

---

# High-Resolution Signal Processing

---

## Preface

High-resolution signal processing is encountered in a wide range of applications, which include in particular localization of objects in certain medium. The medium could be space, air, land, water, or even living tissues. The importance of high-resolution signal processing is indeed recognized in such fields as astronomy, radar, sonar, seismology, and biomedical engineering. High-resolution signal processing aims to retrieve desired information with high accuracy from often very limited data. This is an area closely related to statistical signal processing and spectral analysis. Its mathematical foundation consists mainly of statistics and linear algebra.

---

---

# CONTENTS

<b>1 EEG/MEG SPATIO-TEMPORAL DIPOLE SOURCE ESTIMATION AND SENSOR ARRAY DESIGN</b>	<b>393</b>
1.1 Introduction	393
1.2 Source and Measurement Models	396
1.2.1 Source Model	396
1.2.2 Measurement Model	397
1.3 Maximum Likelihood Estimation	399
1.3.1 Simultaneous Estimation of the Dipole Parameters and Noise Covariance	399
1.3.2 Ordinary and Generalized Least Squares	402
1.3.3 Estimated Generalized Least Squares	403
1.3.4 ML versus OLS	403
1.4 Nonparametric Basis Functions	405
1.5 Scanning Methods	407
1.6 Fisher Information Matrix and Cramér-Rao Bound	409
1.7 Goodness-of-fit Measures	411
1.8 EEG/MEG Sensor Array Design	412
1.8.1 Reparametrization Invariance	414
1.8.2 Relationship between Optimal Array Design and Information Theory	414
1.9 Numerical Examples	416
1.10 Conclusions	427
<b>APPENDICES</b>	<b>428</b>
1.A ML Estimation	428
1.B Parameter Identifiability	429

---

1.C Asymptotic Properties of the OLS Estimates	431
1.D ML versus OLS	432
1.E Nonparametric Basis Functions	433
1.F Scanning	435
1.G Derivation of the Fisher Information Matrix	436
<b>BIBLIOGRAPHY</b>	<b>438</b>
<b>SUBJECT INDEX</b>	<b>444</b>



# EEG/MEG SPATIO-TEMPORAL DIPOLE SOURCE ESTIMATION AND SENSOR ARRAY DESIGN

Aleksandar Dogandžić, *Iowa State University*  
Arye Nehorai, *University of Illinois*

## 1.1 Introduction

The non-invasive techniques of electroencephalography (EEG) and magnetoencephalography (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. This electromagnetic field is generated by neuronal activity in the brain, and provide information about both its spatial distribution and temporal dynamics. This is in contrast to 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 [3].

A significant amount of work has been done on the analysis of brain temporal activity (see [42] and references therein). The most widely used estimator is the sample mean of an ensemble of single evoked responses timelocked to the instant of stimulus application. The mean signal is then fitted to a parametric model. Though exploiting the good temporal resolution of EEG and MEG, these methods do not utilize their spatial resolution.

Spatio-temporal 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 [51], and

are applied to clinical diagnosis in neurology and psychiatry. A current dipole is often used as an equivalent source for a uni-directional 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 [30] as is often satisfied for sources evoked in response to a given sensory stimulus: auditory, visual etc. The dipolar response of the MEG array was first revealed in a study of somatosensory evoked fields [8].

In [54], spatio-temporal 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 [13] extends the above model by allowing the dipole strengths to change arbitrarily. In [55], 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 [7]. 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 [32]. In MEG, environmental noise is an additional important source of spatial correlation [57], especially in an unshielded environment.

One of the first attempts to tackle the problem of correlated noise was by Sekihara *et al.* [57], who assumed known spatial noise covariance. The localization in [57] is performed using a generalized least squares (GLS) method (see also Section 1.3.2) and measurements at only one point in time. In [15], 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 [39] and [40]. The algorithm in [13] is extended in [75] and [76] 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 particular time point of interest. 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 [32], [14], [71]. Thus, the noise covariance may

need to be estimated only from the data containing the response. A major goal of this chapter is to develop algorithms that solve this problem in an efficient way.

An iteratively reweighted generalized least squares (IRGLS) procedure [72, pp. 298–300] is proposed in [32] 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 which, if the noise is Gaussian, converge to the maximum likelihood (ML) estimates. This method, however, does not include temporal evolution. In this chapter (see also [16] and [19]), 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 which 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. 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 to the ordinary least-squares (OLS) methods widely used in the EEG/MEG literature (see e.g. [54] and [13]). 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 1.4, 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 1.5, we derive ML-based methods for scanning the brain response data, which can be used directly for imaging the brain's electromagnetic activity, or to initialize the multi-dimensional search required for obtaining the dipole location estimates.

