meinshausen and Bühmann, 2006; Zhao and Yu, 2006; Zhang, 2010). However, all these works focused on finite dimensional features and are therefore not directly applicable to neuroimaging studies, where features are time series.

In this paper, we propose a nonparametric feature filtering procedure for identifying the sparse coefficients. The proposed procedure is general but was initially designed for magnetoencephalography (MEG) neuroimaging. MEG is a technique for mapping brain activity by measuring magnetic fields produced by electrical currents occurring in the brain, using arrays of superconducting quantum interference sensors (Hamalainen et al., 2010). The MEG neuroimaging

can be employed to study perceptual and cognitive brain processes, to localize regions affected by pathology, and to determine the function of various parts of the brain. While MEG offers a direct measurement of neural activity with very high temporal resolution, its spatial resolution is relatively low. Concerns over its spatial resolution have raised fundamental issues of methodology and theory. In fact, improving its resolution by virtue of source reconstruction lies at the heart of the entire MEG-based brain mapping enterprise.

In the MEG neuroimaging, $Y_i(t_j)$ is the measurement recorded by the MEG sensor i at time t_j for $1 \le i \le n$, $1 \le j \le J$, where the time points $t_j = j/\Delta$, the number of the time instants $J = b\Delta$ is determined by the time window b and the sampling rate Δ per second, and the number of the sensors n is of order hundreds. Let $\mathbf{Y}(t_j)$ denote the measurements from all the sensors at time t_j, which are assumed to be induced by potential sources at locations $r_k, 1 \le k \le p$ along time-invariant orientations $\eta_k \equiv \eta(r_k, t), 1 \le k \le p$ in the brain respectively. Let $\beta(t_j) = (\beta(r_1, t_j), ..., \beta(r_p, t_j))^T$ be the source magnitude vector of these sources at time t_j and $\{\beta(r_k, t_j) : 1 \le j \le J\}$ the source time-course at r_k, where the superscript T indicates the matrix transpose. Let $\mathbf{x}_k$ be the output vector of these sensors that would be induced by a unit-magnitude source located at r_k along orientation η_k and $\mathbf{X} = (\mathbf{x}_1, ..., \mathbf{x}_p)$. As a special case of model (1), the sensor measurements $\mathbf{Y}(t_j)$ may be modeled as

$$\mathbf{Y}(t_j) = \sum_{k=1}^{p} \mathbf{x}_k \beta(r_k, t_j) + \varepsilon(t_j), \tag{2}$$

where $\mathbf{x}_k = l(r_k)\eta_k$, $l(r_k) \in R^{n\times 3}$ (called lead field matrix at location r_k) is derived from a forward physical model of the brain, and $\varepsilon(t_j)$ is the noise vector of the n sensors at time t_j (Sarvas, 1987). To search for unknown sources, a neural activity index for each grid point in the sieve is calculated, creating a source map of brain activity. Important sources can be then identified by filtering the source map. The accuracy of the filtering depends on the sieve size p, and the spatial and temporal dimensions of the MEG measurements (i.e., the number of sensors and the number of time instants). In practice, the sieve size p is often set to a value much larger than n.

The minimum-variance beamforming, a data-adaptive filtering approach to the above source localization problem has been widely used in neuroimaging. In the approach, one scans the source space through a feature space with a series of filters; each is tailored to a particular area of feature (called pass-band) and resistant to confounding effects originating from other areas (called stop-band) (van Veen et al.,1997; Robinson and Vrba, 1998). The scalar minimum variance beamforming aims to estimate the source power at the location r_k by minimizing the sample variance of the projected data $w^T\mathbf{Y}(t_j), 1 \leq j \leq J$ with respect to the weighting vector w, subject to the constraint $w^T\mathbf{x}_k = 1$. In the scalar minimum-variance beamformer, the pass-band is defined by linearly weighting sensor arrays with the constraint $w^T\mathbf{x}_k = 1$, while the stop-band is realized via minimizing the variance of the projected data. The estimated power can be normalized to produce a power map over a given temporal window while the beamformer projected data can provide time course information at each source. We rank these candidate sources by their powers and filter out noisy ones by thresholding.

In recent years, a number of simulation studies and theoretical studies have been conducted to evaluate the performance of a beamformer (e.g., Brookes et al., 2008; Sekihara and Nagarajan, 2010). Despite of this, several issues remain to be addressed. Firstly, there is no rigorous statistical theory available to allow one to examine when the estimated source time-courses are consistent with the true ones. In particular, when there are multiple sources, the accuracy is compromised by confounding effects of multiple sources. It is natural to ask when a beamformer will breakdown in presence of multiple sources and how this effect is determined by the spatial and temporal dimensions of a beamformer. Secondly, it is largely unknown in the literature when the beamformer-based filtering procedure can recover the true sources with an overwhelming probability, although a sure filtering procedure for ordinary linear regression models has been developed by Fan and Lv (2008).

To address these issues, we propose a beamformer filtering procedure which is based on the thresholded sensor covariance estimator. The objective of the

procedure is to identify a set of sources from sparse source model (2), which have nonzero powers or power increases relative to a reference. We develop a sure filtering theory for the proposed procedure under certain conditions. We show that if the true sources are not too close to each other and if n and J are large enough, then these sources can be recovered in a probability tending to one. Furthermore, we provide mean error bounds for source localization, power estimation and time-course estimation in the procedure. We conduct simulation studies and a real data analysis to assess the performance of the proposed procedure.

The paper is organized as follows. The details of the new beamforming methodology are provided in Sections 2 and 4. The asymptotic properties of the proposed procedure are investigated in Section 3. The simulation results and a real MEG data analysis are presented in Section 5. The conclusions are made in Section 6. The proof of Theorem 1 is deferred to the Appendix. The part of numerical results and the proofs of the lemmas, the proposition and the other theorems can be found in the on-line supplementary material.

2. Methodology

In this section we propose a new filtering procedure which uses the thresholded sensor covariance estimator.

2.1. Estimation of sensor covariance matrix

Suppose that $(\mathbf{Y}(t_j) : 1 \le j \le J)$ and $\mathbf{X} = (\mathbf{x}_1, ..., \mathbf{x}_p)$ are randomly sampled from the model (2). Namely,

$$\mathbf{Y}(t_j) = \mathbf{X}\beta(t_j) + \varepsilon(t_j), \quad 1 \le j \le J,$$

where $\mathbf{x}_k = l(r_k)\eta_k$ is the composite lead field vector at location r_k along orientation η_k, $1 \le k \le p$.

To make the above model identifiable, we assume the following condition.

Condition (A1): The source processes $\{\beta(t)\}$ and the noise process $\{\varepsilon(t)\}$ are stationary with $E[\beta(t)] = E[\varepsilon(t)] = 0$. These two processes are uncorrelated

with each other. The sources $\{\beta_k(t)\}, 1 \leq k \leq p$ are also uncorrelated with each other.

Under model (2) and condition (A1), if the noises are uncorrelated across the sensors and white, then the sensor covariance matrix can be expressed as

$$C = \mathbf{X}\text{cov}(\beta(t_j))\mathbf{X}^T + \text{cov}(\varepsilon(t_j)) = \sum_{k=1}^{p} \gamma_k \mathbf{x}_k \mathbf{x}_k^T + \sigma_0^2 I_n,$$

where γ_k denotes the marginal variance (i.e., power) of the k-th time-source, σ_0^2 is the background noise level and I_n is an $n \times n$ identity matrix.

The sensor covariance is traditionally estimated by its sample version,

$$\hat{C} = (\hat{c}_{ij}) = \frac{1}{J} \sum_{j=1}^{J} (\mathbf{Y}(t_j) - \bar{\mathbf{Y}})(\mathbf{Y}(t_j) - \bar{\mathbf{Y}})^T,$$

where $\bar{\mathbf{Y}}$ is the sample mean of $\{\mathbf{Y}(t_j)\}$. It is known that the sample covariance is not a good estimator of the population covariance if its dimension n is large or if the sample covariance is degenerate Bickel and Levina (2008). In MEG neuroimaging, the sample covariance matrix can be nearly singular due to collinearity between nearby voxels, which can have serious effects on estimating the precision matrix used in the source reconstruction. Here, we apply the procedure of Bickel and Levina (2008) to estimate the sensor covariance C, which is given by

$$\hat{C}(\tau_{nJ}) = (\hat{c}_{ij}(\tau_{nJ})),$$

where $\hat{c}_{ij}(\tau_{nJ}) = \hat{c}_{ij} I(|\hat{c}_{ij}| \geq \tau_{nJ})$ and τ_{nJ} is a constant changing in n and J.

Bickel and Levina (2008) showed that for independent and identically distributed (IID) samples, if one chooses $\tau_{nJ} = O(\sqrt{\log(n)/J})$, then under certain regularity conditions the thresholded covariance estimator is consistent with the true one, with the convergence rate of $O_p(m_n \sqrt{\log(n)/J})$, where $m_n = \max_i \sum_{j=1}^{n} I(c_{ij} \neq 0)$. Zhang et al.(2014) extended the above result to non IID samples.

2.2. SAM index

Source localization can be realized in two steps: Step 1, we construct a sieve Φ by partitioning the brain into a regular three dimensional grid and calculate

the source power for each grid point. The size of the sieve is determined by the resolution of the grid. In practice, one often set the resolution level at 1 cm or 1 mm. These powers generate a power distribution overlaid on a structural image of the subject's brain, creating a brain power map. The beamforming method consider here is termed Synthetic Aperture Magnetometry (SAM) in the literature (Robinson and Vrba, 1998). Step 2, we identify significant sources by thresholding the map. The details are given as follows.

