Estimating Evoked Dipole Responses in Unknown Spatially Correlated Noise with EEG/MEG Arrays

Aleksandar Dogandžić, *Student Member, IEEE,* and Arye Nehorai, *Fellow, IEEE*

***Abstract*—We present maximum likelihood (ML) methods for estimating evoked dipole responses using electroencephalography (EEG) and magnetoencephalography (MEG) arrays, which allow for spatially correlated noise between sensors with unknown covariance. The electric source is modeled as a collection of current dipoles at fixed locations and the head as a spherical conductor. We permit the dipoles' moments to vary with time by modeling them as linear combinations of parametric or nonparametric basis functions. We estimate the dipoles' locations and moments and derive the Cramér–Rao bound for the unknown parameters. We also propose an ML-based method for scanning the brain response data, which can be used to initialize the multidimensional search required to obtain the true dipole location estimates. Numerical simulations demonstrate the performance of the proposed methods.**

***Index Terms*—Cramér–Rao bound, dipole source, EEG, evoked responses, maximum likelihood parameter estimation, MEG, sensor array processing, unknown noise covariance.**

I. Introduction

THE NONINVASIVE techniques of electroencephalography (EEG) and magnetoecephalography (MEG) are necessary for understanding both spatial and temporal behavior of the brain. Arrays of EEG and MEG sensors measure electric potential on the scalp and magnetic field around the head, respectively. These two fields are generated by neuronal activity in the brain and provide information about both its spatial distribution and temporal dynamics. This is in contrast with other brain imaging techniques that measure anatomical information (MRI, CT), blood flow or blood volume (fMRI, SPECT), or metabolism of oxygen or sugar (PET). Furthermore, the temporal resolution of EEG/MEG is far superior to that achieved by other modalities.

Spatiotemporal EEG/MEG data analysis is based on modeling a source of brain activity by a primary current distributed over a certain region of the cortex. Evoked responses are used to study sensory and cognitive processing in the brain [1] and are applied to clinical diagnosis in neurology and psychiatry. A current dipole is often used as an equivalent source for a unidirectional primary current that may extend over a few square centimeters of cortex. It is justified when the source dimensions are relatively small compared with the distances from the source to the measurement sensors [2] as is often satisfied for sources evoked in response to a given sensory stimulus: auditory, visual, etc.

In [3], spatiotemporal measurements are incorporated using the common dipoles-in-a-sphere model. The dipoles are assumed to have fixed locations and orientations, whereas their strengths are allowed to change in time according to a parametric model. De Munck [4] extends the above model by allowing the dipole strengths to change arbitrarily. In [5], only the dipole position is fixed, and the orientation and amplitude are allowed to vary in time according to a parametric model.

In all the above models, the noise is assumed to be spatially uncorrelated. As a result, these (and most other) localization procedures are based on minimizing a sum of squared errors. Such a residual function is appropriate only if the brain background noise, which is a major source of noise in EEG/MEG, shows no correlation across the scalp at different electrodes. However, since the background noise arises mostly in the cortex, it is expected to be strongly correlated in space. For example, regular rhythms in spontaneous brain activity, such as alpha waves, are not only large in amplitude but also correlated between neighboring sensors [6]. The correlated-noise problem is important in EEG because of the bipolar nature of the potential field recordings, i.e., the noise at the reference electrode spreads to all other channels [7]. In MEG, environmental noise is an additional important source of spatial correlation [8], especially in an unshielded environment.

One of the first attempts to tackle the problem of correlated noise was by Sekihara *et al.* [8], who assumed known spatial noise covariance. The localization in [8] is performed using a generalized least squares (GLS) method (see also Section III-B) and measurements at only one point in time. In [9], detection algorithms are derived for known spatial noise covariance and multiple time snapshots, whereas the temporal evolutions of the dipole moments are allowed to vary arbitrarily. Lütkenhöner has analyzed the GLS method for multiple time snapshots and applied it to both simulated and real data in [10] and [11]. The algorithm in [4] is extended in [12] and [13] to account for stationary noise correlated in both space and time. However, the noise covariance of such a process has an extremely large number of parameters that need to be determined. It is often estimated from the baseline measurements, i.e., data containing only noise collected before the stimulus is applied, assuming that, statistically, it does not differ between the baseline and a

Manuscript received June 4, 1998; revised July 26, 1999. This work was supported by the National Science Foundation under Grant MIP-9615590, the Air Force Office of Scientific Research under Grants F49620-97-1-0481 and F49620-99-1-0067, the Office of Naval Research under Grant N00014-98-1-0542, and the Aileen S. Andrew Foundation Graduate Fellowship. The associate editor coordinating the review of this paper and approving it for publication was Dr. Lal C. Godara.

The authors are with the Department of Electrical Engineering and Computer Science, University of Illinois at Chicago, Chicago, IL 60607 USA (e-mail: adogandz@eecs.uic.edu; nehorai@eecs.uic.edu).

Publisher Item Identifier S 1053-587X(00)00094-5.

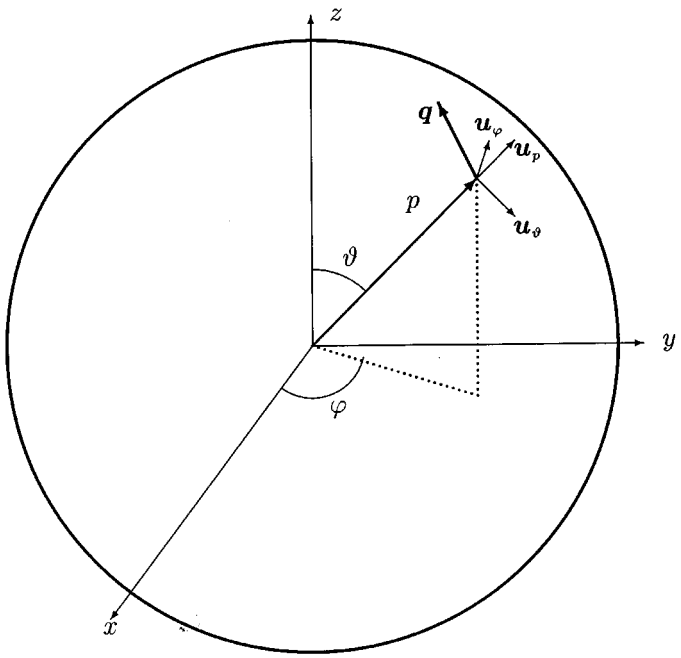


Fig. 1. Dipole in a sphere.

particular time point of interest. However, this method is suboptimal since it does not use the data containing the response for estimating the noise covariance. Furthermore, there are indications that utilizing baseline data may not be justified since it is hypothesized that the noise covariance changes due to dependence on the state of the subject or visual stimulation [7], [14], [15]. Thus, the noise covariance may need to be estimated only from the data containing the response. A major goal of this paper is to develop algorithms that solve this problem in an efficient way.

An iteratively reweighted generalized least squares (IRGLS) procedure [16, pp. 298–300] is proposed in [7] to estimate the noise covariance matrix and fit the dipole locations at a particular time point utilizing multiple trials. It is a two-stage procedure yielding estimates that, if the noise is Gaussian, converge to the maximum likelihood (ML) estimates. This method, however, does not include temporal evolution. In this paper (see also [17] and [18]), we allow temporal evolutions of all dipole moment components by modeling them as linear combinations of basis functions, assuming spatially correlated noise with unknown covariance. The basis functions may be chosen to exploit prior information on the temporal evolutions of the dipole moments, thus improving their estimation accuracy. If such information is not available, we propose and analyze a *nonparametric* basis functions method that exploits repeated trials and the linear dependence of the evoked responses on the dipole moments to estimate the basis functions. In this case, only the number of basis functions needs to be specified. This number is equal to the rank of the moment signal matrix, indicating the level of correlation between the moment components.

We first derive closed-form expressions for the ML estimates when the dipole locations and basis functions are known and then a concentrated likelihood function to be optimized when the dipole locations and basis functions are unknown (for references and relationship to previous work in statistics and signal processing, see Section III). Under statistical normality, this technique gives the ML estimates of the dipole locations and moments, requiring only a one-stage iterative procedure with computational complexity comparable with the ordinary least-squares (OLS) methods widely used in the EEG/MEG literature (see e.g., [3] and [4]). Both the ML and OLS methods are consistent if the noise is spatially correlated; however, we show that the ML is asymptotically more efficient. In Section IV, we derive the concentrated likelihood for the nonparametric basis functions which is a function only of the dipole locations and number of basis functions. Then, in Section V, we derive an ML-based method for scanning the brain response data, which can be used to initialize the multidimensional search required to obtain the true dipole location estimates.

We derive the Fisher information matrix (FIM) and Cramér–Rao bound (CRB) for the proposed model in Section VI (see also [18]) and discuss computational issues in Section VII. Finally, in Section VIII, numerical simulations are used to compare the estimation accuracy of the ML, GLS, OLS, and scanning methods.

II. Source and Measurement Models

A. Source Model

We model the head as a spherically symmetric conductor locally fitted to the head curvature. Let $\boldsymbol{p}$ be the position of a current dipole source relative to the center of the sphere

$$\boldsymbol{p} = p[\sin\vartheta\cos\varphi,\ \sin\vartheta\sin\varphi,\ \cos\vartheta]^T \tag{2.1}$$

where

- ϑ dipole's elevation;
- φ azimuth;
- p distance from the center; see Fig. 1.

Thus, $\boldsymbol{p}$ is fully described by $\boldsymbol{\theta} = [\vartheta,\ \varphi,\ p]^T$. The vectors

$$\begin{aligned} \boldsymbol{u}_\vartheta &= [\cos\vartheta\cos\varphi,\ \cos\vartheta\sin\varphi,\ -\sin\vartheta]^T \\ \boldsymbol{u}_\varphi &= [-\sin\varphi,\ \cos\varphi,\ 0]^T \\ \boldsymbol{u}_p &= \boldsymbol{p}/p \end{aligned} \tag{2.2}$$

form an orthonormal basis (see also [19]). Using this basis, the dipole moment can be written as $q_\vartheta\boldsymbol{u}_\vartheta + q_\varphi\boldsymbol{u}_\varphi + q_p\boldsymbol{u}_p$. We define the vector of moment parameters $\boldsymbol{q} = [q_\vartheta,\ q_\varphi,\ q_p]^T$.

B. Measurement Model

Consider a bimodal array of m_{E} EEG and m_{B} MEG sensors. The subscripts E and B refer to the EEG and MEG sensors, respectively. Let $m = m_{\mathrm{E}} + m_{\mathrm{B}}$. Then, the m-dimensional measurement vector of this array is

$$\boldsymbol{y} = A(\boldsymbol{\theta})\boldsymbol{q} + \boldsymbol{e} \tag{2.3}$$

where $\boldsymbol{y} = [\boldsymbol{y}_{\mathrm{B}}^T,\ \boldsymbol{y}_{\mathrm{E}}^T]^T$, $A(\boldsymbol{\theta})$ is the $m \times 3$ array response matrix, and $\boldsymbol{e} = [\boldsymbol{e}_{\mathrm{B}}^T,\ \boldsymbol{e}_{\mathrm{E}}^T]^T$ is additive noise. The array response matrix is derived using the quasistatic approximation of Maxwell's equations and spherical head model (see [2], [20], and references therein). The radial component of a dipole produces no external magnetic field in the spherical head model [2];

therefore, the last column of the MEG response matrix is zero. Thus, $A(\boldsymbol{\theta}) = [[A_{\mathrm{B}}(\boldsymbol{\theta}), \mathbf{0}_{m_{\mathrm{B}}\times 1}]^T, A_{\mathrm{E}}(\boldsymbol{\theta})^T]^T$, where $A_{\mathrm{B}}(\boldsymbol{\theta})$ and $A_{\mathrm{E}}(\boldsymbol{\theta})$ are the MEG and EEG response matrices with dimensions $m_{\mathrm{B}} \times 2$ and $m_{\mathrm{E}} \times 3$, respectively. The symbol $\mathbf{0}_{m_{\mathrm{B}}\times 1}$ denotes the $m_{\mathrm{B}} \times 1$ vector with zero entries. Define $r = \mathrm{rank}(A(\boldsymbol{\theta}))$; usually, $r = 3$, except when only MEG sensors are employed ($r = 2$).