For any grid point r and orientation η, we define the scaled lead field vector $\mathbf{x} = \mathbf{x}(r,\eta) = l(r)\eta$, where $l(r)$ is the $n \times 3$ lead field matrix. Given η, we calculate the weighting vector $w(r,\eta)$ by minimizing the variance of $w^T\hat{C}(\tau_{nJ})w$ with respect to w, subject to $w^T\mathbf{x} = 1$. This gives rise to $w(r,\eta) = \hat{C}(\tau_{nJ})^{-1}\mathbf{x}/\mathbf{x}^T\hat{C}(\tau_{nJ})^{-1}\mathbf{x}$. The power $\hat{\gamma}(r)$ of the estimated source time-series $\{w(r,\eta)^T\mathbf{Y}(t_j) : 1 \le j \le J\}$ at r is equal to $1/\mathbf{x}^T\hat{C}(\tau_{nJ})^{-1}\mathbf{x}$. The orientation is then estimated by maximizing the signal-to-noise ratio

$$\frac{w(r,\eta)^T\hat{C}(\tau_{nJ})w(r,\eta)}{(\sigma_0^2 w(r,\eta)^T w(r,\eta))}$$

or equivalently by maximizing the normalized power

$$\text{SAM}(\mathbf{x}(r,\eta)) = \frac{w(r,\eta)^T\hat{C}(\tau_{nJ})w(r,\eta)}{w(r,\eta)^T w(r,\eta)} = \frac{(l(r)\eta)^T\hat{C}(\tau_{nJ})^{-1}l(r)\eta}{(l(r)\eta)^T\hat{C}(\tau_{nJ})^{-2}l(r)\eta}.$$

The above maximization can be done by solving a generalized eigenvalue problem: The optimal orientation $\hat{\eta}(r)$ is the eigenvector associated with the minimum non-zero eigenvalue of the matrix $l(r)^T\hat{C}(\tau_{nJ})^{-2}l(r)$ relative to $l(r)^T\hat{C}(\tau_{nJ})^{-1}l(r)$. Denote $\hat{\mathbf{x}}(r) = \mathbf{x}(r,\hat{\eta}(r))$. We call $\text{SAM}(r) = \text{SAM}(\hat{\mathbf{x}}(r))$ the SAM index of neural activity at r. When r is running over the grid points in the brain, $\text{SAM}(r)$ creates a neuronal power map that underlies measured magnetic fields. Zhang et al. (2014) proved that when the underlying true sources are separable in a sense, the SAM index can consistently estimate their powers and therefore identify the true sources from a large number of candidates.

By thresholding the above neuronal power map, we obtain an estimated source set, namely

$$D_n = \{r \in \Omega : \text{SAM}(r) \ge \hat{\sigma}_0^2(1 + h_n)\},$$

where $\hat{\sigma}_0^2$ is the estimated background noise level based on a pre-stimulus dataset and $1+h_n$ is a pre-selected positive constant or a value estimated from the data. The estimated source time-courses and powers are given by

$$\begin{aligned}\Theta_n &= \{\hat{\mathbf{x}}(r)^T\hat{C}(\tau_{nJ})^{-1}\mathbf{Y}(\cdot)/\hat{\mathbf{x}}(r)^T\hat{C}(\tau_{nJ})^{-1}\mathbf{x}(r) : r \in D_n\},\\ \Gamma_n &= \{\hat{\gamma}(r) : r \in D_n\},\end{aligned}$$

respectively. In this paper, we often let h_n increase to infinity with a rate slower than n, i.e., $h_n/n = o_p(1)$. The local peaks of the SAM index over D_n give location estimators of the underlying sources. In particular, we are interested in the global maximum of the SAM index, which produces a point estimator called the maximum location estimator for one of the underlying sources.

In practice, the MEG imaging is often run on a subject first without stimulus and then with stimulus. This allows us to calculate the sample covariance $\hat{C}$ for the MEG data with stimulus as well as the sample covariance $\hat{C}_0$ for the background noises. The latter can provide an estimator of the background noise level. To make the thresholded sample covariance to be convergent, Zhang et al. (2014) chose $\tau_{nJ} = c_0\hat{\sigma}_0^2\sqrt{\log(n)/J}$ with a tuning constant c_0 and threshold $\hat{C}$ by τ_{nJ}, where $\hat{\sigma}_0^2$ is the minimum diagonal element in $\hat{C}_0$. The corresponding SAM index is written as $\text{SAM}_{c_0}(r)$. Note that, when $c_0 \leq 0$, the proposed SAM procedure reduces to the standard SAM implemented in the software FieldTrip. For each value of c_0, we can apply the proposed SAM procedure to the data and obtain the maximum SAM index

$$\text{SAM}_{c_0} = \max\{\text{SAM}_{c_0}(r) : r \text{ in the sieve}\}. \tag{3}$$

In both simulations and a real data analysis, Zhang et al.(2014) showed that $c_0 \in D_0 = \{0, 0.5, 1, 1.5, 2\}$ had covered its useful range. The issue of how to choose the tuning constant c_0 has been addressed in Zhang et al.(2014), where they recommended choosing c_0 so that SAM_{c_0} in (3) attains the maximum. Bickel and Levina (2008) proposed a cross-validation approach for choosing thresholding level in context of covariance matrix estimation. Their method is not directly applicable to the optimization of the SAM mapping as the latter

involves not only covariance estimation but also other steps. In general, an optimal covariance estimator does not lead to an optimal power map. However, as shown in the following sections, the tuning constant c_0 has no effect on the asymptotic theory. We will not discuss the issue further and suppress symbol c_0 in $\text{SAM}_{c_0}(r)$ thereafter.

3. Filter theory

We present an asymptotic analysis for the proposed procedure when both n and J are tending to infinity. Throughout the paper, we denote by $||\mathbf{x}||$ the Euclidean norm of vector $\mathbf{x}$. For an $n \times n$ symmetric matrix M, we use $\lambda_1(M) \geq \cdots \geq \lambda_n(M)$ for its eigenvalues. Define the operator norm for $M = (m_{ij})_{n\times n}$ by $||M|| = \max_{1\leq j\leq p} |\lambda_j(M)|$. It is well-known that $||M|| \leq \max_i \sum_{j=1}^n |m_{ij}|$.

3.1. Reparametrization

To simplify the derivation, assuming that $l(r_k)\eta_k \neq 0, 1 \leq k \leq p$, we reparametrize the model (2) as follows:

$$\mathbf{Y}(t_j) = (\tilde{\mathbf{x}}_1, ..., \tilde{\mathbf{x}}_n) \begin{pmatrix} \tilde{\beta}_1(t_j) \\ \vdots \\ \tilde{\beta}_p(t_j) \end{pmatrix} + \varepsilon(t_j),$$

where $\tilde{\mathbf{x}}_k = \sqrt{n} l(r_k)\eta_k / ||l(r_k)\eta_k||$ and $\tilde{\beta}_k(t_j) = ||l(r_k)\eta_k|| \beta_k(t_j)/\sqrt{n}$. For the notation simplicity, we let $\mathbf{x}_k$ and $\beta_k(t_j)$ stand for $\tilde{\mathbf{x}}$ and $\tilde{\beta}_k(t_j)$ respectively, that is, let

$$\begin{aligned} \mathbf{x}_k &= \sqrt{n} l(r_k)\eta_k / ||l(r_k)\eta_k||, \\ E[\mathbf{Y}(t_j)] &= E[\beta(t_j)] = E[\varepsilon(t_j)] = 0, 1 \leq j \leq J, 1 \leq k \leq p \end{aligned}$$

in the model (2). The original time-course and power can be recovered by multiplying $\beta(r_k, t_j)$ by the scaling factor $\sqrt{n}/||l(r_k)\eta_k||$. However, the normalized power index $\text{SAM}(r)$ is invariant under this reparametrization. Note that in practice we often see that $||l(r_k)\eta_k||^2/n = \sum_{i=1}^n (l_i(r_k)\eta_k)^2/n$ is tending to a constant as n is large. See Zhang et al.(2014).

Assume that the sensor processes are ergodic. Then, the underlying sensor covariance matrix C_p can be written as $C_p = \sum_{k=1}^{p} \gamma_k \mathbf{x}_k \mathbf{x}_k^T + \sigma_0^2 I_n$. To build a brain map, we consider an arbitrary location and orientation (r, η) in the brain with $l(r)\eta \neq 0$. Let $\mathbf{x} = \mathbf{x}(r,\eta) = \sqrt{n} l(r)\eta / ||l(r)\eta||$ denote the scaled lead field vector at r along orientation η. For any two locations r and r_y, the lead field spatial coherence is defined by $\rho(\mathbf{x}(r,\eta), \mathbf{y}) = \mathbf{x}(r,\eta)^T \mathbf{y}/n = 1 - ||\mathbf{x}(r,\eta) - \mathbf{y}||^2/(2n)$. Note that $||x(r,\eta)|| = ||y|| = n$. We define the so-called lead field distance between $x(r,\eta)$ and y by $||\mathbf{x}(r,\eta) - \mathbf{y}||_n = ||\mathbf{x}(r,\eta) - \mathbf{y}||/\sqrt{n}$. When $\rho(\mathbf{x}(r,\eta), \mathbf{y}) \geq 0$, we have

$$\begin{aligned} \frac{1}{2}||\mathbf{x}(r,\eta) - \mathbf{y}||_n^2 &= 1 - \rho(\mathbf{x}(r,\eta), \mathbf{y}) \leq 1 - \rho(\mathbf{x}(r,\eta), \mathbf{y})^2 \\ &\leq 2(1 - \rho(\mathbf{x}(r,\eta), \mathbf{y})) = ||\mathbf{x}(r,\eta) - \mathbf{y}||_n^2. \end{aligned}$$

When $\rho(\mathbf{x}(r,\eta), \mathbf{y}) < 0$, the above inequalities still hold if we replace $\mathbf{y}$ by $-\mathbf{y}$. So $1 - \rho(\mathbf{x}(r,\eta), \mathbf{y})^2$ shows how close (r,η) is to (r_y, η_y) in terms of the lead field distances $||\mathbf{x}(r,\eta) - \mathbf{y}||_n$ and $||\mathbf{x}(r,\eta) + \mathbf{y}||_n$.

3.2. Identifiability

The source identifiability is mainly determined by the lead field matrix. A necessary condition for the source r_1 being identifiable is that the columns in the matrix are independent of each other (Zhang et al., 2014). That is, we assume the following condition.

Condition (A2): For any three different locations r, r_1 and r_2 in the brain, the columns in the matrix $(l(r), l(r_1), l(r_2))$ are linearly independent.

3.3. Convergence of sensor covariance estimator

Following Bickel and Levina (2008), Zhang et al.(2014) established the convergence rate of the thresholded sensor covariance estimator when both n and J are tending to infinity under the following two additional conditions.

Condition (A4): There exist positive constants κ_1 and τ_1 such that for any $u > 0$ and t,

$$\max_{1 \leq i \leq n} P(||\mathbf{y}_i(t)|| > u) \leq \exp(1 - \tau_1 u^{\kappa_1})$$

and $\max_{1\le i\le n} E||\mathbf{y}_i(t)||^2 < +\infty$.

In the second one, we assume that the sensor processes are strong mixing. Let $\mathcal{F}_{-\infty}^0$ and $\mathcal{F}_k^\infty$ denote the σ-algebras generated by $\{\mathbf{y}(t) : -\infty \le t \le 0\}$ and $\{\mathbf{y}(t) : t \ge k\}$ respectively. Define the mixing coefficient

$$\alpha(k) = \sup_{A\in\mathcal{F}_{-\infty}^0, B\in\mathcal{F}_k^\infty} |P(A)P(B) - P(AB)|.$$

The mixing coefficient $\alpha(k)$ quantifies the degree of the temporal dependence of the process $\{\mathbf{y}(t)\}$ at lag k. We assume that $\alpha(k)$ is decreasing exponentially fast as lag k is increasing.