For n distinct dipoles, (2.3) holds with $A(\boldsymbol{\theta})$, $\boldsymbol{\theta}$, and $\boldsymbol{q}$ substituted with $[A(\boldsymbol{\theta}_1)\cdots A(\boldsymbol{\theta}_n)]$, $\boldsymbol{\theta} = [\boldsymbol{\theta}_1^T \cdots \boldsymbol{\theta}_n^T]^T$, and $\boldsymbol{q} = [\boldsymbol{q}_1^T \cdots \boldsymbol{q}_n^T]^T$, respectively. Note that in this case, $\boldsymbol{\theta}$ and $\boldsymbol{q}$ are $3n$- and rn-dimensional vectors, respectively. Since the dipoles are at distinct locations, we assume that $A(\boldsymbol{\theta})$ has full rank equal to rn.

The noise vector $\boldsymbol{e}$ is assumed to be zero-mean with unknown spatial covariance Σ, whereas the source moment signal is deterministic. Thus, the mean and covariance matrix of the snapshot $\boldsymbol{y}$ are $A(\boldsymbol{\theta})\boldsymbol{q}$ and Σ, respectively. The noise is predominantly due to background activity in neurons. An assumption of Gaussianity, which is often used in EEG/MEG literature, may be justified by the additive nature of the noise and the large number of neurons normally active throughout the brain and has been validated in [21]. Tests for the normality of background EEG signals have also been developed in [22].

III. Maximum Likelihood Estimation

A. Simultaneous Estimation of the Dipole Parameters and Noise Covariance

We assume that the evoked field is a result of brain electrical activity that is well modeled by n dipoles at unknown fixed locations with time-varying moments. As is commonly done in analyzing evoked responses, the experiment is repeated K times to improve the signal-to-noise ratio (SNR). The activated dipoles are assumed to have the same locations and temporal patterns in each experiment, i.e., the evoked responses are *homogeneous* (this is a strong assumption that may need to be validated in practice. Homogeneity tests for the evoked responses have been derived in [23]). In the kth trial ($k = 1, \cdots, K$), N temporal data vectors (snapshots) $\boldsymbol{y}_k(1), \boldsymbol{y}_k(2), \cdots, \boldsymbol{y}_k(N)$ are collected. We refer to the matrix $Y_k = [\boldsymbol{y}_k(1)\cdots\boldsymbol{y}_k(N)]$ as the *spatiotemporal data matrix*. We assume that the temporal evolutions of the dipoles' moment components can be described by linear combinations of a set of basis functions $\boldsymbol{q}(t) = X\boldsymbol{\phi}(t, \boldsymbol{\eta})$, where X is a matrix of unknown coefficients with dimensions $rn \times l$ for the function representation described by the $l \times 1$ basis vectors $\boldsymbol{\phi}(t, \boldsymbol{\eta})$, and the parameter vector $\boldsymbol{\eta}$ is unknown in general. This parametrization allows us to exploit prior information on evoked response temporal evolutions and reduce the number of unknown parameters, thus improving the moment estimation accuracy. The measurement model is then

$$\boldsymbol{y}_k(t) = A(\boldsymbol{\theta})X\boldsymbol{\phi}(t, \boldsymbol{\eta}) + \boldsymbol{e}_k(t) \tag{3.1}$$

for $t = 1, \cdots, N$ and $k = 1, \cdots, K$. Here, $\boldsymbol{e}_k(t)$ denotes the noise, which is assumed to be zero mean with unknown spatial covariance Σ and uncorrelated in time and between trials. In reality, the noise is likely to be correlated in time (within a trial) but uncorrelated between trials. The noise covariance matrix Σ is assumed to be positive definite and constant in time and across all trials. If $K = 1$, and $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$ are known, the above model is known as the generalized multivariate analysis of variance (GMANOVA), which was first addressed in [24] (see also [25], [26, ch. 6.4], and [16, ch. 5]). In statistics, it is usually applied to fitting growth curves and thus is also called the growth-curve model [24]–[26].

Define $B(\boldsymbol{\eta}) = [\boldsymbol{\phi}(1, \boldsymbol{\eta})\cdots\boldsymbol{\phi}(N, \boldsymbol{\eta})]$ and $Q_B(\boldsymbol{\eta}) = B(\boldsymbol{\eta})^T[B(\boldsymbol{\eta})B(\boldsymbol{\eta})^T]^{-1}$. The projection matrix on the row space of B is then $P_B(\boldsymbol{\eta}) = Q_B(\boldsymbol{\eta})B(\boldsymbol{\eta})$. In Appendix A (see also [17] and [18]), we extend the GMANOVA equations to multiple trials, i.e., we show that for known $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$, if the noise $\boldsymbol{e}_k(t)$ is normal, the ML estimates of X and Σ are

$$\hat{X} = [A(\boldsymbol{\theta})^T S^{-1} A(\boldsymbol{\theta})]^{-1} A(\boldsymbol{\theta})^T S^{-1} \overline{Y} Q_B(\boldsymbol{\eta}) \tag{3.2a}$$

and

$$\hat{\Sigma}(\boldsymbol{\theta}, \boldsymbol{\eta}) = S + (I_m - TS^{-1})S_1(I_m - TS^{-1})^T \tag{3.2b}$$

where

$$\overline{Y} = \frac{1}{K}\sum_{k=1}^{K} Y_k \tag{3.3a}$$

$$S = \hat{R} - S_1 \tag{3.3b}$$

$$\hat{R} = \frac{1}{NK}\sum_{k=1}^{K} Y_k Y_k^T \tag{3.3c}$$

$$S_1 = \frac{1}{N}\overline{Y}P_B(\boldsymbol{\eta})\overline{Y}^T \tag{3.3d}$$

$$T = A(\boldsymbol{\theta})[A(\boldsymbol{\theta})^T S^{-1} A(\boldsymbol{\theta})]^{-1} A(\boldsymbol{\theta})^T \tag{3.3e}$$

and I_m denotes the identity matrix of size m. Note that S and S_1 are functions of $\boldsymbol{\eta}$ only, and T and $\hat{X}$ are functions of both $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$. To simplify the notation, we omit these dependencies throughout this paper. For the above model (and under the Gaussianity assumption), the sufficient statistics are $\overline{Y}$ and $\hat{R}$. If the matrices Y_k become scalars, i.e., $Y_k = y_k$ and $A(\boldsymbol{\theta}) = a$, we obtain the well–known results from univariate statistics $\hat{\Sigma} = S = 1/K\sum_{k=1}^{K}(y_k - \overline{y})^2$ and $\hat{X} = \overline{y}/a$.

If Σ is known, the ML estimate of X is simply

$$\hat{X} = \left[A(\boldsymbol{\theta})^T\Sigma^{-1}A(\boldsymbol{\theta})\right]^{-1} A(\boldsymbol{\theta})^T\Sigma^{-1}\overline{Y}Q_B(\boldsymbol{\eta}) \tag{3.4}$$

as can easily be shown by differentiating the log-likelihood function (see Appendix A) using the identity $(\partial/\partial X)\,\mathrm{tr}(P^T X Q) = PQ^T$ [27, p. 72].

If $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$ are not known (in addition to X and Σ), their ML estimates $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\eta}}$ are obtained by maximizing the concentrated likelihood function (see Appendix A)

$$l_{\mathrm{ML}}(\boldsymbol{\theta}, \boldsymbol{\eta}) = |\hat{\Sigma}(\boldsymbol{\theta}, \boldsymbol{\eta})|^{-NK/2} \tag{3.5}$$

where $\hat{\Sigma}(\boldsymbol{\theta}, \boldsymbol{\eta})$ is the ML estimate of Σ for known $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$, which is defined in (3.2b). To find the ML estimates of X and Σ, substitute $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$ in (3.2a) and (3.2b) by $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\eta}}$. If $\boldsymbol{q}(t)$ is modeled as a linear combination of l basis functions without any prior on their shape (i.e., nonparametric basis functions), we can concentrate (3.5) with respect to $\boldsymbol{\eta} = \mathrm{vec}(B^T)$ as well, as shown in

Section IV. Here, the vec operator stacks the columns of a matrix below one another into a single column vector.

The case of $K = 1$ and unknown $\boldsymbol{\theta}$ was first addressed in [28], where a concentrated likelihood function of a similar form was obtained and applied to direction-of-arrival (DOA) estimation. In addition, a signal subspace fitting (SSF) criterion approximating this likelihood function was proposed. Here (see also [17] and [18]), we consider a more general model with multiple trials (suitable for analyzing evoked responses), multirank (vector) source moment signals, and parametric and nonparametric basis functions. This formulation also includes as special cases the ML radar array processing methods in [29] and [30]. In Section VII, we show how to maximize the function in (3.5) with respect to both $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$ in a computationally efficient way. For a discussion on identifiability of the unknown parameters, see [18, App. B].

Modeling the dipole moments by linear combinations of parametric basis functions allows us to exploit prior information on the temporal evolutions of the evoked responses, which improves the moment estimation accuracy. Since the temporal evolutions can be described with a small number of parameters, the parametric basis function models may also be used as feature extractors in a pattern recognition scheme. However, their disadvantage is that the likelihood function often needs to be maximized with respect to the nonlinear basis parameters $\boldsymbol{\eta}$, in addition to the unknown dipole location parameters $\boldsymbol{\theta}$. This can be avoided by using nonparametric basis functions; see Section IV.

The above estimators have good asymptotic properties even when the noise is not Gaussian: Theorem 1 in the following section states that regardless of the noise distribution, the covariance of these estimators asymptotically achieves the CRB calculated under normality. In the sequel, we do not restrict ourselves to a particular distributional assumption, except when discussing the FIM and CRB, which is justified by the above comment. Thus, we slightly abuse the terminology by refering to these estimators as the ML estimators. Having a similar terminology problem, some authors refer to these methods as extended least squares (ELS).

B. Ordinary and Generalized Least Squares

The (nonlinear) ordinary least squares method [16, pp. 447–448] applied to the above model gives the residual sum of squares $\mathrm{tr}\{\hat{R} - (1/N)A(\boldsymbol{\theta})[A(\boldsymbol{\theta})^T A(\boldsymbol{\theta})]^{-1} A(\boldsymbol{\theta})^T \overline{Y} P_B(\boldsymbol{\eta}) \overline{Y}^T\}$ as a cost function to be minimized with respect to $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$. This expression is easily derived by substituting $\Sigma = \sigma^2 I_m$ and identities (3.4), (A.4a), and (A.4b) from Appendix A into the likelihood function in (A.1b); thus, the OLS is ML for Gaussian spatially uncorrelated noise. Obviously, the OLS method does not account for the spatial correlation in the noise covariance. Further, the OLS estimates are not based on the sufficient statistics since $\hat{R}$ does not affect the above minimization. If $K = 1$ and $B = I_N$, i.e., $\boldsymbol{q}(t)$ is an arbitrary vector at each time point $t = 1, \cdots, N$, this method coincides with the deterministic maximum likelihood in, e.g., [31].

The GLS method [16, pp. 448–449] is the ML method for spatially correlated Gaussian noise with known spatial covariance Σ. It is a simple extension of OLS since it reduces to applying OLS to the spatially prewhitened data. Detection methods for this case with $B = I_N$ are derived in [9].

C. ML Versus OLS

In this section, we show consistency and asymptotic normality of the ML estimates, as well as consistency of the OLS estimates. We then show that the ML estimates are asymptotically more efficient than the OLS estimates.

Define $\boldsymbol{\rho} = [\mathrm{vec}(X)^T, \boldsymbol{\theta}^T, \boldsymbol{\eta}^T]^T$ and $\boldsymbol{\psi} = \mathrm{vech}(\Sigma)$. Here, the vech operator creates a single column vector by stacking the elements below the main diagonal columnwise.

The vector of all unknown parameters is $\boldsymbol{\gamma} = [\boldsymbol{\rho}^T, \boldsymbol{\psi}^T]^T$. Assume that the true values of the parameter vectors $\boldsymbol{\rho}$ and $\boldsymbol{\gamma}$ are $\boldsymbol{\rho}_0 = [\mathrm{vec}(X_0)^T, \boldsymbol{\theta}_0{}^T, \boldsymbol{\eta}_0{}^T]^T$ and $\boldsymbol{\gamma}_0 = [\boldsymbol{\rho}_0{}^T, \boldsymbol{\psi}_0{}^T]^T$, respectively.

Let $\mathcal{I}_{\mathrm{signal}}(\boldsymbol{\gamma})$ be the Fisher information matrix [31] of the signal parameters $\boldsymbol{\rho}$. The exact expression will be given in Section VI; see (6.2) and (6.3). To establish asymptotic properties of the ML and OLS methods, we need the following regularity conditions:

- R1) The parameter space of $\boldsymbol{\rho}$ is compact, and the true parameter value $\boldsymbol{\rho}_0$ is an interior point.
- R2) The noise vectors $\boldsymbol{e}_k(t), k = 1, \cdots, K, t = 1, \cdots, N$ are independent, identically distributed (i.i.d.) with zero mean and arbitrary positive-definite covariance Σ.
- R3) $A(\boldsymbol{\theta})$ and $B(\boldsymbol{\eta})$ are continuous and have continuous first and second partial derivatives with respect to $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$.
- R4) The matrix $\mathcal{I}_{\mathrm{signal}}(\boldsymbol{\gamma})$ is nonsingular.
- R5) $\mathrm{tr}\{\Sigma^{-1}[A(\boldsymbol{\theta})XB(\boldsymbol{\eta}) - A(\boldsymbol{\theta}_0)X_0B(\boldsymbol{\eta}_0)][A(\boldsymbol{\theta})XB(\boldsymbol{\eta}) - A(\boldsymbol{\theta}_0)X_0B(\boldsymbol{\eta}_0)]^T\} = 0$ if and only if $\boldsymbol{\rho} = \boldsymbol{\rho}_0$.

The regularity condition R5) is essentially an identifiability condition for $\boldsymbol{\rho}$, requiring uniqueness of the mean response corresponding to the true value of the parameter vector $\boldsymbol{\rho}_0$ (see also [18, App. B]). Observe that the above conditions do not require specific distributional assumptions on the noise vectors $\boldsymbol{e}_k(t)$.

Theorem 1: Under the regularity conditions R1)–R5), the ML estimate of $\boldsymbol{\rho}$ satisfies (as $K \to \infty$)

$$\hat{\boldsymbol{\rho}} \overset{\text{a.s.}}{\to} \boldsymbol{\rho}_0 \tag{3.6a}$$

$$\hat{\boldsymbol{\rho}} = \boldsymbol{\rho}_0 + O_{\mathrm{p}}(K^{-1/2}) \tag{3.6b}$$

and

$$\sqrt{NK}(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}_0) \overset{\mathrm{d}}{\to} \mathcal{N}(\boldsymbol{0}, NK\mathcal{I}_{\mathrm{signal}}(\boldsymbol{\gamma}_0)^{-1}) \tag{3.7}$$

where $\overset{\text{a.s.}}{\to}$ indicates almost sure convergence, and $\overset{\mathrm{d}}{\to}$ indicates convergence in distribution.

Proof: The proof follows from [33, ch. 5.6], [34], where it is shown for a more general case; see also [16, pp. 300–301], [28], and [35]. ■

Theorem 2: Under the regularity conditions R1)–R5) [where R4) and R5) should be checked using $\Sigma = I_m$ instead of the

actual noise covariance], the OLS estimates of $\boldsymbol{\rho}$ satisfy (as $K \to \infty$)

$$\hat{\boldsymbol{\rho}}_{\mathrm{OLS}} \overset{\text{a.s.}}{\to} \boldsymbol{\rho}_0 \tag{3.8a}$$

$$\hat{\boldsymbol{\rho}}_{\mathrm{OLS}} = \boldsymbol{\rho}_0 + \left[K \sum_{t=1}^{N} D(t, \boldsymbol{\rho}_0)^T D(t, \boldsymbol{\rho}_0)\right]^{-1} \times \sum_{k=1}^{K} \sum_{t=1}^{N} D(t, \boldsymbol{\rho}_0)^T \boldsymbol{e}_k(t) + o_{\mathrm{p}}(K^{-1/2}) \tag{3.8b}$$

$$= \boldsymbol{\rho}_0 + O_{\mathrm{p}}(K^{-1/2}) \tag{3.8c}$$

where

$$D(t, \boldsymbol{\rho}) = \frac{\partial (A(\boldsymbol{\theta}) X \phi(t, \boldsymbol{\eta}))}{\partial \boldsymbol{\rho}^T}. \tag{3.9}$$

Here, $K \sum_{t=1}^{N} D(t, \boldsymbol{\rho})^T D(t, \boldsymbol{\rho})$ equals $\mathcal{I}_{\mathrm{signal}}(\boldsymbol{\gamma})$ in (6.2) and (6.3) when $\Sigma = I_m$.

Proof: See [18, App. C]. ∎

If $\boldsymbol{e}_k(t)$ are i.i.d. normal, it can be shown that the ML estimate $\hat{\boldsymbol{\rho}}$ is asymptotically efficient in the sense of first-order efficiency [27, Sec. 5c.2 and 5f.2].

In Appendix B, we prove that under the above regularity conditions, the ML estimate of $\boldsymbol{\rho}$ is asymptotically more efficient than the OLS estimate, i.e., the difference between the asymptotic covariances of the ML and the OLS estimates is negative semidefinite, where equality is achieved if $\Sigma = \sigma^2 I_m$.

The superior asymptotic performance of the ML compared with the OLS can be explained by the fact that the OLS estimator does not utilize information contained in the second-order moment matrix $\hat{R}$; see also Section III-B.

IV. Nonparametric Basis Functions

In this section, we obtain the ML estimates of nonparametric basis functions (i.e., $\boldsymbol{\eta} = \mathrm{vec}(B^T)$ with certain identifiability constraints; see discussion below), where only their number l is specified; here, l equals the rank of the moment signal matrix and is a measure of the level of correlation between the moment components; see below. This method exploits multiple trials and the linearity of the dipole moments to compute closed-form solutions for the basis function estimates. As a result, the corresponding concentrated likelihood function becomes a function of $\boldsymbol{\theta}$ only. The disadvantage of this method is that $\boldsymbol{\eta}$ may contain a large number of parameters compared with a suitable nonlinear parametrization that may utilize prior information on the temporal evolutions and improve the estimation accuracy of the dipole moments. However, the use of nonparametric basis functions does not deteriorate the asymptotic accuracy of dipole location, as shown in Section VI. This result is also confirmed by simulation results; see Section VIII.

Note that the nonparametric basis functions require using multiple trials or known signal corrupted by noise (i.e., training data, e.g., baseline), or both, as shown below. Otherwise, the concentrated likelihood function would go to infinity; see Appendix C.

In Appendix C, we derive the concentrated likelihood function for the nonparametric basis functions in the form of a generalized likelihood ratio (GLR) test statistic [27, p. 418], [36] for testing H_0: $X = 0$ versus H_1: $X \neq 0$. First, the GLR is suitably rewritten [see (C.2)] and then maximized with respect to B using the Poincaré separation theorem [27, pp. 64–65]; see also Appendix C. The resulting GLR is given by the product of the l largest generalized eigenvalues of the matrices $I_N - (1/N)\overline{Y}^T W(\boldsymbol{\theta}) \overline{Y}$ and $I_N - (1/N)\overline{Y}^T \hat{R}^{-1} \overline{Y}$, where

$$W(\boldsymbol{\theta}) = \hat{R}^{-1} - \hat{R}^{-1} A(\boldsymbol{\theta}) \left[A(\boldsymbol{\theta})^T \hat{R}^{-1} A(\boldsymbol{\theta})\right]^{-1} \cdot A(\boldsymbol{\theta})^T \hat{R}^{-1}. \tag{4.1}$$

The rows of the ML estimate $\hat{B}$ are the corresponding generalized eigenvectors of the above two matrices. Note that assuming $\mathrm{rank}(A(\boldsymbol{\theta})) \leq N$ (which holds in most practical applications), there can be only $\mathrm{rank}(A(\boldsymbol{\theta})) = nr$ generalized eigenvalues greater than one (and the rest are equal to one); thus, $1 \leq l \leq nr$; see also (C.7) in Appendix C. If $l = 1$, all the components of the dipole moments have the same temporal evolution (up to a scaling factor), and thus, they are fully correlated. On the other hand, $l = nr$ allows as many basis functions as the number of moment components, in which case, the concentrated likelihood is simply $|I_N - (1/N)\overline{Y}^T W(\boldsymbol{\theta}) \overline{Y}| / |I_N - (1/N)\overline{Y}^T \hat{R}^{-1} \overline{Y}|$ (which follows from the fact that the determinant of a matrix equals to the product of its eigenvalues). Further, this expression is equal to the GLR for known basis functions in the form of Dirac pulses, i.e., $B = I_N$ [see (C.2) in Appendix C]. Thus, moment components can be completely uncorrelated. The choice of l allows us to specify the level of correlation between the moment components, ranging from fully correlated ($l = 1$) to uncorrelated ($l = nr$). This is a useful property since the sources of evoked responses are often correlated.

Unless suitably constrained, the ML estimates $\hat{B}$ are not unique. However, in Appendix C, we show that regardless of which ML estimate of B is chosen, the concentrated likelihood function and the estimated dipole moment temporal evolution $\hat{Q} = \hat{X}\hat{B}$ are unique. The orthonormal set of the ML estimates of the basis functions can be constructed from the above ML estimates as $\hat{B}_{\mathrm{orth}} = [\hat{B}\hat{B}^T]^{-1/2}\hat{B}$. Here, $H^{1/2}$ denotes a symmetric square root of a symmetric matrix H, and $H^{-1/2} = (H^{1/2})^{-1}$; this notation will be used throughout the paper.

Consider now the case where the data set of each trial contains a part with baseline data. Thus, $Y_k = [Y_{1k}, Y_{2k}]$, $k = 1, \cdots, K$, where Y_{1k} is a spatiotemporal data matrix of size $m \times N_1$ containing the background noise only, whereas Y_{2k} is of size $m \times N_2$ containing the evoked response modeled as $A(\boldsymbol{\theta})XB_2$ corrupted by noise. The statistical properties of the noise, which are described in Section III, are assumed to be the same for both Y_{1k} and Y_{2k}. Thus, $B = [0, B_2]$, $N = N_1 + N_2$, and $\overline{Y} = [\overline{Y}_1, \overline{Y}_2]$. A simple extension of the above results shows that the concentrated likelihood function $\mathrm{GLR}(\boldsymbol{\theta})$ is the product of the l largest generalized eigenvalues of $I_{N_2} - (1/N)\overline{Y}_2^T W(\boldsymbol{\theta}) \overline{Y}_2$ and $I_{N_2} - (1/N)\overline{Y}_2^T \hat{R}^{-1} \overline{Y}_2$ (see Appendix

C). This GLR can be viewed as an extension of the detectors in [37] and [38].

It is also possible to estimate a nonparametric array response matrix A if the basis functions B are known (or possibly parametric); see [38, p. 25] and [39]. Note that in this problem, only the rank of A needs to be specified, and it is not necessary to use multiple trials or training data. This model has been used in radar array processing for the robust estimation of range and velocity; see [29] and [30].

V. Scanning

We propose a scheme to obtain initial estimates of the dipole locations by "scanning" the evoked response data with the likelihood function for a single dipole with fixed (in time) unknown orientation using the above concentrated likelihood function for nonparametric basis functions. In the EEG/MEG literature, scanning has often been performed using the MUSIC cost function [40]. However, MUSIC does not perform well when the sources are correlated [31], [41], the noise is spatially correlated [42], or both.