Condition (A5): There exist positive constants κ_2 and τ_2 such that for $k \ge 0$,

$$\alpha(k) \le \exp(-\tau_2 k^{\kappa_2}).$$

Write $\tau_{nJ} = A\sqrt{\log(n)/J}$, where A is a constant. Let $\bar{y}_i$ be the sample mean of the i-th sensor as before and

$$\begin{aligned} C &= C_p = (c_{ij}), \quad m_n = \max_{1\le i\le n} \sum_{j=1}^{n} I(c_{ij} \ne 0), \\ \hat{c}_{ij} &= \frac{1}{J}\sum_{t=1}^{J}(y_i(t) - \bar{y}_i)(y_j(t) - \bar{y}_j), \\ \hat{C}(\tau_{nJ}) &= \hat{C}_p(\tau_{nJ}) = (\hat{c}_{ij} I(|\hat{c}_{ij}| \ge \tau_{nJ})), \end{aligned}$$

where $I(\cdot)$ is the indicator. Let $\kappa_3 = \max\{2(2/\kappa_1 + 1/\kappa_2) - 1, (4/3)(1/\kappa_1 + 1/\kappa_2) - 1/3, 1\}$. We adopt the following result from Zhang et al.(2014).

Proposition 3.1. *Under conditions (A1)∼(A5), if* $(\log(n))^{\kappa_3}/J = o(1)$ *and* $\tau_{nJ} m_n = o(1)$ *as* $n \to \infty$ *and* $J \to \infty$, *then* $||\hat{C}(\tau_{nJ})^{-1} - C^{-1}|| = O_p(m_n \tau_{nJ})$ *and* $||\hat{C}(\tau_{nJ})^{-2} - C^{-2}|| = O_p(m_n \tau_{nJ})$.

In practice, $E[\beta(t)] = 0$ may not be true even after a centralization. For example, $\beta(t)$ may have a nonlinear trend $\alpha(t)$, i.e., $\beta(t)$ can be expressed as $\alpha(t)$ plus a random process $\zeta(t)$ with $E[\zeta(t)] = 0$. In this situation, we modify condition (A1) as follows.

Condition (A1'): The process $\zeta(t)$ and the noise process $\varepsilon(t)$ are stationary. The noise process is temporally uncorrelated with $\alpha(t)$ and $\zeta(t)$. The components of $\alpha(t) - \sum_{j=1}^{J} \alpha(t_j)/J$ are orthogonal when J is large enough. The components of $\zeta(t) - \sum_{j=1}^{J} \zeta(t_j)/J$ are asymptotically uncorrelated as $J \to \infty$. Also $\alpha(t) - \sum_{j=1}^{J} \alpha(t_j)/J$ and $\zeta(t) - \sum_{j=1}^{J} \zeta(t_j)/J$ are asymptotically uncorrelated as $J \to \infty$. The limit $\lim_{J\to\infty} \sum_{j=1}^{J} (\alpha_k(t_j) - \bar{\alpha})^2/J$ exists, where $\alpha_k(t)$ is the k-th component of $\alpha(t)$.

We re-define the C matrix by letting $\gamma_k = \text{var}(\zeta(t)) + \lim_{J\to\infty} \sum_{j=1}^{J} (\alpha_k(t_j) - \bar{\alpha})^2/J$, where $\alpha_k(t)$ is the k-th component of $\alpha(t)$ and $\bar{\alpha} = \sum_{j=1}^{J} \alpha(t_j)/J$. Then, Proposition 3.1 still holds under conditions (A1') and (A2)$\sim$(A5).

3.4. Bounds for filtering errors

Under similar conditions to (A1)$\sim$(A5), where the true sources are well separated, Zhang et al. (2014) showed the following local consistency result for the SAM index: For non-null true sources,

$$\frac{1}{\mathbf{x}_j^T \hat{C}(\tau_{nJ})^{-1} \mathbf{x}_j} = \gamma_j + o_p(1), \quad \frac{\mathbf{x}_j^T \hat{C}(\tau_{nJ})^{-1} \mathbf{x}_j}{\mathbf{x}_j^T \hat{C}(\tau_{nJ})^{-2} \mathbf{x}_j} = \sigma_0^2 \left(n\gamma_j v_{1j} + 1 + 2\gamma_j v_{2j}\right) + o_p(1),$$

where v_{1j} and v_{2j} are constants depending on the lead field vectors. In contrast, for null sources,

$$\frac{1}{\mathbf{x}^T \hat{C}(\tau_{nJ})^{-1} \mathbf{x}} = o_p(1), \quad \frac{\mathbf{x}^T \hat{C}(\tau_{nJ})^{-1} \mathbf{x}}{\mathbf{x}^T \hat{C}(\tau_{nJ})^{-2} \mathbf{x}} = \sigma_0^2 + o_p(1),$$

where σ_0^2 is the background noise level. In the following, we will present a global screening properties for the SAM index.

Suppose that there are only two non-null sources among p potential sources in model (2), which are located at r_1 and r_2 with orientations η_1 and η_2 respectively. Let $\mathbf{x}_1 = l(r_1)\eta_1$, $\mathbf{x}_2 = l(r_2)\eta_2$, and $\psi_n = n(1 - \rho(\mathbf{x}_1, \mathbf{x}_2)^2)$. We assume that the two underlying sources are apart away from each other by a lead field distance of an order larger than $O(1/n)$ (i.e., $\psi_n \to \infty$ as n tends ∞). We choose the two-latent-source model for our study because it is more amenable to theoretical analysis and sharper statements are possible. Although the above

model is simple, it can provide insight into the more general setting when more than two latent sources exist.

For the simplicity of notations, we let Ω denote a bounded region of interest in the brain. Let $\mathbf{x} = \mathbf{x}(r, \eta)$ denote the scaled lead field vector at location r along orientation η and let $\Phi = \{\mathbf{x} \in B : ||\mathbf{x}|| = n\}$, where B is a set of all the scaled lead field vectors. Let $\hat{\mathbf{x}}(r) = \mathbf{x}(r, \hat{\eta}(r))$ as before. We introduce the following notations.

$$\begin{aligned}
\rho_{12} &= \rho(\mathbf{x}_1, \mathbf{x}_2), \quad \psi_n = n(1 - \rho_{12}^2), \\
\rho(\mathbf{x}, \mathbf{x}_1, \mathbf{x}_2) &= \rho(\mathbf{x}, \mathbf{x}_2) - \rho(\mathbf{x}, \mathbf{x}_1)\rho_{12}, \quad \rho(\mathbf{x}, \mathbf{x}_2, \mathbf{x}_1) = \rho(\mathbf{x}, \mathbf{x}_1) - \rho(\mathbf{x}, \mathbf{x}_2)\rho_{12}, \\
\delta_n(r|12) &= n\rho(\hat{\mathbf{x}}(r), \mathbf{x}_1, \mathbf{x}_2), \quad \delta_n(r|21) = n\rho(\hat{\mathbf{x}}(r), \mathbf{x}_2, \mathbf{x}_1), \\
\zeta_n(r|1) &= n(1 - \rho(\hat{\mathbf{x}}(r), \mathbf{x}_1)^2), \quad \zeta_n(r|2) = n(1 - \rho(\hat{\mathbf{x}}(r), \mathbf{x}_2)^2) \\
\kappa_n(r|12) &= \zeta_n(r|1) - \delta_n(r|12)^2/\psi_n, \quad \kappa_n(r|21) = \zeta_n(r|2) - \delta_n(r|21)^2/\psi_n, \\
\delta_n(\mathbf{x}|12) &= n\rho(\mathbf{x}, \mathbf{x}_1, \mathbf{x}_2), \quad \delta_n(\mathbf{x}|21) = n\rho(\mathbf{x}, \mathbf{x}_2, \mathbf{x}_1), \\
\zeta_n(\mathbf{x}|1) &= n(1 - \rho(\mathbf{x}, \mathbf{x}_1)^2), \quad \zeta_n(\mathbf{x}|2) = n(1 - \rho(\mathbf{x}, \mathbf{x}_2)^2) \\
\kappa_n(\mathbf{x}|12) &= \zeta_n(\mathbf{x}|1) - \delta_n(\mathbf{x}|12)^2/\psi_n, \quad \kappa_n(\mathbf{x}|21) = \zeta_n(\mathbf{x}|2) - \delta_n(\mathbf{x}|21)^2/\psi_n.
\end{aligned}$$

Note that $\kappa_n(r|12)$ gauges the closeness of $\mathbf{x}(r)$ to $\mathbf{x}_1$ adjusted by the interference from $\mathbf{x}_2$ and $\kappa_n(r|21)$ gauges the closeness of $\mathbf{x}(r)$ to $\mathbf{x}_2$ adjusted by the interference from $\mathbf{x}_1$. By definition, $\kappa_n(r|12) \to 0$ when $r \to r_1$ and $\kappa_n(r|21) \to 0$ when $r \to r_2$. By Lemma 6.1 in the Appendix, under condition (A2), $\kappa_n(r|12) \geq 0$ and $\kappa_n(r|21) \geq 0$. Therefore, $\delta_n(r|12) \leq (\zeta_n(r|1) + \psi_n)/2$ and $\delta_n(r|21) \leq (\zeta_n(r|2) + \psi_n)/2$. Here, the following regularity condition is imposed on the lead field matrix, which states that the lead field distance $\zeta_n(r|1)$ ($\zeta_n(r|2)$) and the adjusted lead field distance $\kappa_n(r|12)$ ($\kappa_n(r|21)$) are of the same rate of convergence as $n \to \infty$. For any constants $a_n \to \infty$ and $b_n = O(1)$, let

$$\begin{aligned}
\Phi_2 &= \{\mathbf{x} \in \Phi : ||\mathbf{x}|| = n, \zeta_n(\mathbf{x}|k) \geq a_n, k = 1, 2\}. \\
\Phi_{3|k} &= \{\mathbf{x} \in \Phi : ||\mathbf{x}|| = n, \psi_n\zeta_n(\mathbf{x}|k) \geq a_n, \zeta_n(\mathbf{x}|k) \leq b_n\}, k = 1, 2.
\end{aligned}$$

Condition (A6): If $\psi_n \to \infty$, then there is a constant $0 < c_{12} < 1$ such that

$$\sup_{\mathbf{x}\in\Phi_{3|1}} \kappa_n(\mathbf{x}|12)/\zeta_n(\mathbf{x}|1) \geq c_{12}.$$

If $\psi_n \to \infty$, then there is a constant $0 < c_{21} < 1$ such that

$$\sup_{\mathbf{x}\in\Phi_{3|2}} \kappa_n(\mathbf{x}|21)/\zeta_n(\mathbf{x}|2) \geq c_{21}.$$

If $\psi_n \to \infty$, then

$$\sup_{\mathbf{x}\in\Phi_2} \kappa_n(\mathbf{x}|12) \to \infty, \quad \sup_{\mathbf{x}\in\Phi_2} \kappa_n(\mathbf{x}|21) \to \infty.$$

For any positive constants $a_n \to \infty$ and $b_n = O(1)$ and for $k = 1, 2$, define

$$S_{1|k} = \{r \in \Omega : \psi_n\zeta_n(r|k) \leq a_n\}, \quad S_{2|k} = \{r \in \Omega : \psi_n\zeta_n(r|k) \leq b_n\},$$

which is an $O(1/(n\psi_n))$-lead field neighborhood of the source r_k. The following theorem implies that with an overwhelming probability, D_n includes $S_{2|k}, k = 1, 2$ as two sub-sets, while being hold within the $O(a_n/(n\psi_n))$-lead field neighborhoods of the true sources.