When there is only one dipole with unknown fixed moment orientation, we can recast the array response matrix by incorporating this orientation, which gives a response vector $\boldsymbol{a}(\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a $(3+r-1)\times 1$ parameter vector consisting of three location and $r-1$ orientation parameters. Then, $l=1$, and the ML estimate of B becomes a row vector $\hat{\boldsymbol{b}}$, which can be obtained in a closed form (see Appendix C) and the concentrated likelihood function (in the form of GLR) becomes

$$\begin{aligned}\mathrm{GLR}(\boldsymbol{\theta}) = 1 + \frac{1}{N\boldsymbol{a}(\boldsymbol{\theta})^T\hat{R}^{-1}\boldsymbol{a}(\boldsymbol{\theta})}\cdot \boldsymbol{a}(\boldsymbol{\theta})^T\hat{R}^{-1}\overline{Y} \\ \times\left[I_N - \frac{1}{N}\overline{Y}^T\hat{R}^{-1}\overline{Y}\right]^{-1}\overline{Y}^T\hat{R}^{-1}\boldsymbol{a}(\boldsymbol{\theta})\end{aligned} \tag{5.1}$$

which equals the concentrated likelihood for $B = I_N$, as observed in Section IV; see also Section A of Appendix C. This GLR is a function of $\boldsymbol{\theta}$ only; we use it to scan the evoked response data and choose parameter vectors $\boldsymbol{\theta}$ for which it achieves peak values as initial estimates in optimizing the concentrated likelihood function for multiple dipoles. The scanning procedure requires only a $(3+r-1)$-D search over the dipole location and orientation parameters.

Observe that the above expression has a Capon-like structure [43], [44] in the denominator. In the scalar case, i.e., for $\boldsymbol{a}(\boldsymbol{\theta}) = a$ and $N=1$, and after a linear transformation, the GLR expression (5.1) reduces to the familiar t-test: $\overline{y}^2/s^2$, where $\overline{y} = [\sum_{t=1}^K y(t)]/K$, $s^2 = \sum_{t=1}^K [y(t)-\overline{y}]^2$. This scanning is a reasonable method: It evaluates the likelihood of a dipole for particular location and orientation while simultaneously estimating the unknown noise covariance, which accounts for the sources of brain activity at other locations.

To further reduce the dimensionality of the search to 2-D, it is possible to use the anatomic constraints [45], i.e., assume that sources can lie only on the surface of the cortex with moments orthogonal to the cortex. However, such a method relies on the validity of the above constraints and would require using patient-specific MRI images to extract the necessary information.

If the data sets contain parts with baseline data (see also Section IV), $\overline{Y}$ in (5.1) would simply need to be substituted by $\overline{Y}_2$.

VI. Fisher Information Matrix and Cramér–Rao Bound

The FIM can be viewed as a measure of the intrinsic accuracy of a distribution [27]. Its inverse is the CRB, which is a lower bound on the covariance matrix of any unbiased estimator. It is achieved asymptotically by the ML estimator; see (3.7).

Denote the Kronecker product (which is also known as the direct product) between two matrices by $\otimes$; see [16, p. 11] for the definition and some properties. In [18, App. D], we derive the FIM for the above model as

$$\mathcal{I}(\boldsymbol{\gamma}) = \begin{bmatrix}\mathcal{I}_{\mathrm{signal}}(\boldsymbol{\gamma}) & 0 \\ 0 & \mathcal{I}_{\mathrm{noise}}(\boldsymbol{\psi})\end{bmatrix} \tag{6.1}$$

where

$$\mathcal{I}_{\mathrm{signal}}(\boldsymbol{\gamma}) = \begin{bmatrix}\mathcal{I}_{xx} & \mathcal{I}_{\theta x}^T & \mathcal{I}_{\eta x}^T \\ \mathcal{I}_{\theta x} & \mathcal{I}_{\theta\theta} & \mathcal{I}_{\eta\theta}^T \\ \mathcal{I}_{\eta x} & \mathcal{I}_{\eta\theta} & \mathcal{I}_{\eta\eta}\end{bmatrix} \tag{6.2}$$

and

$$\mathcal{I}_{xx} = KB(\boldsymbol{\eta})B(\boldsymbol{\eta})^T \otimes A(\boldsymbol{\theta})^T\Sigma^{-1}A(\boldsymbol{\theta}) \tag{6.3a}$$

$$\mathcal{I}_{\theta x} = KD_A(\boldsymbol{\theta})^T[XB(\boldsymbol{\eta})B(\boldsymbol{\eta})^T \otimes \Sigma^{-1}A(\boldsymbol{\theta})] \tag{6.3b}$$

$$\mathcal{I}_{\eta x} = KD_B(\boldsymbol{\theta})^T[B(\boldsymbol{\eta})^T \otimes X^TA(\boldsymbol{\theta})^T\Sigma^{-1}A(\boldsymbol{\theta})] \tag{6.3c}$$

$$\mathcal{I}_{\theta\theta} = KD_A(\boldsymbol{\theta})^T[XB(\boldsymbol{\eta})B(\boldsymbol{\eta})^TX^T \otimes \Sigma^{-1}]D_A(\boldsymbol{\theta}) \tag{6.3d}$$

$$\mathcal{I}_{\eta\theta} = KD_B(\boldsymbol{\theta})^T[B(\boldsymbol{\eta})^TX^T \otimes X^TA(\boldsymbol{\theta})^T\Sigma^{-1}]D_A(\boldsymbol{\theta}) \tag{6.3e}$$

$$\mathcal{I}_{\eta\eta} = KD_B(\boldsymbol{\theta})^T[I_N \otimes X^TA(\boldsymbol{\theta})^T\Sigma^{-1}A(\boldsymbol{\theta})X]D_B(\boldsymbol{\theta}) \tag{6.3f}$$

$$D_A(\boldsymbol{\theta}) = \frac{\partial\,\mathrm{vec}(A(\boldsymbol{\theta}))}{\partial\boldsymbol{\theta}^T}, \quad D_B(\boldsymbol{\theta}) = \frac{\partial\,\mathrm{vec}(B(\boldsymbol{\eta}))}{\partial\boldsymbol{\eta}^T} \tag{6.3g}$$

whereas the $(i,\,j)$th entry of $\mathcal{I}_{\mathrm{noise}}(\boldsymbol{\psi})$ is [32] , [46]

$$[\mathcal{I}_{\mathrm{noise}}(\boldsymbol{\psi})]_{ij} = \frac{NK}{2}\,\mathrm{tr}\left[\Sigma^{-1}\frac{\partial\Sigma}{\partial\psi_i}\,\Sigma^{-1}\frac{\partial\Sigma}{\partial\psi_j}\right]. \tag{6.4}$$

Further, let $\Sigma^{-1} = [\sigma^{ij}]$ and $\Sigma = [\sigma_{ij}]$; then, the following simple formula solves (6.4):

$$\begin{aligned}&\mathrm{tr}\left[\Sigma^{-1}\frac{\partial\Sigma}{\partial\sigma_{pq}}\,\Sigma^{-1}\frac{\partial\Sigma}{\partial\sigma_{rs}}\right] \\ &\quad= \begin{cases}2(\sigma^{qr}\sigma^{ps} + \sigma^{pr}\sigma^{qs}), & p\neq q,\ r\neq s \\ 2\sigma^{pr}\sigma^{qr}, & p\neq q,\ r=s \\ (\sigma^{pr})^2, & p=q,\ r=s\end{cases}\end{aligned} \tag{6.5}$$

where $p,\ q,\ r,\ s \in \{1,\cdots,m\}$.

As expected, the information increases linearly with the number of trials K. The information on noise $\mathcal{I}_{\mathrm{noise}}(\boldsymbol{\psi})$ increases linearly with N as well. In the sequel, we use the same block partitioning of the CRB as for the above FIM matrix.

Due to the block-diagonal structure of $\mathcal{I}$ that separates the signal and noise parts, its inverse is computed by simply inverting the two diagonal blocks. Thus, $\mathrm{CRB}_{\mathrm{signal}}(\boldsymbol{\gamma})$ for the unknown noise covariance equals the corresponding CRB for known noise covariance. Therefore, the ML method and GLS with correctly specified Σ (see Section III-B) have the same asymptotic covariance.

Using the matrix inversion formula [16, result vi, p. 8], it can be shown that

$$\begin{aligned}\mathrm{CRB}_{\theta\theta} = \frac{1}{K}\{&D_A(\boldsymbol{\theta})^T[XB(\boldsymbol{\eta})B(\boldsymbol{\eta})^TX^T \otimes [\Sigma^{-1} - \Sigma^{-1} \\ &\times A(\boldsymbol{\theta})(A(\boldsymbol{\theta})^T\Sigma^{-1}A(\boldsymbol{\theta}))^{-1} \\ &\times A(\boldsymbol{\theta})^T\Sigma^{-1}]]D_A(\boldsymbol{\theta})\}^{-1} \qquad (6.6a)\end{aligned}$$

$$\begin{aligned}\mathrm{CRB}_{\eta\eta} = \frac{1}{K}\{&D_B(\boldsymbol{\eta})^T[[I_N - P_B(\boldsymbol{\eta})] \\ &\otimes X^TA(\boldsymbol{\theta})^T\Sigma^{-1}A(\boldsymbol{\theta})X]D_B(\boldsymbol{\eta})\}^{-1} \qquad (6.6b)\end{aligned}$$

and

$$\mathrm{CRB}_{\theta\eta} = 0. \qquad (6.6c)$$

The expression in (6.6a) implies that $\mathrm{CRB}_{\theta\theta}$ is independent of the choice of basis functions as long as the dipole moment temporal evolutions can be expressed exactly as their linear combination, i.e., $Q = XB(\boldsymbol{\eta}) = [\boldsymbol{q}(1) \cdots \boldsymbol{q}(N)]$. (Note that this condition is satisfied for a trivial choice of basis functions in the form of Dirac pulses, i.e., $B = I_N$; then $Q = X$.) After substituting $Q = XB(\boldsymbol{\eta})$ and $K = 1$, (6.6a) equals the deterministic CRB for known Σ [31], [47].

Of course, the choice of basis functions is important for the asymptotic accuracy of estimating X and $\boldsymbol{\eta}$; thus, it affects the accuracy of estimating dipole moments' temporal evolutions. The fact that the CRB submatrix for $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$ is block diagonal [see (6.6c)] is a generalization of similar results in [29] and [30], where it was shown for a particular choice of a basis function suitable for radar array processing.

For the nonparametric basis functions, the expression in (6.6a) remains valid, but we need to be careful in handling the CRB expressions related to $\boldsymbol{\eta}$ due to the fact that when $\boldsymbol{\eta} = \mathrm{vec}(B^T)$, we need to restrict the parameter space of $\boldsymbol{\eta}$ to achieve identifiability, as discussed in more detail in [18, App. B]. Thus, we need to compute the constrained CRB. For simplicity, we may choose the orthonormality constraints, i.e., $\boldsymbol{b}_i^T\boldsymbol{b}_j = \delta_{ij}$, $i, j = 1, 2, \cdots, l$, where $B = [\boldsymbol{b}_1 \cdots \boldsymbol{b}_l]^T$, and δ_{ij} is the Kronecker delta symbol. These constraints can be written in vector form as $\boldsymbol{c}(\boldsymbol{\eta}) = \boldsymbol{0}$, where $\boldsymbol{c}(\boldsymbol{\eta})$ is of size $l(l+1)/2 \times 1$. Define the gradient matrix of the constraints as $C(\boldsymbol{\eta}) = \partial\boldsymbol{c}(\boldsymbol{\theta})/\partial\boldsymbol{\rho}^T = [0, \partial\boldsymbol{c}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}^T]$, where 0 is a matrix of zeros of dimension $l(l+1)/2 \times (rnl + 3n)$. Further, define the matrix U whose columns form a basis for the nullspace of $C(\boldsymbol{\eta})$. Following [48], the constrained CRB on the signal parameters is then

$$\mathrm{CCRB}_{\mathrm{signal}} = U[U^T\mathcal{I}_{\mathrm{signal}}(\boldsymbol{\gamma})U]^{-1}U^T \qquad (6.7)$$

where $\mathcal{I}_{\mathrm{signal}}(\boldsymbol{\gamma})$ is computed using (6.2) and (6.3) with $\boldsymbol{\eta} = \mathrm{vec}(B^T)$.

VII. Computational Issues

In this section, we derive a computationally efficient method for computing the concentrated log-likelihood function

$$l(\boldsymbol{\theta}, \boldsymbol{\eta}) = -\frac{NK}{2}\ln|\hat{\Sigma}(\boldsymbol{\theta}, \boldsymbol{\eta})| \qquad (7.1)$$

and its derivatives with respect to $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$. The maximization of (7.1) can be split into a two-step procedure: with respect to $\boldsymbol{\theta}$ with $\boldsymbol{\eta}$ fixed, and vice versa. For each case, we suitably decompose the expression in (7.1) and apply QR decompositions, as suggested in [49]. For a general discussion on the use of Cholesky and QR decompositions for regression calculations, see [49] and [50]. A QR decomposition of a matrix L of size $r \times s$ ($r \geq s$) is

$$L = \tilde{Q}\tilde{R} = \tilde{Q}\begin{bmatrix} R \\ 0 \end{bmatrix} = QR \qquad (7.2)$$

where $\tilde{Q}$ is an $r \times r$ orthogonal matrix, and $\tilde{R}$ is zero below the main diagonal. Thus, Q consists of the first s columns of $\tilde{Q}$, and R consists of the first s rows of $\tilde{R}$ (and is upper triangular).

Assume first that $\boldsymbol{\eta}$ is fixed. Then, the concentrated log-likelihood depends only on $\boldsymbol{\theta}$ and can be rewritten as [using (3.2b)]

$$\begin{aligned}l(\boldsymbol{\theta}) &= -\frac{NK}{2}\ln|S + (I_m - TS^{-1})PP^T(I_m - TS^{-1})^T| \\ &= -\frac{NK}{2}\ln|L^TL| \qquad (7.3)\end{aligned}$$

where

$$P = \overline{Y}B^T(BB^T)^{-1/2}/\sqrt{N} \qquad (7.4)$$

and $L^T = [S^{1/2}, (I_m - TS^{-1})P] = [{L_1}^T, L_2(\boldsymbol{\theta})^T]$. Observe that $L_1 = S^{1/2}$ does not depend on $\boldsymbol{\theta}$, whereas the dependence of $L_2(\boldsymbol{\theta})$ on $\boldsymbol{\theta}$ is only through $T = T(\boldsymbol{\theta})$. The QR decomposition of L is $L = QR$, where Q is of size $(m + l) \times m$, and R is upper triangular of size $m \times m$. It easily follows (see also [49] and [50]) that

$$l(\boldsymbol{\theta}) = -\frac{NK}{2}\ln\left(|R|^2\right) = -NK\ln\left(\left|\prod_{i=1}^{m} R_{ii}\right|\right) \qquad (7.5)$$

where R_{ii} is the ith diagonal entry of R. Now, following [49], we have

$$\frac{dl}{dL} = -NK(L^+)^T = -NKQ(R^T)^{-1} \qquad (7.6)$$

where $L^+ = (L^TL)^{-1}L^T$ is a pseudo-inverse of L, which is easily computed as $L^+ = R^{-1}Q^T$. Further, the gradient of $l(\boldsymbol{\theta})$ follows from

$$\begin{aligned}\frac{\partial l(\boldsymbol{\theta})}{\partial\theta_k} &= -NK\sum_{i=1}^{2m}\sum_{j=1}^{m}\frac{\partial L_{ij}}{\partial\theta_k}[(L^+)^T]_{ij} \\ &= -NK\,\mathrm{tr}\left(L^+\frac{\partial L}{\partial\theta_k}\right). \qquad (7.7)\end{aligned}$$

The above expression can be simplified by observing that

$$\frac{\partial L_1}{\partial \theta_k} = 0 \tag{7.8a}$$

$$\frac{\partial L_2(\boldsymbol{\theta})}{\partial \theta_k} = \frac{\partial (P^T - P^T S^{-1/2} \Pi S^{1/2})}{\partial \theta_k} = -P^T S^{-(1/2)} \frac{\partial \Pi}{\partial \theta_k} S^{1/2} \tag{7.8b}$$

where $\Pi = S^{-1/2} T S^{-1/2}$ is a projection matrix, and thus (see, e.g., [28])

$$\frac{\partial \Pi}{\partial \theta_k} = \dot{\Pi}_k + \dot{\Pi}_k^T \tag{7.9a}$$

$$\dot{\Pi}_k = (I_m - \Pi) S^{-(1/2)} \frac{\partial A(\boldsymbol{\theta})}{\partial \theta_k} [S^{-(1/2)} A(\boldsymbol{\theta})]^+. \tag{7.9b}$$

Since L_1 does not depend on $\boldsymbol{\theta}$, we can decompose it as $L_1 = Q_1 R_1$ and retain only R_1 for further calculation since

$$|L^T L| = |{L_1}^T L_1 + L_2(\boldsymbol{\theta})^T L_2(\boldsymbol{\theta})| = |{R_1}^T R_1 + L_2(\boldsymbol{\theta})^T L_2(\boldsymbol{\theta})|. \tag{7.10}$$

Therefore, instead of factoring L, we can factor $L' = [R_1^T, L_2(\boldsymbol{\theta})^T]^T = QR'$, and then, $l(\boldsymbol{\theta}) = -NK \ln(|\prod_{i=1}^m R'_{ii}|)$.

We now fix $\boldsymbol{\theta}$ and maximize (7.1) with respect to $\boldsymbol{\eta}$. Using (C.1), and (C.2) from Appendix C, we rewrite (7.1) as

$$l(\boldsymbol{\eta}) = \frac{NK}{2} \left[\ln \left| B(\boldsymbol{\eta}) \left(I_N - \frac{1}{N} \overline{Y}^T W \overline{Y} \right) B(\boldsymbol{\eta})^T \right| - \ln \left| B(\boldsymbol{\eta}) \left(I_N - \frac{1}{N} \overline{Y}^T \hat{R}^{-1} \overline{Y} \right) B(\boldsymbol{\eta})^T \right| - \ln |\hat{R}| \right]. \tag{7.11}$$

Since W does not depend on $\boldsymbol{\eta}$, we can do the following decompositions only once in this step: $I_N - (1/N)\overline{Y}^T W \overline{Y} = M_W M_W^T$ and $I - (1/N)\overline{Y}^T \hat{R}^{-1} \overline{Y} = M_R M_R^T$, which yields

$$l(\boldsymbol{\eta}) = \frac{NK}{2} \left[\ln \left(|L_W(\boldsymbol{\eta})^T L_W(\boldsymbol{\eta})| \right) - \ln(|L_R(\boldsymbol{\eta})^T L_R(\boldsymbol{\eta})|) - \ln |\hat{R}| \right]. \tag{7.12}$$

where $L_W(\boldsymbol{\eta}) = M_W^T B(\boldsymbol{\eta})^T$, and $L_R(\boldsymbol{\eta}) = M_R^T B(\boldsymbol{\eta})^T$. We then proceed in a similar manner as before; find the QR decompositions $L_W = Q_W R_W$ and $L_R = Q_R R_R$ (where Q_W and Q_R are of size $N \times l$, and R_W and R_R are upper triangular of size $l \times l$). Then, it easily follows that

$$\frac{\partial l(\boldsymbol{\eta})}{\partial \eta_k} = NK \left[\text{tr} \left(L_W^+ \frac{\partial L_W}{\partial \eta_k} \right) - \text{tr} \left(L_R^+ \frac{\partial L_R}{\partial \eta_k} \right) \right] \tag{7.13}$$

where $L_W^+ = R_W^{-1} Q_W^T$, $L_R^+ = R_R^{-1} Q_R^T$, and

$$\frac{\partial L_W}{\partial \eta_k} = M_W^T \frac{\partial B(\boldsymbol{\eta})^T}{\partial \eta_k} \tag{7.14a}$$

$$\frac{\partial L_R}{\partial \eta_k} = M_R^T \frac{\partial B(\boldsymbol{\eta})^T}{\partial \eta_k}. \tag{7.14b}$$

If $\boldsymbol{\eta} = \text{vec}(B^T)$, there is a closed-form solution for optimizing $l(\boldsymbol{\eta})$, as shown in Section IV.

VIII. Numerical Simulations

In this section, we compare the localization accuracy of the ML, GLS, OLS, and scanning methods when spatially correlated noise is added to a simulated evoked response. Our simulations confirm the theoretical results presented in Sections III and VI.

The simulation was performed for an MEG configuration of 37 radial magnetometers located on a spherical helmet of radius $r = 10$ cm with a single sensor at the pole of the cap and three rings at elevation angles of $\pi/12$, $\pi/6$, and $\pi/4$ rad, containing, respectively 6, 12, and 18 sensors equally spaced in the azimuthal direction. This arrangement is similar to an array made commercially by Biomagnetic Technologies, Inc. (BTI), San Diego, CA.

We generated two coherent dipole sources. The components q_θ and q_φ of the first dipole change in time according to

$$q_\theta = 15 \exp(-(t-60)^2/8^2) - 5 \exp(-(t-40)^2/17^2) \ [\text{nA} \cdot \text{m}] \tag{8.1a}$$

$$q_\varphi = 13 \exp(-(t-60)^2/12^2) - 3 \exp(-(t-40)^2/17^2) \ [\text{nA} \cdot \text{m}] \tag{8.1b}$$

and the corresponding components of the second dipole as $q_\theta(t)$ and $-q_\varphi(t)$, i.e., the sources are correlated. The dipoles are symmetric relative to the midsagittal plane with locations $\boldsymbol{\theta}_1 = [\pi/6, -\pi/3, 5 \text{ cm}]$ and $\boldsymbol{\theta}_2 = [\pi/6, \pi/3, 5 \text{ cm}]$.

We simulated 50 runs, each consisting of $K = 10$ trials and $N = 100$ snapshots per trial. To approximate realistic spatially correlated noise, we generated 400 random dipoles uniformly distributed on a sphere with a radius of 5 cm (for a discussion on random dipole modeling of spontaneous brain activity, see [14]). For each noise dipole, we assumed that its two tangential moment components were uncorrelated and distributed as $\mathcal{N}(0, \sigma_{\text{m}}^2)$. For $\sigma_{\text{m}} = 1$ nA $\cdot$ m, the total noise standard deviation at the sensors was approximately 110 fT, which is consistent with 25 fT $/\sqrt{\text{Hz}}$ one-sided white noise spectral density bandlimited to 20 Hz. We justify this choice by the fact that typically recorded background noise spectral density is 20–40 fT $/\sqrt{\text{Hz}}$ below 20 Hz [2]. The peak value of the signal at the sensor with the largest response was around 270 fT, which is consistent with typical values measured in practical applications.

In the EEG/MEG literature, several parametric models have been used to model temporal evolution of the evoked responses: decaying sinusoids (see [51]); double Gaussian (see [52]); or Hermite wavelets (see [53]). In this example, we choose a combination of Gaussian and harmonic terms, i.e., $\boldsymbol{\phi}(t, \boldsymbol{\eta}) = [\exp(-(t-\tau_1)^2/\sigma_1^2), \exp(-(t-\tau_2)^2/\sigma_2^2), 1, \sin(\omega t), \sin(2\omega t), \sin(3\omega t), \cos(\omega t), \cos(2\omega t), \cos(3\omega t)]^T$. Hence, the unknown parameter vector describing the temporal evolution is $\boldsymbol{\eta} = [\tau_1, \sigma_1, \tau_2, \sigma_2, \omega]^T$. The two Gaussian functions were used to model peaks in the response, and the sine and cosine terms model the low-pass signal component. Such components are typical in evoked responses.

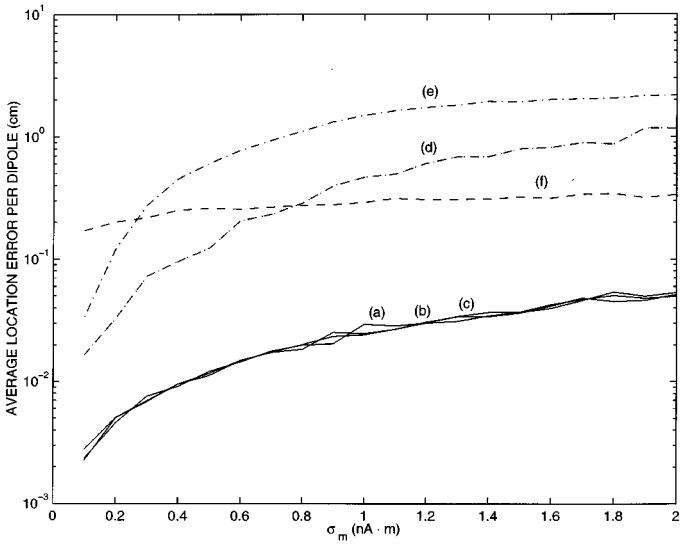


Fig. 2. Average location error per dipole as a function of the noise level $\sigma_{\rm m}$ for (a) ML method with parametric basis functions. (b) ML method with $l = 2n$ nonparametric basis functions. (c) GLS with $B = I_N$. (d) OLS with parametric basis functions. (e) OLS with $B = I_N$. (f) Scanning with unknown fixed dipole orientation.