Theorem 3.1. *Under conditions (A1)∼(A6), if* $(\log(n))^{\kappa_3}/J = o(1)$, $n\sqrt{\log(n)/J} = o(1)$, $h_n/\psi_n = o_p(1)$, *and* $\psi_n/(h_n a_n) = o_p(1)$, *then as* $n \to \infty$ *and* $J \to \infty$,

$$P\left((S_{2|1} \cup S_{2|2}) \subseteq D_n \subseteq (S_{1|1} \cup S_{1|2})\right) \to 1,$$

and

$$SAM(r) = \begin{cases} \frac{\gamma_1\psi_n(1+O(1/\psi_n))}{1+\gamma_1^2(\psi_n\zeta_n(\mathbf{X}|1)-\delta_n(\mathbf{X}|12)^2)/\sigma_0^4}, & r \in S_{2|1}, \\ \frac{\gamma_2\psi_n(1+O(1/\psi_n))}{1+\gamma_2^2(\psi_n\zeta_n(\mathbf{X}|2)-\delta_n(\mathbf{X}|21)^2)/\sigma_0^4}, & r \in S_{2|2}. \end{cases}$$

Remark 3.1. *Theorem 3.1 suggests that by high enough thresholding (i.e., choosing* $h_n \to \infty$ *in probability), we are able to detect two contiguous groups of active locations,*$S_{2|1}$ *and* $S_{2|2}$, *where the SAM index asymptotically attains the local maximums at the true source locations. For all but a small set of* r *in* D_n, *we have the following inequalities* $O\left(1/(n\psi_n)\right) \leq \min\{|1 - \rho(\hat{\mathbf{x}}(r), \mathbf{x}_1)^2|, |1 - \rho(\hat{\mathbf{x}}(r), \mathbf{x}_2)^2|\} \leq O\left(a_n/(n\psi_n)\right)$ *for the lead field distance from* r *to* $\{r_1, r_2\}$.

We define the mean errors of the power and the time-course estimation by $\text{MER}(\Gamma_n) = \sup_{r\in D_n} \min\{|\hat{\gamma}(r) - \gamma_1|, |\hat{\gamma}(r) - \gamma_2|\}$ and $\text{MER}(\Theta_n) = \max\{||\theta - \beta_1|| \wedge ||\theta - \beta_2|| : \theta \in \Theta_n\}$ respectively, where $a \wedge b = \min\{a, b\}$, and

$$||\theta - \beta_k||^2 = \frac{1}{J}\sum_{j=1}^{J}(\theta(t_j) - \beta_k(t_j))^2, \quad k = 1, 2.$$

We have the following theorem on the mean errors.

Theorem 3.2. *Under the conditions in Theorem 3.1, we have*

$$\begin{aligned} MER(\Gamma_n) &= O_p\left(a_n/\psi_n + n\sqrt{\log(n)/J}\right), \\ MER(\Theta_n) &= O_p\left(\psi_n^{-1/2} + n\sqrt{\log(n)/J} + \sqrt{a_n/(n\psi_n)}\right). \end{aligned}$$

Remark 3.2. *If two underlying sources are well separated from each other in the sense that $1 - \rho_{12}^2 \geq A/n^\alpha$ with $0 \leq \alpha < 1$ and $A > 0$, and if letting $h_n = O_p(n^{1-\alpha}/(\log(\log(n)))$ and $a_n = n^\alpha(\log(\log(n)))^2$, then we have $h_n/\psi_n = o_p(1)$ and $\psi_n/(h_n a_n) = o_p(1)$. The above mean errors are bounded by $O_p\left((\log(\log(n)))^2/n^{1-2\alpha} + n\sqrt{\log(n)/J}\right)$ and $O_p\left(1/n^{(1-\alpha)/2} + n\sqrt{\log(n)/J}\right)$ respectively.*

Remark 3.3. *Note that the SAM index is asymptotically flat in the $O(1/(n\psi_n))$ neighborhoods of the true source locations. This suggests that the sieve grid points should be distributed with a spacing of order $O(1/(n\psi_n))$.*

We now turn to the case where two underlying sources are not well separated in the sense that $\psi_n = O(1)$. For any positive constants $b_n = O(1)$, define sets $S_{4|k} = \{r \in \Omega : \zeta_n(r|k) \leq b_n\}$ for $k = 1, 2$.

Theorem 3.3. *If $\psi_n \to \psi_0$ and $n\sqrt{\log(n)/J} = o(1)$ as n and J tend to infinity, then under the conditions in Theorem 3.1, uniformly for $r \in S_{4|1} \cup S_{4|2}$, we have*

$$\hat{\gamma}(r) = \begin{cases} u_n(r|1)^{-1}\left(\gamma_1 + \gamma_2\left(1 + \psi_n\gamma_1/\sigma_0^2\right)\right) \\ \quad \times(1 + o_p(1)), \ \textit{when } r \in S_{4|1}, \\ u_n(r|2)^{-1}\left(\gamma_2 + \gamma_1\left(1 + \psi_n\gamma_2/\sigma_0^2\right)\right) \\ \quad \times(1 + o_p(1)), \ \textit{when } r \in S_{4|2}, \end{cases}$$

$$
SAM(r) \;=\; \begin{cases} v_n(r|1)^{-1}u_n(r|1)\sigma_0^4\left(r_1 + r_2(1+r_1\psi_n/\sigma_0^2)\right) \\ \qquad \times(1+o_p(1)),\ \textit{when } r \in S_{4|1}, \\ v_n(r|2)^{-1}u_n(r|2)\sigma_0^4\left(r_2 + r_1(1+r_2\psi_n/\sigma_0^2)\right) \\ \qquad \times(1+o_p(1)),\ \textit{when } r \in S_{4|2}, \end{cases}
$$

where

$$
\begin{aligned}
u_n(r|1) &= 1 + \frac{\gamma_1\zeta_n(r|1)}{\sigma_0^2} + \frac{\gamma_2}{\sigma_0^2}\left(\zeta_n(r|1) + \psi_n - 2\delta_n(r|12)\right) \\
&\quad + \frac{\gamma_1\gamma_2}{\sigma_0^4}(\zeta_n(r|1)\psi_n - \delta_n^2(r|12)), \\
u_n(r|2) &= 1 + \frac{\gamma_2\zeta_n(r|2)}{\sigma_0^2} + \frac{\gamma_1}{\sigma_0^2}\left(\zeta_n(r|2) + \psi_n - 2\delta_n(r|21)\right) \\
&\quad + \frac{\gamma_1\gamma_2}{\sigma_0^4}(\zeta_n(r|2)\psi_n - \delta_n^2(r|21)). \\
v_n(r|1) &= \zeta_n(r|1)\left(r_1 + r_2(1+r_1\psi_n/\sigma_0^2)\right)^2 + r_2(1+r_1\delta_n(r|12)/\sigma_0^2) \\
&\quad \times\left(\psi_n r_2 - \delta_n(r|12)\left(2(r_1+r_2) + r_1r_2\psi_n/\sigma_0^2\right)\right), \\
v_n(r|2) &= \zeta_n(r|2)\left(r_2 + r_1(1+r_2\psi_n/\sigma_0^2)\right)^2 + r_1(1+r_2\delta_n(r|21)/\sigma_0^2) \\
&\quad \times\left(\psi_n r_1 - \delta_n(r|21)\left(2(r_1+r_2) + r_1r_2\psi_n/\sigma_0^2\right)\right).
\end{aligned}
$$

335 For any positive constants $a_n \to \infty$, define sets $S_{5|k} = \{r \in \Omega : \zeta_n(r|k) \le 1/a_n\}$.

Corollary 3.1. *If $\psi_n \to \psi_0$ and $n\sqrt{\log(n)/J} = o(1)$ as n and J tend to infinity, then uniformly for $r \in S_{5|1} \cup S_{5|2}$, we have*

$$
\hat{\gamma}(r) = \begin{cases} \left(\gamma_1 + \frac{\gamma_2}{1+\psi_0\gamma_2/\sigma_0^2}\right)(1+o_p(1)), & r \in S_{5|1} \\ \left(\gamma_2 + \frac{\gamma_1}{1+\psi_0\gamma_1/\sigma_0^2}\right)(1+o_p(1)), & r \in S_{5|2}. \end{cases}
$$

$$
SAM(r) \;=\; \begin{cases} \frac{\sigma_0^2}{\gamma_2^2\psi_0}(\sigma_0^2 + \gamma_2\psi_0)(\gamma_1 + \gamma_2(1+\gamma_1\psi_0/\sigma_0^2)) \\ \qquad \times(1+o_p(1)),\ \textit{when } r \in S_{5|1}, \\ \frac{\sigma_0^2}{\gamma_1^2\psi_0}(\sigma_0^2 + \gamma_1\psi_0)(\gamma_2 + \gamma_1(1+\gamma_2\psi_0/\sigma_0^2)) \\ \qquad \times(1+o_p(1)),\ \textit{when } r \in S_{5|2}. \end{cases}
$$