In Fig. 2, we compare the localization accuracies of the ML, GLS, OLS, and scanning methods by showing the mean localization errors per dipole $(1/2)(\|\hat{\boldsymbol{p}}_1 - \boldsymbol{p}_1\|^2 + \|\hat{\boldsymbol{p}}_2 - \boldsymbol{p}_2\|^2)^{1/2}$ (averaged over the 50 runs) as functions of $\sigma_{\rm m}$. Here, $\|\cdot\|$ denotes the Euclidean norm, and $\boldsymbol{p}_1, \boldsymbol{p}_2$, and $\hat{\boldsymbol{p}}_1, \hat{\boldsymbol{p}}_2$ are the location vectors of the two dipoles and the corresponding ML estimates; see also (2.1). The standard deviation of the localization error curves is the largest for OLS with parametric basis functions (up to 0.3 mm).

It is interesting to note that as $\sigma_{\rm m}$ increases, the dipole location estimates obtained by the OLS methods move toward the center of the head. Thus, the error values become comparable to the head's dimensions for large $\sigma_{\rm m}$ (as shown in Fig. 2), whereas the ML estimation errors remain very small, showing the robustness of the ML method. A similar trend was also observed in [10] and [11].

The average location error is approximately the same for the ML methods with parametric and nonparametric basis functions and the GLS method, which is consistent with the asymptotic results in Section VI, where we show that the ML and GLS methods have the same asymptotic accuracy of the signal parameters $\boldsymbol{\rho}$ (due to the block-diagonal structure of the FIM for signal and noise) and the parametric and nonparametric ML have the same asymptotic location accuracy (since $\mathrm{CRB}_{\theta\theta}$ is independent of the choice of basis functions as long as the dipole moment temporal evolutions can be expressed exactly as their linear combination).

We have applied the scanning algorithm in Section V, which requires a 4-D search (3-D for the location and 1-D for orientation). As shown in Fig. 2, this algorithm is robust to the increase in the noise level σ_m because it acounts for spatially correlated noise. Further, for larger values of σ_m, it outperforms the OLS algorithms, which do not account for the correlation in the noise. This is an important result since scanning is computationally simpler than OLS (OLS with $B = I_N$ requires a 6-D search for the two-dipole fit). Note that for small values of σ_m, the OLS algorithms perform better than scanning because they fit the exact noiseless response (two dipoles in this case), which becomes more important than the noise correlation when the noise level is small.

In this example, we have used a very small number of trials ($K = 10$). As $K \to \infty$, both the ML and OLS estimates converge to the true parameters, as shown in Theorems 1 and 2. In some real data applications, the number of trials is $K = 100$ or more; then, the ML and OLS results may differ only by a few millimeters [54].

IX. Concluding Remarks

We proposed maximum likelihood methods for estimating evoked dipole responses using a combination of EEG and MEG arrays, assuming spatially correlated noise with unknown covariance. To exploit prior information on the shapes of the evoked responses and improve the estimation of the dipole moments, we modeled them as linear combinations of parametric basis functions. Utilizing multiple trials, we also derived the estimation method for nonparametric basis functions, which allows for computation of the concentrated likelihood function that depends only on the dipole locations (but needs many parameters to describe the moment evolutions). We further showed how to obtain initial estimates of the dipole locations using scanning. We presented a computationally efficient method for implementing the ML estimation. Cramér–Rao bounds for the proposed model were derived. We also showed that the proposed estimators are asymptotically more efficient compared with the nonlinear OLS estimators.

We presented numerical simulations demonstrating the performance of the ML and scanning methods. The ML and OLS methods were compared; the ML was more accurate and robust, confirming the theoretical results in Section III.

In [55], we extended the above method to solve the problem of dipoles having fixed orientations in time, whereas their strengths were modeled by a linear combination of basis functions.

We applied the proposed algorithms to real EEG/MEG data; see [18] (some results are also available at http://www.eecs.uic.edu/~nehorai/MEG.html). Further research will include

- analysis of the proposed methods in the presence of more realistic noise and signal models (e.g., temporally correlated noise, latency jitters, etc.);
- tracking moving dipoles [56], [57];
- classifying evoked responses for diagnostic purposes;
- optimal design of EEG/MEG sensor arrays [55] and novel performance measures [19], [58];
- more realistic array response modeling (e.g., incorporating a realistic patient-specific head model obtained from MRI scans in source estimation and performance analysis following [57] and [59]);
- more extensive applications to real data.

Appendix A
ML Estimation

We derive the ML estimates of the matrix of basis function coefficients X and noise covariance for known $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$. Then,

we present the concentrated likelihood function that should be maximized when $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$ are unknown to obtain their ML estimates.

Stack all the measurement matrices into one matrix $\tilde{Y} = [Y_1, Y_2 \cdots Y_K]$ of size $m \times NK$. Similarly, put K basis function matrices B into a matrix $\tilde{B} = [B \cdots B]$ of size $l \times NK$. The likelihood function for this model is

$$f(X,\Sigma) = \frac{1}{(2\pi)^{(1/2)mNK}|\Sigma|^{(1/2)NK}} \times \exp\left\{-\tfrac{1}{2}\,\mathrm{tr}\left[\Sigma^{-1}(\tilde{Y}-AX\tilde{B})(\tilde{Y}-AX\tilde{B})^T\right]\right\}. \tag{A.1b}$$

Then, according to [25] and [26, ch. 6.4], the ML estimates of X and Σ are

$$\hat{X} = (A^T S^{-1} A)^{-1} A^T S^{-1} \tilde{Y}\tilde{B}^T(\tilde{B}\tilde{B}^T)^{-1} \tag{A.2a}$$
$$\hat{\Sigma} = S + (I_m - TS^{-1}) S_1 (I_m - TS^{-1})^T \tag{A.2b}$$

where

$$\hat{R} = \frac{1}{NK}\tilde{Y}\tilde{Y}^T = \frac{1}{NK}\sum_{k=1}^{K} Y_k Y_k^T \tag{A.3a}$$
$$S = \hat{R} - S_1 \tag{A.3b}$$
$$S_1 = \frac{1}{NK}\tilde{Y}\tilde{B}^T(\tilde{B}\tilde{B}^T)^{-1}\tilde{B}\tilde{Y}^T \tag{A.3c}$$
$$T = A(\boldsymbol{\theta})\left[A(\boldsymbol{\theta})^T S^{-1} A(\boldsymbol{\theta})\right]^{-1} A(\boldsymbol{\theta})^T. \tag{A.3d}$$

Observe that

$$\tilde{B}\tilde{B}^T = KBB^T \tag{A.4a}$$
$$\tilde{B}\tilde{Y}^T = \sum_{k=1}^{K} BY_k^T = KB\overline{Y}^T \tag{A.4b}$$

directly yields (3.3). Substituting the above estimates into the likelihood function (A.1b), we obtain the concentrated likelihood function

$$f(\hat{X},\hat{\Sigma}) = \frac{1}{(2\pi)^{(1/2)mNK}|\hat{\Sigma}|^{(1/2)NK}}\ \exp\left[-\frac{1}{2}mNK\right] \tag{A.5}$$

which is proportional to the expression in (3.5).

Appendix B
ML Versus OLS

We show that the ML estimates of $\boldsymbol{\rho}$ derived in this paper are asymptotically more efficient than the OLS estimates, i.e., the difference in their asymptotic variances is negative semidefinite.

Assume that the regularity conditions R1)–R5) hold. Theorem 1 implies that the asymptotic covariance matrix of $\sqrt{NK}\hat{\boldsymbol{\rho}}$ is (see also [18, App. D])

$$\begin{aligned} C_{\mathrm{ML}}^{\infty} &= NK\left[\mathcal{I}_{\mathbf{signal}}(\boldsymbol{\theta})\right]^{-1} \\ &= NK\left[\sum_{t=1}^{N} D(t,\boldsymbol{\rho})^T \Sigma^{-1} D(t,\boldsymbol{\rho})\right]^{-1} \\ &= NK\left[D^T\mathcal{S}^{-1}D\right]^{-1} \end{aligned} \tag{B.1}$$

where

$$D(t,\boldsymbol{\rho}) = \frac{\partial(A(\boldsymbol{\theta})X\boldsymbol{\phi}(t,\boldsymbol{\eta}))}{\partial\boldsymbol{\rho}^T} \tag{B.2a}$$
$$D = [D(1,\boldsymbol{\rho})^T \cdots D(N,\boldsymbol{\rho})^T]^T \tag{B.2b}$$
$$\mathcal{S} = I_N \otimes \Sigma \tag{B.2c}$$

which can be further simplified [see (6.2) and (6.3)].

The asymptotic covariance matrix of $\sqrt{NK}\hat{\boldsymbol{\rho}}_{\mathrm{OLS}}$ is (see Theorem 2)

$$C_{\mathrm{OLS}}^{\infty} = NK\left[D^TD\right]^{-1} D^T\mathcal{S}D\left[D^TD\right]^{-1}. \tag{B.3}$$

Let T_d be an arbitrary full-rank matrix such that its columns span the space orthogonal to the column space of D; thus, $D^TT_d = 0$. Then

$$\begin{aligned} T_d(T_d^T\mathcal{S}T_d)^{-1}T_d^T &= \mathcal{S}^{-1} - \mathcal{S}^{-1}D \\ &\quad\times (D^T\mathcal{S}^{-1}D)^{-1}D^T\mathcal{S}^{-1} \end{aligned} \tag{B.4}$$

which is Lemma 1 in [25] (see also [27, p. 77]). It follows that

$$\begin{aligned} \left[D^T\mathcal{S}^{-1}D\right]^{-1} &= \left[D^TD\right]^{-1}D^T\mathcal{S}D\left[D^TD\right]^{-1} \\ &\quad - \left[D^TD\right]^{-1}D^T\mathcal{S}T_d\left[T_d^T\mathcal{S}T_d\right]^{-1}T_d^T \\ &\quad\times \mathcal{S}D\left[D^TD\right]^{-1} \end{aligned} \tag{B.5}$$

and thus, $C_{\mathrm{ML}}^{\infty} - C_{\mathrm{OLS}}^{\infty} \le 0$. Note that equality holds if $\Sigma = \sigma^2 I_m$.

Appendix C
Nonparametric Basis Functions

To maximize the concentrated likelihood with respect to nonparametric basis functions, we express it as a function of $\hat{R}$ instead of S (since S is a function of B) by repeatedly applying the matrix inversion lemma, see [38, pp. 26–28]

$$\begin{aligned} \mathrm{GLR}(\boldsymbol{\theta},\boldsymbol{\eta}) &= |\hat{R}|\cdot l_{\mathrm{ML}}(\boldsymbol{\theta},\boldsymbol{\eta})^{2/NK} \\ &= \frac{|\hat{R}|}{|\hat{\Sigma}(\boldsymbol{\theta},\boldsymbol{\eta})|} \\ &= \frac{\left|\tilde{B}\left(I_{NK} - \frac{1}{NK}\tilde{Y}^TW(\boldsymbol{\theta})\tilde{Y}\right)\tilde{B}^T\right|}{\left|\tilde{B}\left(I_{NK} - \frac{1}{NK}\tilde{Y}^T\hat{R}^{-1}\tilde{Y}\right)\tilde{B}^T\right|} \end{aligned} \tag{C.1}$$

where $W(\boldsymbol{\theta})$ is defined in (4.1). Compared with $l_{\mathrm{ML}}(\boldsymbol{\theta}, \boldsymbol{\eta})$, the above concentrated likelihood function is normalized by $|\hat{R}|$ and represents an expression for the GLR test for testing $\mathrm{H}_0 : X = 0$.

Using the lemma in (B.4), we can compute $W(\boldsymbol{\theta})$ in the following alternative way: $W(\boldsymbol{\theta}) = T_a(T_a^T \hat{R} T_a)^{-1} T_a$, where T_a is an arbitrary full-rank $m \times (m - nr)$ matrix such that $A(\boldsymbol{\theta})^T T_a = 0$ [assuming that $A(\boldsymbol{\theta})$ is full-rank], i.e., T_a spans the space orthogonal to the column space of $A(\boldsymbol{\theta})$.

The matrix in the denominator $I_{NK} - 1/(NK)\tilde{Y}^T \hat{R}^{-1} \tilde{Y}$ is a projection matrix with rank $NK - m$. Thus, for fixed $\boldsymbol{\eta}$ and if $NK - m - l \geq 0$, the denominator of the above expression is nonzero with probability one. Note also that for $K = 1$, the GLR in (C.1) would go to infinity if we choose the rows of $\tilde{B} = B$ from the row space of $\tilde{Y} = Y$.

Using (C.1), (A.4a), and (A.4b), we get

$$\mathrm{GLR}(\boldsymbol{\theta}, \boldsymbol{\eta}) = \frac{\left| B \left(I_N - \frac{1}{N} \overline{Y}^T W(\boldsymbol{\theta}) \overline{Y} \right) B^T \right|}{\left| B \left(I_N - \frac{1}{N} \overline{Y}^T \hat{R}^{-1} \overline{Y} \right) B^T \right|}. \tag{C.2}$$

Further, $\hat{X}$ can also be computed as a function of $\hat{R}$ instead of S as

$$\begin{aligned} \hat{X} = \sqrt{N} \Big[& A(\boldsymbol{\theta})^T \hat{R}^{-1} A(\boldsymbol{\theta}) + A(\boldsymbol{\theta})^T \hat{R}^{-1} P \\ & \times \left[I_l - P^T \hat{R}^{-1} P \right]^{-1} P^T \hat{R}^{-1} A(\boldsymbol{\theta}) \Big]^{-1} \\ & \times A(\boldsymbol{\theta})^T \hat{R}^{-1} P \left[I_l - P^T \hat{R}^{-1} P \right]^{-1} (BB^T)^{-1/2} \end{aligned} \tag{C.3}$$

where P is defined in (7.4). Equation (C.3) follows from (3.2a) and

$$S^{-1} = \hat{R}^{-1} + \hat{R}^{-1} P (I_l - P^T \hat{R}^{-1} P)^{-1} P^T \hat{R}^{-1} \tag{C.4}$$

which is obtained by using the matrix inversion lemma.

Consider the basis function matrix of the form $B = C\Delta^T (I_N - (1/N)\overline{Y}^T \hat{R}^{-1} \overline{Y})^{-1/2}$, where C is an $l \times N$ matrix of full rank l, and Δ is the matrix whose columns are the (normalized) eigenvectors of $\Xi(\boldsymbol{\theta}) = (I_N - (1/N)\overline{Y}^T \hat{R}^{-1} \overline{Y})^{-1/2} [I_N - (1/N)\overline{Y}^T W(\boldsymbol{\theta}) \overline{Y}] (I_N - (1/N)\overline{Y}^T \hat{R}^{-1} \overline{Y})^{-1/2}$ that are ordered to correspond to the eigenvalues of $\Xi(\boldsymbol{\theta})$ (which are denoted by λ_j, $j = 1, \cdots, N$) sorted in nonincreasing order, i.e., $\lambda_1 \geq \lambda_2 \cdots \geq \lambda_N$. Thus, $\Xi(\boldsymbol{\theta}) = \Delta \, \mathrm{diag}\{\lambda_1, \cdots, \lambda_N\} \Delta^T$. In addition, denote by Δ_l the matrix containing the first l columns of Δ, which are the eigenvectors corresponding to the largest l eigenvalues of $\Xi(\boldsymbol{\theta})$. Then, (C.2) reduces to

$$\begin{aligned} \mathrm{GLR}(\boldsymbol{\theta}, \boldsymbol{\eta}) &= \frac{|C\Delta^T \Xi(\boldsymbol{\theta}) \Delta C^T|}{|CC^T|} \\ &= \frac{|C \, \mathrm{diag}\{\lambda_1, \cdots, \lambda_N\} C^T|}{|CC^T|} \end{aligned} \tag{C.5}$$

which is maximized for $C = H \cdot [I_l, 0]$, where H is an arbitrary $l \times l$ matrix of full rank, and the maximum is equal to $\prod_{j=1}^{l} \lambda_j$. Thus

$$\hat{B} = H \Delta_l^T \left(I_N - \frac{1}{N} \overline{Y}^T \hat{R}^{-1} \overline{Y} \right)^{-1/2}. \tag{C.6}$$

For $H = I_l$, the rows of $\hat{B}$ are the generalized eigenvectors of the matrices $I_N - (1/N)\overline{Y}^T W(\boldsymbol{\theta}) \overline{Y}$ and $I_N - (1/N)\overline{Y}^T \hat{R}^{-1} \overline{Y}$ that correspond to the largest l generalized eigenvalues of these two matrices; the product of these eigenvalues is $\mathrm{GLR}(\boldsymbol{\theta}) = \prod_{j=1}^{l} \lambda_j$.

Note that $\Xi(\boldsymbol{\theta})$ can be written as

$$\begin{aligned} \Xi(\boldsymbol{\theta}) = & I_N + \frac{1}{N} \left(I_N - \frac{1}{N} \overline{Y}^T \hat{R}^{-1} \overline{Y} \right)^{-1/2} \overline{Y}^T \hat{R}^{-1} A(\boldsymbol{\theta}) \\ & \times \left[A(\boldsymbol{\theta})^T \hat{R}^{-1} A(\boldsymbol{\theta}) \right]^{-1} A(\boldsymbol{\theta})^T \hat{R}^{-1} \\ & \times \overline{Y} \left(I_N - \frac{1}{N} \overline{Y}^T \hat{R}^{-1} \overline{Y} \right)^{-1/2}. \end{aligned} \tag{C.7}$$

The second term in (C.7) is a positive semidefinite symmetric matrix with rank $\min(\mathrm{rank}(A(\boldsymbol{\theta})), N)$, which equals $\mathrm{rank}(A(\boldsymbol{\theta})) = nr$ in most practical applications.

We now show that although $\hat{B}$ is not unique, the moment temporal evolution $\hat{Q} = \hat{X}\hat{B}$ is. Using (3.2a), we get

$$\hat{Q} = \hat{X}\hat{B} = [A(\boldsymbol{\theta})^T S^{-1} A(\boldsymbol{\theta})]^{-1} A(\boldsymbol{\theta})^T S^{-1} \overline{Y} P_{\hat{B}} \tag{C.8}$$

where $P_{\hat{B}}$, which is the projection matrix on the row space of $\hat{B}$, is independent of H because it cancels out. Since S depends on B only through P_B [see (3.3)], $\hat{Q}$ is also independent of H.

Substituting $H = [\Delta_l^T (I_N - (1/N)\overline{Y}^T \hat{R}^{-1} \overline{Y})^{-1} \Delta_l]^{-1/2}$ into (C.6), we get orthonormal basis functions, i.e., $\hat{B}\hat{B}^T = I_l$.

When $Y_k = [Y_{1k}, Y_{2k}]$, $k = 1, \cdots, K$, where Y_{1k} contains baseline data and Y_{2k} contains the evoked response, the (C.2) becomes

$$\mathrm{GLR}(\boldsymbol{\theta}, \boldsymbol{\eta}) = \frac{\left| B_2 \left(I_{N_2} - \frac{1}{N} \overline{Y}_2^T W(\boldsymbol{\theta}) \overline{Y}_2 \right) B_2^T \right|}{\left| B_2 \left(I_{N_2} - \frac{1}{N} \overline{Y}_2^T \hat{R}^{-1} \overline{Y}_2 \right) B_2^T \right|}. \tag{C.9}$$

The concentrated likelihood function $\mathrm{GLR}(\boldsymbol{\theta})$ is the product of l largest eigenvalues of $(I_{N_2} - (1/N)\overline{Y}_2^T\hat{R}^{-1}\overline{Y}_2)^{-1/2}[I_{N_2} - (1/N)\overline{Y}_2^T W(\boldsymbol{\theta})\overline{Y}_2](I_{N_2} - (1/N)\overline{Y}_2^T\hat{R}^{-1}\overline{Y}_2)^{-1/2}$. Thus, the result in Section IV follows.

A. Scanning

For a single source with fixed orientation, i.e., when $A(\boldsymbol{\theta}) = \boldsymbol{a}(\boldsymbol{\theta})$, B reduces to a row vector $\boldsymbol{b}^T$, and (C.2) becomes

$$\mathrm{GLR}(\boldsymbol{\theta}, \boldsymbol{b}) = 1 + \frac{1}{N\boldsymbol{a}(\boldsymbol{\theta})^T\hat{R}^{-1}\boldsymbol{a}(\boldsymbol{\theta})} \cdot \frac{(\boldsymbol{a}(\boldsymbol{\theta})^T\hat{R}^{-1}\overline{Y}\boldsymbol{b})^2}{\boldsymbol{b}^T\left(I_N - \frac{1}{N}\overline{Y}^T\hat{R}^{-1}\overline{Y}\right)\boldsymbol{b}}. \quad \text{(C.10)}$$

Application of the Cauchy–Schwartz inequality [27, p. 54] to the above expression gives the ML estimate of $\boldsymbol{b}$

$$\begin{aligned}\hat{\boldsymbol{b}}^T &= [\phi(1), \cdots, \phi(N)] \\ &= \boldsymbol{a}(\boldsymbol{\theta})^T\hat{R}^{-1}\overline{Y}\left[I_N - \frac{1}{N}\overline{Y}^T\hat{R}^{-1}\overline{Y}\right]^{-1} \quad \text{(C.11)}\end{aligned}$$

and the GLR in (5.1), which can also be obtained by substituting $B = I_N$ and $A(\boldsymbol{\theta}) = \boldsymbol{a}(\boldsymbol{\theta})$ into (C.2) and using the formula for the determinant of a partitioned matrix, see, e.g., [16, result v, p. 8].

ACKNOWLEDGMENT

The authors are grateful to Dr. B. Lütkenhöner from the University of Münster, Münster, Germany, Prof. C. H. Muravchik from the University of La Plata, La Plata, Argentina, and the anonymous reviewers for their helpful comments.