340 **Remark 3.4.** *Denote $\alpha^2 = \delta_n^2(r|12)/(\psi_n\zeta_n)$. In Figures 1 and 2 in the online supplemental material, we plotted the asymptotic SAM derived in Theorem 3.3*

against the asymptotic value of $\zeta_n(r|1)$ *for various combinations of* (ψ_0, α^2)*, where* $\psi_0 \in \{0, 0.5, 1, 2, 3, 4\}$*,* $\alpha^2 \in \{0, 0.0392,\ 0.1584, 0.3576,\ 0.4872, 0.6368,\ 0.8064, 0.9960\}$*. We assumed that there were two nonzero underlying sources with powers* $\gamma_1 = 2$ *and* $\gamma_2 = 3$ *respectively and that the noise level* $\sigma_0^2 = 1$*. Note that* ψ_0 *shows how close two underlying sources are to each other, while* α^2 *measures the degree of coherence of lead field vectors in the neighborhoods of the underlying sources. Let* $\hat{r}$ *denote a local peak location in the SAM curve. Let* ζ *be the limit of* ζ_n*. If the asymptotic SAM attains the peak at* $\zeta > 0$*, then* $\hat{r}$ *is asymptotically inconsistent with the true source location* r_1 *in the sense that the lead field discrepancy* $n(1 - \rho(\mathbf{x}(\hat{r}), \mathbf{x}_1)^2)$ *is not close to zero. Figures 1 and 2 in the online supplemental material indicate that the consistency of the SAM-based localization depends on the degree of the separateness between the underlying sources as well as the degree of the coherence among* $\mathbf{x}(\hat{r}), \mathbf{x}_1$*, and* $\mathbf{x}_2$*. For example, Figures 1 and 2 show that the SAM peak can occur at nonzero* ζ *if* $\psi_0 = 0.5$ *and* $\alpha^2 > 0.1584$*, or if* $\psi_0 = 1, \alpha^2 > 0.3576$*, or if* $\psi_0 = 2, \alpha^2 > 0.4872$*, or if* $\psi_0 = 3, \alpha^2 > 0.63668$*, or* $\psi_0 = 4, \alpha^2 > 0.8064$*. In another word, if the underlying sources are not well separated (i.e.,* $\psi_n \to \psi_0 > 0$*), the local peaks in the SAM map may not asymptotically occur at the true source locations. Therefore, in these cases, we are unable to localize these sources even after reducing the threshold* h_n *to a lower level. Furthermore, the above corollary shows that if* $\psi_n \to \psi_0 > 0$*, then the power estimators are always inconsistent with the true ones due to the signal cancellation between the sources, even when their true positions and orientations are known.*

Remark 3.5. *Theorems 3.1∼3.3 still hold if we replace condition (A1) by condition (A1').*

4. Choice of threshold in mapping

The main product of the proposed screening procedure is the estimated source set D_n, whose performance depends on the threshold $1 + h_n$. We choose the threshold based on the idea of clustering as follows.

It follows from Theorem 3.1 and Lemma 6.4 in the Appendix that if the underlying sources are well separated with moderate or large signal-to-noise ratios, then the values of $\text{SAM}(r)/\hat{\sigma}_0^2$ at non-source locations are likely to be wandering at around 1, whereas the values of $\text{SAM}(r)/\hat{\sigma}_0^2$ in neighborhoods of source locations are typically much larger than 1. Therefore, the values of $\text{SAM}(r)/\hat{\sigma}_0^2$ can be grouped into two general clusters, one of which is a noise cluster. The positions of the peaks in the clusters give a clue to the source locations. The lower bound of the $\text{SAM}(r)/\hat{\sigma}_0^2$ peaks in the non-noise clusters provides a natural estimate of the threshold value for the screening. In our implementation, we refine the above idea by taking into account the following structure of the SAM mapping, i.e., the grouping of the SAM values often be affected by the spatial coherence of grid points. In neuroimaging, it is a common practice to reduce the complexity of the problem by first partitioning the grid points into a number of transverse slices along the z-axis of the brain. Except the noise slices, the SAM values in each of brain slices except noisy ones often peak toward a source location. Therefore, these peak values can be used as summary statistics for the non-noise group. In practice, we need to calculate the peak values $s_1, \cdots, s_{k_0}$ for these slices. Sorting these values in descending order, we have order statistics $s_{(1)}, \cdots, s_{(k_0)}$. Plotting these ordered values against their indices $k = 1, \cdots, k_0$ gives rise to a scree plot, where the peaks are decreasing as one moves to the right. When the dropping rate changes, the curve in the plot has an elbow or a change-point. See Figure 3 in the online supplemental material for a typical scree plot in our simulations conducted in the next section. The peak values around the change-point can be used to estimate $1 + h_n$. For simplicity, we estimate the change-point by searching for k at which $s_{(k+1)} - s_{(k)}$ attains the maximum in our simulations. A more refined but time-consuming way to determine the change-point would be to employ the gap-statistic based procedure, where we could adjust for spatially varying baseline-effects by using some reference distribution derived from multiple trial information (Tibshinari et al., 2001). See the real data analysis in the next section for the details.

5. Numerical results

5.1. Simulation results

In this section, we examined the finite-sample performance of the proposed beamformer procedure under various scenarios via simulations, where $n\sqrt{\log(n)/J}$ is not small. We need more notations as follows. For any estimator $\hat{r}_1$ of source location r_1, we denote by $E|\hat{r}_1 - r_1|$ the mean localization bias, where $|\hat{r}_1 - r_1|$ is the L_1-distance between $\hat{r}_1$ and r_1. We denote by $\rho_{\max}$ the maximum coherence between two locations r_1 and r_2 with

$$\rho_{\max}(r_1, r_2) = \max_{||\eta_1||=1, ||\eta_2||=1} \frac{(l(r_1)\eta_1)^T l(r_2)\eta_2}{||l(r_1)\eta_1|| \cdot ||l(r_2)\eta_2||}.$$

We assessed the behavior of the beamformer-based maximum location estimator by its mean L_1-distance to the underlying sources. Similarly, the accuracy of the estimated source set D_n derived from the proposed screening can be measured by its closeness to the underlying sources, namely $d(D_n, \{r_1, r_2\}) = max_{r_0 \in \{r_1, r_2\}} \min_{r \in D_n} |r - r_0|$, and by its size. Good D_n requires a trade-off between the size and the closeness to the underlying sources. By using the simulations, we attempt to answer the following questions: (1) How does the tuning constant c_0 improve the performance of the SAM screening procedure? (2) Is the screening procedure a valuable complement to the SAM-based maximum location estimator? (3) To what extent will the performance of the proposed procedure deteriorate by source cancellations and correlations? (4) What is the performance of the proposed procedure when the assumptions we made in Theorems 3.1∼3 are invalid?

We simulated a 184-sensor MEG system (CTF/VSM) using a real adult human subject head shape which was approximated by multiple local spheres and downloaded at http://fieldtrip.fcdonders.nl/download. We constructed $p = 2905$ regular 3-D grid points of resolution 1 cm within the head. These candidate source positions were aligned with the axes of the head coordinate system. A lead field matrix L (184×8715) between the $n = 184$ sensors and the 2905 grid points was then calculated by using the software FieldTrip. Note that

the sieve size p is determined by the resolution. If we refine the sieve so that it is of resolution 1 mm, we can obtain 10 times more grid points. Then, the computation in beamforming is too time-consuming to be realized in an ordinary PC.

We assumed that there were two non-zero sources $\theta_1(t)$ and $\theta_2(t)$, which were located at $r_1 = (5, -2, 8)^T$ and $r_2 = (-5, -2, 8)$ in the auditory area with the dumping cosine patterns $\theta_1(t) = \eta_1\beta_1(t)$ and $\theta_2(t) = \eta_2\beta_2(t)$ respectively. The L_1 distance and the maximum coherence between the two sources are 10 cm and 0.3632 respectively. Here,

$$
\begin{aligned}
\eta_1 &= \left(\frac{10}{\sqrt{102}}, \frac{1}{\sqrt{102}}, \frac{1}{\sqrt{102}}\right)^T, \quad \eta_2 = (1, 0, 0), \\
a_1 &= \sqrt{102}, \quad a_2 = 8, \\
\beta_k(t)/a_k &= z_k(t) \\
&+ \begin{cases}
\exp\left(-\left(t - \frac{m_k}{m_0} - \pi\right)^2\right) \\
\times \left(1 - \frac{(t - m_k/m_0 - m_0/60000)^2}{(1 - 1/m_0 + m_0/60000)^2}\right) \\
\times \cos\left(2\pi f_k \left(t - \frac{m_k}{m_0}\right) - \pi\right), \\
\qquad \frac{m_k}{m_0} \le t \le 1 + \frac{(m_k - 1)}{m_0} \\
0, \qquad \text{Otherwise,}
\end{cases}
\end{aligned}
$$

where m_0, m_k and f_k are two factors related to time-shifts and frequencies of the cosine waves, and $z_k(t) \sim \text{AR}(1)$, i.e., $z_k(t) = 0.2z_k(t-1) + e(t)$, $\{e(t)\}$ is a white noise process with mean 0 and variance 0.1^2. Note that $E[\beta_k(t)] \neq 0$. Therefore, condition (A1) may not hold true. Let $J = 2m_0$. We considered three scenarios: (1) $(m_1, m_2) = m_0 \times (1/10, 1/2)$ and $(f_1, f_2) = (1, 1)$; (2) $(m_1, m_2) = m_0 \times (1/8, 1/4)$ and $(f_1, f_2) = (1, 3)$; (3) $(m_1, m_2) = m_0 \times (1/8, 1/8)$ and $(f_1, f_2) = (1, 1)$. We generated a pair of signals for each scenario with $m_0 = 240, 500, 1000$, and 1500 respectively. As examples, the pairs of signal curves in these scenarios for $m_0 = 500$ were plotted in Figure 3 in the online supplemental material. The temporal correlation coefficients between these paired curves are -0.462, 0.043, 0.706 respectively. These paired curves stand for the three cases, where the paired sources are negatively correlated, positively and weak

correlated, and positively and strong correlated respectively. As $n\sqrt{\log(n)/J}$ is not small in these scenarios, the conditions for Theorem 3.1 do not hold.

The sensor measurements at t follow the model

$$Y(t) = l(r_1)\eta_1\beta_1(t) + l(r_2)\eta_2\beta_2(t) + \varepsilon(t), \quad 0 \le t \le 1, \tag{4}$$

where $\varepsilon(t)$ is the sensor noise vector. Let the time window width $w = 2$. That is, the sensors were measured at the time instants $t_k = 2k/J, k = 0, 1, 2, ..., J-1$. The signal strength (SS) in the sensor space was defined by

$$\text{SS} = \sqrt{\sum_{k=0}^{J-1} \frac{||l(r_1)\eta_1\beta_1(t_k)||^2}{J}}.$$

For each k, we sampled N_{nk} from an n-dimensional standard Normal and set $\varepsilon(t_k) = \text{SS} \times N_{nk}/\sqrt{\text{SNR}}$. Here, we considered three values of SNR, 1.5625, 1, and 0.01, standing for the cases with moderate signals and weak signals respectively.

For each combination of SNR, (m_1, m_2) and (f_1, f_2), we independently generated 50 datasets of $\{(Y(t_k), 0 \le k \le J-1\}$ paired with $\{\varepsilon(t_k), 0 \le k \le J-1\}$ by (4). Here, we imitated the practical situation, where the MEG imaging was run on a subject first without stimulus and then with stimulus. The former provides an estimator of the background noise level. For each dataset, we calculated the sample covariance $\hat{C}$ of $\{(Y(t_k), 0 \le k \le J-1\}$ and the corresponding sample covariance $\hat{C}_0$ of the background noises. We set $\tau_{nJ} = c_0\hat{\sigma}_0^2\sqrt{\log(n)/J}$ with the tuning constant c_0 and thresholded $\hat{C}$ by τ_{nJ}, where $\hat{\sigma}_0^2$ is the minimum diagonal element in $\hat{C}_0$. We considered four values for c_0: $0, 0.5, 1.5,$ and 2. Note that, when $c_0 = 0$, the proposed SAM procedure reduces to the standard SAM implemented in the FieldTrip. For each value of c_0, we applied the proposed SAM procedure to the 50 datasets, where the threshold $1 + h_n$ is determined by the scree algorithm described in the previous section. For each dataset, we obtained the SAM-based maximum location estimate and selected the source set D_n by thresholding. We calculated the average minimum localization biases of the SAM-based maximum location estimates to r_1 and r_2 as

well as the standard errors over these 50 datasets. For each dataset, we also calculated the maximum L_1-distances from D_n to the two underlying sources r_1 and r_2 as well as the size of D_n, where the distance from D_n to r_k is defined by $d(D_n, r_k) = \min\{|r - r_k| : r \in D_n\}$. We calculated the averages and standard errors of these quantities over these 50 datasets respectively. Note that, when both the mean and variance of the maximum L_1-distance from D_n to the underlying sources are zero, D_n recovers all the true sources. By dividing the average size of D_n by the total number of candidate sources, 2905, we obtain the average proportion of the candidate sources being selected. The maximum L_1-distance from D_n to r_1 and r_2 measures the approximate coverage of the true sources, while the size of D_n implies the false discovery rate. So, given the distance from D_n to the true sources, the smaller size the better D_n is.

The results, summarized in Table 1 and Tables $1 \sim 2$ in the online supplemental material, show that: (1) Using the thresholded covariance, we can improve the performances of the maximum location estimator and the estimated source set slightly in terms of localization bias when signals are weak. In practice, the preferred values of c_0 may be lying between 0.5 and 1.5 for weak signals. (2) The performances were robust to the potential source correlations. (3) For strong or moderately strong sources, the estimated source set contained or was very close to the true sources most of the time as the maximum-minimum distances from the estimated source sets to the underlying two sources have an mean 0 and a variance 0 approximately. For weak sources, the estimated source set also had reasonable closeness to the true sources if choosing c_0 properly. For moderate source signals, the size of D_n is around 7, which is extremely tiny, compared to the number of the grids, 2905, indicating that the screening result is very accurate. However, the size of the estimated source set may increase to a few hundreds when the signal-to-noise ratio is low, reflecting that D_n contained many false discoveries. In Scenario 3 with SNR= 0.01, the SAM-based maximum location estimator has a L_1-bias ranging from 2 to 5, suggesting that it has missed the true sources on average. See Table 1 in the online supplemental material. Therefore, the screening procedure may be better than the

single-point estimator in identifying the underlying sources.

The results, displayed in Tables $3 \sim 6$ in the online supplemental material, demonstrate that if there is only one source, say $\beta_2(t)$ in the data, we can estimate the source very accurately either by using the maximum location estimator or by the screening. However, compared to the results in Table 1 and Tables $1 \sim 2$ in the online supplemental material, we can see that if we added another source $\beta_1(t)$ to the model, then the coherence between the lead field vectors can have a serious effect on the estimation of source $\beta_2(t)$. Sometimes, source $\beta_2(t)$ can be completely masked by the interference from source $\beta_1(t)$.

In summary, the simulation results suggest that:

- The source interference due to the lead field coherence has sever effects on recovering sources. In fact, as demonstrated in Table 1 and Tables $1 \sim 6$ in the online supplemental material, without the interference, we can accurately localize the source by using the SAM-based maximum location estimator or by using the screening. In particular, the size of the selected set D_n can be substantially inflated by source interference. In general, when the sample size J is larger than 1000, the screening procedure is extremely good at detecting sources with moderate or strong signals.

- The proposed screening procedure offers a better source estimation than a point estimator such as the maximum location estimator in the presence of multiple sources. Although the accuracy of the proposed source screening can be affected by the signal-to-noise ratio and by the degree of source interference, it is robust to the underlying source correlations. On average, the accuracy can be improved by choosing the tuning constant $c_0 = 0.5$, or 1, or 1.5, or 2 in the sensor covariance estimation when the SNR is low. However, when the SNR is not low, it may be better to choose $c_0 = 0$. The results also imply that there is not a universal choice of c_0. Rather than choosing a universal one, we should choose c_0 adaptive to the SAM indices over a range of c_0. Zhang et al.(2014) provided two of such selection procedures for the maximum location estimator.

Table 1: *Simulation results for the estimated source set D_n in the two-source setting.* The average maximum L_1-distance from D_n to the underlying sources, $\max\{d(D_n, r_k) : k = 1, 2\}$, and the average size of D_n over 50 repetitions are provided for Scenarios 1 to 3 when SNR= 0.01 and 1 respectively. The standard errors are given in the parentheses.

J	Scen.	$c_0 = 0$	$c_0 = 0.5$	$c_0 = 1$	$c_0 = 1.5$	$c_0 = 2$
		SNR=0.01				
		Average $\max_k d(D_n, r_k)$ (Std. error)				
480	1	1.76 (0.45)	2.56 (0.669)	2.12 (0.536)	6.04 (0.885)	11.6 (1.00)
1000	1	0.46 (0.25)	0.18 (0.16)	1.34 (0.46)	5.66 (0.67)	10.9 (0.77)
2000	1	0.04 (0.03)	0.02 (0.02)	3.3 (0.65)	9.04 (0.31)	9.26 (0.48)
3000	1	0.74 (0.34)	0.2 (0.18)	1.5 (0.5)	7.5 (0.53)	8.8 (0.43)
480	2	2.02 (0.43)	1.88 (0.40)	0.40 (0.16)	6.1 (0.92)	10.3 (0.86)
1000	2	1.06 (0.37)	0.66 (0.31)	2.14 (0.53)	4.2 (0.67)	10 (0.88)
2000	2	0.8 (0.33)	0.32 (0.22)	1.26 (0.42)	6.3 (0.63)	9.1 (0.68)
3000	2	0.36 (0.21)	0 (0)	1.14 (0.43)	7.3 (0.52)	9.02 (0.50)
480	3	1.98 (0.32)	2.16 (0.33)	1.44 (0.30)	2.54 (0.61)	11.2 (0.89)
1000	3	1.56 (0.26)	0.96 (0.17)	0.3 (0.11)	2.52 (0.48)	9.78 (0.87)
2000	3	1.12 (0.17)	0.54 (0.13)	0.36 (0.09)	2.36 (0.46)	8.5 (0.79)
3000	3	0.88 (0.13)	0.58 (0.12)	0.52 (0.17)	3.64 (0.54)	8.32 (0.76)
		Average size (Std. error)				
480	1	351 (73.7)	414 (86)	559 (97)	309 (75)	141 (54)
1000	1	281 (83)	449 (104)	305 (82)	138 (40)	62 (23)
2000	1	269 (76)	276 (77)	35 (5.7)	13 (1.2)	48 (20)
3000	1	197 (58)	235 (71)	53 (10)	19 (2.4)	21 (5)
480	2	311 (78)	553 (103)	707 (95)	278 (64)	144 (52)
1000	2	261 (72)	509 (109)	406 (90)	361 (76)	143 (51)
2000	2	120 (34)	281 (77)	228 (73)	117 (43)	68 (27)
3000	2	439 (100)	318 (87)	60 (10)	29 (7.5)	22 (5.3)
480	3	518 (91)	536 (95)	838 (130)	496 (82)	128 (40)
1000	3	419 (94)	631 (120)	1008 (123)	414 (77)	143 (39)
2000	3	303 (83)	574 (114)	490 (96)	260 (60)	134 (33)
3000	3	432 (101)	522 (110)	387 (81)	234 (66)	82 (18)
		SNR=1				
		Average $\max_k d(D_n, r_k)$ (Std. error)				
480	1	0 (0)	0 (0)	0.04 (0.028)	0.18 (0.18)	0.3 (0.204)
1000	1	0 (0)	0 (0)	0.02 (0.02)	0.04 (0.03)	0.18 (0.18)
2000	1	0.02 (0.02)	0.02 (0.02)	0.04 (0.028)	0.02 (0.02)	0 (0)
3000	1	0 (0)	0.06 (0.03)	0 (0)	0.02 (0.02)	0.02 (0.02)
480	2	0.06 (0.03)	0 (0)	0.04 (0.03)	0 (0)	0.2 (0.2)
1000	2	0.02 (0.02)	0 (0)	0 (0)	0 (0)	2 (0.53)
2000	2	0.06 (0.03)	0 (0)	0.02 (0.02)	0.06 (0.03)	0 (0)
3000	2	0.04 (0.03)	0.02 (0.02)	0.02 (0.02)	0.02 (0.02)	0 (0)
480	3	0.02 (0.02)	0.06 (0.03)	0.3 (0.18)	0.94 (0.31)	0.26 (0.18)
1000	3	0 (0)	0.02 (0.02)	0.08 (0.04)	0.08 (0.04)	0.48 (0.20)
2000	3	0.06 (0.03)	0.06 (0.03)	0.02 (0.02)	0.08 (0.04)	0.1 (0.04)
3000	3	0.02 (0.10)	0.10 (0.04)	0.02 (0.02)	0.06 (0.03)	0.02 (0.02)
		Average size (Std. error)				
480	1	7.4 (0.09)	7.3 (0.10)	6.92 (0.06)	5.1 (0.10)	4.5 (0.26)
1000	1	7.14 (0.08)	7.08 (0.08)	7 (0.06)	6.84 (0.05)	4.58 (0.15)
2000	1	7.26 (0.07)	7.30 (0.09)	7 (0.09)	6.88 (0.05)	6.92 (0.04)
3000	1	7.28 (0.08)	7.2 (0.09)	7.2 (0.09)	6.98 (0.05)	6.86 (0.05)
480	2	6.84 (0.08)	7 (0.05)	6.88 (0.07)	5.16 (0.14)	4.7 (0.13)
1000	2	6.96 (0.06)	6.98 (0.05)	6.94 (0.07)	6.9 (0.04)	4.58 (0.15)
2000	2	6.82 (0.05)	6.86 (0.05)	6.90 (0.05)	6.82 (0.08)	6.86 (0.05)
3000	2	6.82 (0.05)	6.82 (0.05)	6.86 (0.06)	6.82 (0.05)	6.9 (0.04)
480	3	6.62 (0.08)	6.72 (0.07)	6.36 (0.11)	17.38 (2.77)	8.3(1.18)
1000	3	6.86 (0.05)	6.80 (0.06)	6.76 (0.07)	6.90 (0.63)	14.1 (2.75)
2000	3	6.78 (0.08)	6.78 (0.06)	6.90 (0.05)	6.82 (0.06)	7.24 (1.17)
3000	3	6.82 (0.05)	6.74 (0.07)	6.8 (0.06)	6.72 (0.06)	6.7 (0.07)