REFERENCES

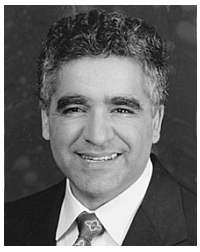
[1] J. Raz and B. Turetsky, "Event-related potentials," in *Encyclopedia of Biostatistics*. Chichester, U.K.: Wiley, 1998, vol. 2, pp. 1407–1409.
[2] M. S. Hamalainen, R. Hari, R. Ilmoniemi, J. Knuutila, and O. V. Lounasmaa, "Magnetoencephalography—Theory, instrumentation, and applications to noninvasive studies of signal processing of the human brain," *Rev. Mod. Phys.*, vol. 65, no. 2, pp. 413–497, Apr. 1993.
[3] M. Scherg and D. Von Cramon, "Two bilateral sources of the late AEP as identified by a spatiotemporal dipole model," *Electroencephalogr. Clin. Neurophysiol.*, vol. 62, pp. 32–44, 1985.
[4] J. C. de Munck, "The estimation of time varying dipoles on the basis of evoked potentials," *Electroencephalogr. Clin. Neurophysiol.*, vol. 77, pp. 156–160, 1990.
[5] M. Scherg and D. Von Cramon, "Evoked dipole source potentials of the human auditory cortex," *Electroencephalogr. Clin. Neurophysiol.*, vol. 65, pp. 344–360, 1986.
[6] C. Braun, S. Kaiser, W. E. Kincses, and T. Elbert, "Confidence interval of single dipole locations based on EEG data," *Brain Topogr.*, vol. 10, no. 1, pp. 31–39, 1997.
[7] H. M. Huizenga and P. C. M. Molenaar, "Equivalent source estimation of scalp potential fields contaminated by heteroscedastic and correlated noise," *Brain Topogr.*, vol. 8, no. 1, pp. 13–33, 1995.
[8] K. Sekihara, Y. Ogura, and M. Hotta, "Maximum-likelihood estimation of current-dipole parameters for data obtained using multichannel magnetometer," *IEEE Trans. Biomed. Eng.*, vol. 39, pp. 558–562, June 1992.
[9] A. Dogandžić and A. Nehorai, "Detecting a dipole source by MEG/EEG and generalized likelihood ratio tests," in *Proc. 30th Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, Nov. 1996, pp. 1196–1200.
[10] B. Lütkenhöner, "Dipole source localization by means of maximum likelihood estimation I–Theory and simulations," *Electroencephalogr. Clin. Neurophysiol.*, vol. 106, pp. 314–321, 1998.
[11] ——, "Dipole source localization by means of maximum likelihood estimation II–Experimental evaluation," *Electroencephalogr. Clin. Neurophysiol.*, vol. 106, pp. 322–329, 1998.
[12] T. Yamazaki, B. W. van Dijk, and H. Spekreijse, "Confidence limits for the parameter estimation in the dipole localization method on the basis of spatial correlation of background EEG," *Brain Topogr.*, vol. 5, no. 2, pp. 195–198, 1992.
[13] ——, "The accuracy of localizing equivalent dipoles and the spatiotemporal correlations of background EEG," *IEEE Trans. Biomed. Eng.*, vol. 45, pp. 1114–1121, Sept. 1998.
[14] J. C. de Munck, P. C. Vijn, and F. H. Lopes da Silva, "A random dipole model for spontaneous brain activity," *IEEE Trans. Biomed. Eng.*, vol. 39, pp. 791–804, Aug. 1992.
[15] P. C. Vijn, B. W. van Dijk, and H. Spekreijse, "Topography of occipital EEG-reduction upon visual stimulation," *Brain Topogr.*, vol. 5, no. 2, pp. 177–181, 1992.
[16] E. F. Vonesh and V. M. Chinchilli, *Linear and Nonlinear Models for the Analysis of Repeated Measurements*. New York: Marcel Dekker, 1997.
[17] A. Dogandžić and A. Nehorai, "Estimating evoked dipole responses by MEG/EEG for unknown noise covariance," in *Proc. 19th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Chicago, IL, Oct. 1997, pp. 1224–1227.
[18] ——, "Estimating evoked dipole responses in unknown spatially correlated noise with EEG/MEG arrays," Dept. Elect. Eng. Comput. Sci., Univ. Illinois at Chicago, Chicago, IL, Rep. UIC-EECS-98-1, May 1998.
[19] B. Hochwald and A. Nehorai, "Magnetoencephalography with diversely-oriented and multi-component sensors," *IEEE Trans. Biomed. Eng.*, vol. 44, pp. 40–50, Jan. 1997.
[20] Z. Zhang, "A fast method to compute surface potentials generated by dipoles within multilayer anisotropic spheres," *Phys. Med. Biol.*, vol. 40, pp. 335–349, 1995.
[21] T. Gasser, J. Möchs, and W. Köhler, "Amplitude probability distribution of noise for flash-evoked potentials and robust response estimates," *IEEE Trans. Biomed. Eng.*, vol. BME-33, pp. 579–584, June 1986.
[22] J. A. McEwen and G. B. Anderson, "Modeling the stationarity and Gaussianity of spontaneous electroencephalographic activity," *IEEE Trans. Biomed. Eng.*, vol. BME-22, pp. 361–369, Sept. 1975.
[23] J. Möchs, D. T. Pham, and T. Gasser, "Testing for homogeneity of noisy signals evoked by repeated stimuli," *Ann. Stat.*, vol. 12, pp. 193–209, 1984.
[24] R. F. Potthoff and S. N. Roy, "A generalized multivariate analysis of variance model useful especially for growth curve problems," *Biometrika*, vol. 51, pp. 313–326, 1964.
[25] C. G. Khatri, "A note on a MANOVA model applied to problems in growth curve," *Ann. Inst. Statist. Math.*, vol. 18, pp. 75–86, 1966.
[26] M. S. Srivastava and C. G. Khatri, *An Introduction to Multivariate Statistics*. New York: North–Holland, 1979.
[27] C. R. Rao, *Linear Statistical Inference and Its Applications*, 2nd ed. New York: Wiley, 1973.
[28] M. Viberg, P. Stoica, and B. Ottersten, "Maximum likelihood array processing in spatially correlated noise fields using parameterized signals," *IEEE Trans. Signal Processing*, vol. 45, pp. 996–1004, Apr. 1997.
[29] A. L. Swindlehurst and P. Stoica, "Maximum likelihood methods in radar array signal processing," *Proc. IEEE*, vol. 86, pp. 421–441, Feb. 1998.
[30] A. Dogandžić and A. Nehorai, "Estimating range, velocity, and direction with a radar array," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, Phoenix, AZ, Mar. 1999, pp. 2773–2776.
[31] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood and Cramér–Rao bound," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 720–741, May 1989.
[32] B. Porat, *Digital Processing of Random Signals: Theory and Methods*. Englewood Cliffs, NJ: Prentice-Hall, 1994.
[33] A. R. Gallant, *Nonlinear Statistical Models*. New York: Wiley, 1987.

[34] C. Gennings, V. M. Chinchilli, and W. H. Carter, "Response surface analysis with correlated data: A nonlinear model approach," *J. Amer. Stat. Assoc.*, vol. 84, pp. 805–890, 1989.
[35] A. R. Gallant, "Seemingly unrelated nonlinear regressions," *J. Econometr.*, vol. 3, pp. 35–50, 1975.
[36] H. L. Van Trees, *Detection, Estimation and Modulation Theory*. New York: Wiley, 1968, pt. I.
[37] E. J. Kelly, "An adaptive detection algorithm," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-22, pp. 115–127, Mar. 1986.
[38] E. J. Kelly and K. M. Forsythe, "Adaptive detection and parameter estimation for multidimensional signal models," Lincoln Lab., Mass. Inst. Technol., Cambridge,, Tech. Rep. 848, Apr. 1989.
[39] P. Stoica and M. Viberg, "Maximum likelihood parameter and rank estimation in reduced-rank multivariate linear regressions," *IEEE Trans. Signal Processing*, vol. 44, pp. 3069–3078, Dec. 1996.
[40] J. C. Mosher, P. S. Lewis, and R. M. Leahy, "Multiple dipole modeling and localization from spatiotemporal MEG data," *IEEE Trans. Biomed. Eng.*, vol. 39, pp. 541–557, June 1992.
[41] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood and Cramér–Rao bound–Further results and comparisons," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 2140–2150, Dec. 1990.
[42] P. Stoica, M. Viberg, and B. Ottersten, "Instrumental variable approach to array processing in spatially correlated noise fields," *IEEE Trans. Signal Processing*, vol. 42, pp. 121–133, Jan. 1994.
[43] D. H. Johnson and D. E. Dudgeon, *Array Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
[44] K. Sekihara and B. Scholz, "Generalized Wiener estimation of three-dimensional current distribution from biomagnetic measurements," *IEEE Trans. Biomed. Eng.*, vol. 43, pp. 281–291, Mar. 1996.
[45] A. M. Dale and M. I. Sereno, "Improved localization of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: A linear approach," *J. Cognit. Neurosci.*, vol. 5, no. 2, pp. 162–176, 1993.
[46] W. J. Bangs, "Array processing with generalized beamformers," Ph.D. dissertation, Dept. Elect. Eng., Yale Univ., New Haven, CT, 1971.
[47] J. C. Mosher, M. E. Spencer, R. M. Leahy, and P. S. Lewis, "Error bounds for MEG and EEG dipole source localization," *Electroencephalogr. Clin. Neurophys.*, vol. 86, pp. 303–321, 1993.
[48] P. Stoica and B. C. Ng, "On the Cramér–Rao bound under parametric constraints," *IEEE Signal Processing Lett.*, vol. 5, pp. 177–179, July 1998.
[49] D. Bates, "The derivative of $|X'X|$ and its uses," *Technometrics*, vol. 25, pp. 373–376, Nov. 1983.
[50] G. H. Golub and G. P. H. Styan, "Numerical computations for univariate linear models," *J. Statist. Comput. Simul.*, vol. 2, pp. 253–274, 1973.
[51] B. Turetsky, J. Raz, and G. Fein, "Representation of multi-channel evoked potential data using a dipole component model of intracranial generators: Application to the auditory P300," *Electroencephalogr. Clin. Neurophysiol.*, vol. 76, pp. 540–556, 1990.
[52] M. E. Spencer, "Spatiotemporal modeling, estimation, and detection for EEG and MEG," Ph.D. dissertation, Dept. Elect. Comput. Eng., Univ. Southern Calif., Los Angeles, 1995.
[53] A. B. Geva, H. Pratt, and Y. Y. Zeevi, "Spatiotemporal multiple source localization by wavelet-type decomposition of evoked potentials," *Electroencephalogr. Clin. Neurophysiol.*, vol. 96, pp. 278–286, 1995.
[54] B. Lutkenhöner and O. Steinsträter, "High-precision neuromagnetic study of the functional organization of the human auditory cortex," *Audiol. Neurootol.*, vol. 3, pp. 191–213, 1998.
[55] A. Dogandžić and A. Nehorai, "Localization of evoked electric sources and design of EEG/MEG sensor arrays," in *Proc. 9th IEEE SP Workshop Stat. Signal Array Process.*, Portland, OR, Sept. 1998, pp. 228–231.
[56] A. Nehorai and A. Dogandžić, "Estimation of propagating dipole sources by EEG/MEG sensor arrays," in *Proc. 32nd Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, Nov. 1998, pp. 304–308.
[57] O. Bria, C. Muravchik, and A. Nehorai, "EEG/MEG error bounds for a dynamic dipole source with a realistic head model," *Methods Inform. Med.*, to be published.
[58] A. Nehorai and M. Hawkes, "Performance bounds for estimating vector systems," *IEEE Trans. Signal Processing*, to be published.
[59] C. H. Muravchik and A. Nehorai, "MEG/EEG numerical error bounds for a dipole source with a realistic head model," in *Proc. 19th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Chicago, IL, Oct. 1997, pp. 1233–1236.

Aleksandar Dogandžić (S'96) received the Dipl. Eng. degree (summa cum laude) in electrical engineering from the University of Belgrade, Yugoslavia, in 1995 and the M.S. degree in electrical engineering and computer science from the University of Illinois, Chicago (UIC), in 1997. He is presently a Ph.D. candidate at UIC.

His research interests are in statistical signal processing and its applications to biomedicine, antenna arrays, and wireless communications.

Mr. Dogandžić received the Distinguished Electrical Engineering M.S. Student Award from the Chicago Chapter of the IEEE Communications Society. In 1997, he was awarded the Aileen S. Andrew Foundation Graduate Fellowship.



Arye Nehorai (S'80–M'83–SM'90–F'94) received the B.Sc. and M.Sc. degrees in electrical engineering from the Technion–Israel Institute of Technology, Haifa, in 1976 and 1979, respectively, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, in 1983.

After graduation, he worked as a Research Engineer for Systems Control Technology, Inc., Palo Alto, CA. From 1985 to 1995, he was with the Department of Electrical Engineering, Yale University, New Haven, CT, where he became an Associate Professor in 1989. In 1995, he joined the Department of Electrical Engineering and Computer Science, The University of Illinois, Chicago (UIC), as a Full Professor. He holds a joint professorship with the Bioengineering Department at UIC. His research interests are in signal processing, communications, biomedicine, and the environment.

Dr. Nehorai is the Editor-in-Chief of the IEEE TRANSACTIONS ON SIGNAL PROCESSING and an Associate Editor of the IEEE SIGNAL PROCESSING LETTERS, the IEEE JOURNAL OF OCEANIC ENGINEERING, *Circuits, Systems, and Signal Processing*, and *the Journal of the Franklin Institute.* He is also a Member of the Editorial Board of *Signal Processing* and has previously been an Associate Editor of the IEEE TRANSACTIONS ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING and the IEEE TRANSACTIONS ON ANTENNAS AND PROPAGATION. He served as Chairman of the Connecticut IEEE Signal Processing Chapter from 1986 to 1995 and is currently a Founding Member and the Vice-Chair of the IEEE Signal Processing Society's Technical Committee on Sensor Array and Multichannel (SAM) Processing. He is the Cogeneral Chair of the *First IEEE SAM Signal Processing Workshop* to be held in 2000. He received the 1979/1980 Rothschild Fellowship in science and engineering, which is awarded annually to eight new graduates throughout Israel. He was co-recipient, with P. Stoica, of the 1989 IEEE Signal Processing Society's Senior Award for Best Paper. He has been a Fellow of the Royal Statistical Society since 1996.