5.2. Source analysis on a real MEG dataset

We applied the proposed methodology to a human MEG dataset provided by Professor Richard Henson from the MRC Cognition and Brain Sciences Unit Volunteer Panel (Henson et al., 2011). The study subject, a healthy young adult underwent the following face perception test which includes two different stimuli (faces and scrambled faces). A central fixation cross (presented for a random duration of 400 to 600 ms) was followed by a face or scrambled face (presented for a random duration of 800 to 1000 ms), followed by a central circles for 1700 ms. As soon as he saw a face or a scrambled face, the subject used either their left or right index finger to report whether he thought it was symmetrical or asymmetrical vertically through its center. There were 96 trials labeled as Face and 50 labeled as Scrambled Face. The MEG data were collected with a VectorView system, containing a magnetometer and two orthogonal, planar gradiometers located at each of 102 positions within a hemispherical array situated in a light, magnetically shielded room. The sample rate was set at 1100Hz. Here, we investigated the dataset recorded by the 102 magnetometer. We want to identify voxels which showed power increases for the faces relative to the scrambled faces.

We first created a grid system of 1 cm resolution with 1487 grid points, using the subject's anatomical magnetic resonance imaging (MRI) data. Then, we applied the neuroimaging software SPM8 to read and preprocess the recorded data, and to epoch and average the data over the trials for the face stimulus and the scrambled face stimulus respectively. This gives rise to 146 epochs of 700ms (770 time instants) with 200ms pre-stimulus and 500ms post-stimulus. For each of the two stimulus, we calculated the sample covariance $\hat{C}$ from the post-stimulus data and the noise covariance $\hat{C}_0$ from pre-stimulus data. The signal-to-noise ratios for both datasets are close to 1. We thresholded $\hat{C}$ by $c_0\hat{\sigma}_0^2\sqrt{\log(n)/J}$, where $n = 102$, $J = 551$, and $\hat{\sigma}_0^2$ is the minimum diagonal element in $\hat{C}_0$. We only reported the results when setting $c_0 = 0.5$, since the results are similar for other values of c_0.

After the pre-processing above, we performed the proposed SAM procedure

on the face dataset and the scrambled face dataset respectively and calculated the logarithm of the SAM index ratio (i.e., the log-contrast) for each of 1487 grid points. However, we decided not to apply our screening procedure directly to the above observed log-ratios, as we note that the baseline distribution at each grid point can spatially vary. To adjust for such a spatial heterogeneity, we first generated an empirical baseline distribution by randomly partitioning the 196 trials into two groups sized 96 and 50 respectively. We labeled the two groups by the faces and by the scrambled faces respectively. The above operation was repeated 80 times, producing 80 paired datasets. We then calculated the SAM log-ratios for these paired datasets respectively, which generated an empirical baseline distribution for each grid. In Figure 1, we plotted the observed log-contrasts against the baselines, suggesting a striking difference between them.

Finally, we adjusted the observed log-ratio for each grid by subtracting the corresponding empirical baseline mean. In Figure 1, we also plotted the adjusted log-ratios on 20 traverse slices and on the three orthogonal slices through the maximum location estimate. The plots suggest two regions of the SAM-index increasing, which are around the grid points $(0, 6, 6)$ and $(-2, 6, 5)$ respectively.

In Figure 1, the scree plot in the top-left clearly suggests that a change-point might have occurred at the ordered slice 5. Therefore, we used the peak 0.0698 on this slice to estimate the threshold for screening. The screening provides a source set containing 22 grid points, where the SAM index shows increases for the faces relative to the scrambled faces. The estimated time-courses at (0,6,6) and (-2,6,5) plotted in Figure 1 show large differential responses of the brain to the faces and the scrambled faces at these two positions. These results have not been found by Henson et al.(2011) via a parametric Bayesian approach.

6. Discussion and conclusion

The beamformer-based screening has been widely used by neuron-scientists for analyzing neuroimaging data in an ad hoc way (e.g., Quraan et al., 2011). In this paper, we have reformulated the problem as finding sparse coefficients

in a time-varying coefficient model. Therefore, the proposed procedure is immediately applicable to estimating sparse time-varying coefficient models. We have improved the commonly used screening procedure by adding a data-driven choice of threshold. We have proved the sure screening property for the proposed procedure under some regularity conditions. We have also provided asymptotic bounds for screening errors in the two-source setting when both the number of sensors and the sampling rate are tending to infinity. To assess the finite sample performance of the proposed screening procedure, we have conducted simulation studies. The simulation results have suggested that the proposed screening procedure is promising in detecting sources of moderate signal-to-noise ratios.

The aim of beamforming is to allow the neuronal signal of interest to pass through in certain source location and orientation, while suppressing noise or unwanted signal in other source location or orientation. In practice, the unwanted signals from other locations cannot be fully blocked. Therefore, the source interference may prevent one from localizing the underlying sources. The existing theoretical analysis on the above issue was based on the unrealistic assumption that the sensor population covariance matrix is known (e.g., Sekihara and Nagarajan, 2010). In general, there is lack of a statistically sounding theory which allows one to address: (1) how to conduct a sure screening on a beamformer map, (2) how the spatial and temporal dimensions determine the accuracy of the estimated sources derived from the screening. Here, we have offered a theoretical analysis based on the estimated sensor covariance matrix in which we have allowed both the number of sensors and the sampling rate to be varying. The new analysis has drawn a clear picture on how the spatial and temporal dimensions are related to the accuracy of the beamformer-based screening. We have illustrated the proposed procedure by using a real MEG neuroimage analysis. Two interesting theoretical topics remain untouched: one is on the asymptotic behavior of the proposed data-driven procedure for choosing the threshold and the other is on determining the number of sources in a neuroimage. However, extending the theory to cover these two topics is beyond the scope of the paper.

Appendix

In the Appendix, suppressing c_0 in $\text{SAM}_{c_0}(r)$, we let $\text{SAM}(r)$ denote $\text{SAM}(r)$. Note that c_0 will not affect the convergence rate of the index.

Lemma 6.1. *For any lead field vectors* $\mathbf{x}, \mathbf{x}_1,$ *and* $\mathbf{x}_2$, *we have*

$$(\rho(\mathbf{x}, \mathbf{x}_2) - \rho(\mathbf{x}, \mathbf{x}_1)\rho(\mathbf{x}_1, \mathbf{x}_2))^2 \leq (1 - \rho(\mathbf{x}, \mathbf{x}_1)^2)(1 - \rho(\mathbf{x}_1, \mathbf{x}_2)^2).$$

The two sides are equal if and only if $\mathbf{x}$ *is a linear combination of* $\mathbf{x}_1$ *and* $\mathbf{x}_2$.

In the following lemmas, for simplicity, let $\hat{C}_2$ denote $\hat{C}(\tau_{nJ})$, where $\tau_{nJ} = O(n\sqrt{\log(n)/J})$.

Lemma 6.2. *Under conditions (A1)∼(A5), if* $(\log(n))^{\kappa_3}/J = o(1)$ *and* $n\sqrt{\log(n)/J} = o(1)$, *then we have*

$$\left| \frac{\mathbf{x}^T \hat{C}_2^{-1} \mathbf{x}}{\mathbf{x}^T \hat{C}_2^{-2} \mathbf{x}} - \frac{\mathbf{x}^T C_2^{-1} \mathbf{x}}{\mathbf{x}^T C_2^{-2} \mathbf{x}} \right| = O_p(n\sqrt{\log(n)/J}),$$

n and J tend to infinity.

In the following lemma, for $k = 1, 2$, we investigate the behavior of the SAM index in an $O(n^{-1})$-neighborhood of r_k but not in an $O(1/(n\psi_n))$-lead field neighborhood of r_k. For this purpose, for any positive constants $a_n \to \infty$ and $b_n = O(1)$, define

$$\Phi_{11|k} = \{\mathbf{x} \in \Phi : \psi_n \zeta_n(\mathbf{x}|k) \geq a_n, \zeta_n(\mathbf{x}|k) \leq b_n\}, k = 1, 2.$$

Lemma 6.3. *Under conditions (A1)∼(A6), if* $\psi_n \to \infty$, *then*

$$\sup_{\mathbf{x} \in \Phi_{11|k}} \frac{\mathbf{x}^T C_2^{-1} \mathbf{x}}{\mathbf{x}^T C_2^{-2} \mathbf{x}} \leq \sigma_0^2 \left(1 + \frac{\sigma_0^2}{\gamma_k \zeta_n(\mathbf{x}|k)}\right) |O(1)|$$

as n and J tend to infinity.

For any constant $a_n \to \infty$, we consider a set outside neighborhoods of r_1 and r_2, defined by

$$\Phi_2 = \{\mathbf{x} \in \Phi : \zeta_n(\mathbf{x}|1) \geq a_n, \zeta_n(r|2) \geq a_n\}.$$

We have the following lemma.

Lemma 6.4. *Under conditions (A1)∼(A6), if $\psi_n \to \infty$, then uniformly for $\mathbf{x} \in \Phi_2$,*

$$\frac{\mathbf{x}^T C_2^{-1} \mathbf{x}}{\mathbf{x}^T C_2^{-2} \mathbf{x}} = \sigma_0^2 \left(1 + O\left(1 + O(\frac{1}{n} + \frac{1}{\psi_n^2})\right)\right),$$

as n and J tend to infinity.

To examine the behavior of the SAM index in the $O(1/n^2)$ neighborhood of r_k, for any positive constant $b_n = O(1)$, we consider the neighborhood of r_k defined by

$$\Phi_{3|k} = \{\mathbf{x} \in \Phi : \psi_n \zeta_n(\mathbf{x}|k) \leq b_n\}, k = 1, 2.$$

Lemma 6.5. *Under conditions (A1)∼(A6), if $\psi_n \to \infty$, then uniformly for $\mathbf{x} \in \Phi_{3|1}$,*

$$\frac{\mathbf{x}^T C_2^{-1} \mathbf{x}}{\mathbf{x}^T C_2^{-2} \mathbf{x}} = \frac{\gamma_1 \psi_n (1 + O(1/\psi_n))}{1 + \gamma_1^2 (\psi_n \zeta_n(\mathbf{x}|1) - \delta_n(\mathbf{x}|12)^2)/\sigma_0^4},$$

as n tends to infinity. The similar result holds uniformly for $\mathbf{x} \in \Phi_{3|2}$.

Proof of Theorem 3.1: Let

$$\begin{aligned}
S^c_{11|1} &= \{r \in \Omega : \psi_n \zeta_n(r|1) > a_n, \zeta_n(r|1) = O(1)\}, \\
S^c_{12|1} &= \{r \in \Omega : \psi_n \zeta_n(r|1) > a_n, \zeta_n(r|1) \to \infty\}, \\
S^c_{11|2} &= \{r \in \Omega : \psi_n \zeta_n(r|2) > a_n, \zeta_n(r|2) = O(1)\}, \\
S^c_{12|2} &= \{r \in \Omega : \psi_n \zeta_n(r|1) > a_n, \zeta_n(r|2) \to \infty\}.
\end{aligned}$$

Then

$$(S_{1|1} \cup S_{1|2})^c = S^c_{1|1} \cap S^c_{1|2} \subseteq S^c_{11|1} \cup S^c_{11|2} \cup \left(S^c_{12|1} \cap S^c_{12|2}\right).$$

In the following, we show that

$$P(D_n \cap S^c_{11|1} = \emptyset) \to 1, \quad P(D_n \cap S^c_{11|2} = \emptyset) \to 1,$$
$$P(D_n \cap S^c_{12|1} \cap S^c_{12|2} = \emptyset) \to 1,$$

respectively.

By Lemmas 6.2 and 6.3, we have for any constant c, with a probability tending to one, uniformly for $r \in S^c_{11|1}$,

$$\begin{aligned} \text{SAM}(r) &\leq \sigma_0^2 \left(1 + \frac{\sigma_0^2}{\gamma_1 \zeta_n(r)}\right) |O(1)| + c(n\sqrt{\log(n)/J}) \\ &\leq \sigma_0^2 \left(1 + \frac{\sigma_0^2 n}{\gamma_1 a_n}\right) |O(1)| + c(n\sqrt{\log(n)/J}), \end{aligned}$$

which is smaller than $\sigma_0^2(1 + h_n)$ by the definition. Therefore, $r \notin D_n$ when n and J are large enough. This implies $P(D_n \cap S^c_{11|1} = \emptyset) \to 1$. Similarly, we can show that $P(D_n \cap S^c_{11|2} = \emptyset) \to 1$.

By Lemmas 6.2 and 6.4, we have for any constant c, with a probability tending to one, uniformly for $r \in S^c_{12|1} \cap S^c_{12|2}$,

$$\begin{aligned} \text{SAM}(r) &\leq \sigma_0^2(1 + o(1)) + c(n\sqrt{\log(n)/J}) \\ &< \sigma_0^2(1 + h_n), \end{aligned}$$

as n and J tend to infinity. This yields

$$P(D_n \cap S^c_{12|1} \cap S^c_{12|2} = \emptyset) \to 1.$$

For $r \in S_{2|1}$, without loss of generality, we assume that $\psi_n \zeta_n(r) = \zeta + o(1)$ and $n\delta_n(r) = \delta + o(1)$. We have

$$|\rho(\hat{\mathbf{x}}, \mathbf{x}_2) - \rho_{12}| \leq \sqrt{2(1 - \rho(\hat{\mathbf{x}}(r), \mathbf{x}_1))} = O(1/n).$$

By Lemmas 6.2 and 6.5, for any constant $c > 0$, with a probability tending to one, uniformly for $r \in S_{2|1}$, we have

$$\text{SAM}(r) \geq \frac{\gamma_1 \psi_n}{1 + \zeta \gamma_1^2 / \sigma_0^4 + o(1)} + c(n\sqrt{\log(n)/J})$$

which is larger than $\sigma_0^2(1 + h_n)$ when $h_n/\psi_n = o_p(1)$. Therefore, $r \in D_n$. This yields $P(S_{2|1} \subseteq D_n) \to 1$. Similarly, we can show that $P(S_{2|2} \subseteq D_n) \to 1$.

The proof is completed.

Acknowledgements

We thank Professor Richard Henson for sharing with us his MEG neuroimaging data.

[1] Bickel, P. and Levina, E., 2008. Covariance regularization by thresholding. *Ann. Stat.*, **36**, 2577-2604.

[2] Brookes, M.J., Vrba,J., Robinson,S.E., Stevenson, C.M., Peters, A.M., Barnes, G.R., Hillebrand, A. and Morris,P.G.,2008. Optimizing experimental design for MEG beamformer imaging. *NeuroImage*, **39**, 1788-1802.

[3] Fan, J. and Li, R., 2001. *Variable selection via nonconcave penalised likelihood and its oracle properties. Jour Ameri Statist Assoc*, **96**, 1348-1360.

[4] Fan, J. and Lv, J., 2008. *Sure independence screening for ultra-high dimensional feature space (with discussions). J our Roy Statist Soc B* , **70**, 849-911.

[5] Hamalainen, M., Hari, R., Ilmoniemi,R.J., Knuutila, J. and Lounasmaa, O.V., 2010. Magnetoencephalography theory, instrumentation, and applications to noninvasive studies of the working human brain. *Rev. Modern Phys.*, **21**, 413-460.

[6] Henson, R.N., Wakeman,D.G., Litvak,V. and Friston,K,J., 2011. A parametric em- pirical Bayesian framework for the EEG/MEG inverse problem: generative models for multi-subject and multi-modal integration. Front. Hum. Neuroscience,**5**, 1-16.

[7] Meinshausen, N. and Bühlmann, P., 2006. *High dimensional graphs and variable selection with the Lasso. Ann. Statist.*, **34**, 1436-1462.

[8] Quraan, M.A., Moses,S.N., Hung,Y., Mills, T. and Taylor, M.J., 2011. Detection and localization of hippocampal activity using beamformers with MEG:

A detailed investigation using simulations and empirical data. Human Brain Mapping, **32**, 812- 827.

[9] Robinson, S. and Vrba, J., 1998. Functional neuroimaging by synthetic aperture magnetometry. In: *Recent Advances in Biomagnetism. Yoshimoto, T., Kotani, M., Kuriki, S. Karibe, H., Nakasato, N. (Eds.), Tohoku University Press,* Sendai, Japan, pp. 302-305.

[10] Sarvas, J., 1987. Basic mathematical and electromagnetic concepts of the biomagnetic inverse problem. *Phys. Med. Biol.*, **32**, 11-22.

[11] Sekihara, K. and Nagarajan, S.S., 2010. *Adaptive Spatial Filters for Electromagnetic Brain Imaging.* Springer-Verlag, Berlin.

[12] Tibshirani, R., 1996. *Regression shrinkage and selection via the Lasso. Jour Roy Statist Soc B*, **58**, 267-288.

[13] van Veen, B.D., van Drongelen, W., Yuchtman, M. and Suzuki, A., 1997. Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Trans. Biomed. Eng.*, **44**, 867-880.

[14] Tibshinari,R., Walther, J. and Hastie, T., 2001. Estimating the number of clusters in a data set via the gap statistic. *Jour. Roy. Stat. Soc. B*, **63**, 411-423.

[15] Zhang, C.H., 2010. *Nearly unbiased variable selection under minimax concave penalty. Ann Statist.*, **38**, 894-942.

[16] Zhang,J., Liu, C. and Green, G., 2014. Source localization with MEG data: A beamforming approach based on covariance thresholding. *Biometrics* , 70, 121-131.

[17] Zhao, P. and Yu, B., 2006. *On model selection consistency of LASSO. Jour Machine Learning Research*, **7**, 2541-2567.

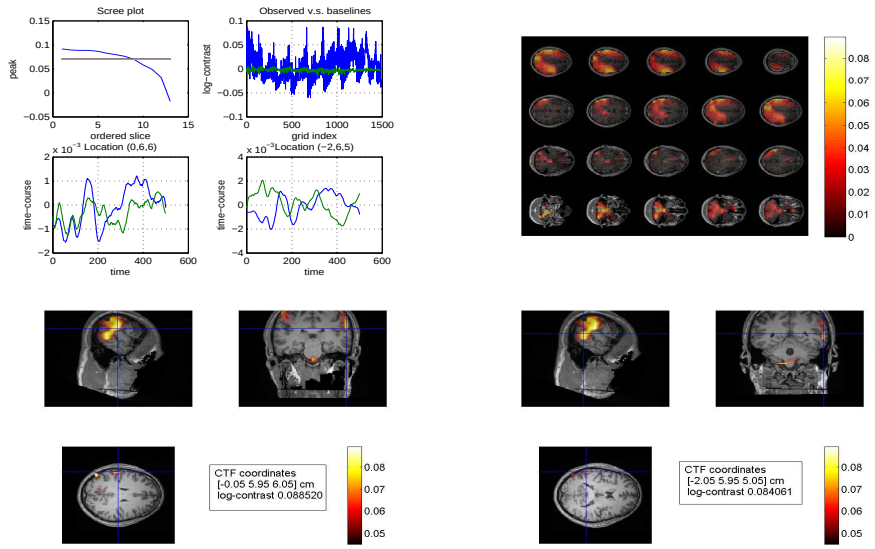

Figure 1: The top row in the left two columns: The plot in the left is a scree plot for the peaks of the adjusted SAM log-contrasts against the ordered slices, where the yellow-colored line indicates the estimated threshold level. The plot in the right provides a comparison between the observed log-contrasts and the baselines. The second row in the left two columns: The plots from left to right show the estimated time-courses at the two estimated source locations, where the face and the scrambled time-courses are colored by blue and green respectively. The remaining plots are derived from the adjusted SAM log-ratio-based mapping for the faces relative to the scrambled faces. The upright plots show the adjusted SAM log-ratios on 20 transverse slices evenly distributed from the top to bottom along the z-axis, where the upper, middle and bottom panels show the map on the slices along the top-to-bottom direction of the z-axis respectively. The (i, j, k) stands for the coordinates of grid point (r_x, r_y, r_z) in the CTF system of the brain. See the software FieldTrip for the definition. The bottom two rows: The adjusted SAM log-ratio-based maximum-mapping for scrambled faces relative to faces. The plots show the adjusted SAM log-ratios the three orthogonal slices through the locations (0,6,6) cm and (-2,6,5)cm respectively. In the plots, the adjusted log-contrasts are increasing in the color order from blue to light blue to green to light yellow to dark red.