In Section 1.7 we discuss goodness-of-fit measures which account for spatially correlated noise. We derive the Fisher information matrix (FIM) and Cramér-Rao bound (CRB) for the proposed model in Section 1.6. We then use the CRB to construct methods for EEG/MEG array optimization. The proposed optimization methods are applicable to sensor array optimization in general; for application to radar, see [20].

Finally, in Section 1.9, we compare the estimation accuracy of the ML, GLS, OLS, and scanning methods for simulated data and apply the ML methods to real auditory evoked MEG responses.

## Symbol List

For readers' convenience, symbols used in this chapter are listed below.

$m$	Total number of sensors.
$n$	Number of dipoles.
$r$	Number of detectable moment components per dipole; $r = 3$ except when only MEG sensors are employed ( $r = 2$ ).
$K$	Number of trials.
$N$	Number of snapshots per trial.
$l$	Number of basis functions.
$Y_k$	An $m \times N$ spatio-temporal data matrix collected in the $k$ th trial, $k = 1, \dots, K$ .
$W_k$	An $m \times N$ noise matrix in the $k$ th trial, $k = 1, \dots, K$ .
$\Sigma$	An $m \times m$ spatial noise covariance matrix.
$A(\boldsymbol{\theta})$	An $m \times nr$ array response matrix.
$\boldsymbol{\theta}$	Vector of dipole location parameters.
$\mathbf{s}(t)$	Vector of dipole moment components at time $t$ , $t = 1, \dots, N$ .
$X$	An $nr \times l$ matrix of basis-function coefficients.
$\Phi(\boldsymbol{\eta})$	An $l \times N$ basis-function matrix.
$\boldsymbol{\eta}$	Vector of basis-function parameters.
$\xrightarrow{\text{a.s.}}$	Almost sure convergence.
$\xrightarrow{\text{d}}$	Convergence in distribution.

## 1.2 Source and Measurement Models

### 1.2.1 Source Model

We model the head as a spherically symmetric conductor locally fitted to the head curvature. This model is used in most clinical and research applications of EEG and MEG. It is often a reasonably good approximation, particularly for MEG, which is less sensitive than EEG to modeling inaccuracies, see e.g. [3]. (Note that the source estimation algorithms presented in this chapter can be applied to realistic head models as well.)

Let  $\mathbf{p}$  be the position of a current dipole source relative to the center of the sphere

$$\mathbf{p} = p [\sin \vartheta \cos \varphi, \sin \vartheta \sin \varphi, \cos \vartheta]^T, \quad (1.2.1)$$

where

- $\vartheta$  is the dipole's elevation,
- $\varphi$  its azimuth, and
- $p$  its distance from the center; see Figure 1.1.

Thus,  $\mathbf{p}$  is fully described by

$$\boldsymbol{\theta} = [\vartheta, \varphi, p]^T. \quad (1.2.2)$$

The vectors

$$\begin{aligned} \mathbf{u}_\vartheta &= [\cos \vartheta \cos \varphi, \cos \vartheta \sin \varphi, -\sin \vartheta]^T, \\ \mathbf{u}_\varphi &= [-\sin \varphi, \cos \varphi, 0]^T, \\ \mathbf{u}_p &= \mathbf{p}/p \end{aligned} \quad (1.2.3)$$

form an orthonormal basis (see also [31]). Using this basis, the dipole moment can be written as  $s_\vartheta \mathbf{u}_\vartheta + s_\varphi \mathbf{u}_\varphi + s_p \mathbf{u}_p$ . We define the vector of moment parameters  $\mathbf{s} = [s_\vartheta, s_\varphi, s_p]^T$ .

## 1.2.2 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  $m = m_E + m_B$ . Then, the  $m$ -dimensional measurement vector of this array is

$$\mathbf{y} = A(\boldsymbol{\theta})\mathbf{s} + \mathbf{w}, \quad (1.2.4)$$

where  $\mathbf{y} = [\mathbf{y}_B^T, \mathbf{y}_E^T]^T$ ,  $A(\boldsymbol{\theta})$  is the  $m \times 3$  array response matrix and  $\mathbf{w} = [\mathbf{w}_B^T, \mathbf{w}_E^T]^T$  is additive noise. The array response matrix is derived using the quasistatic approximation of Maxwell's equations and spherical head model (see [30], [77], and references therein). The quasistatic approximation is justified by the fact that the useful frequency spectrum for electrophysiological signals in EEG/MEG is typically below 1 kHz. As a consequence

- the capacitive component of the tissue impedance is negligible [56], and
- electromagnetic wave effects in the head can be neglected, i.e. the signals are assumed to propagate effectively at infinite velocity and the currents everywhere in the head depend only on the sources at that instant of time (and are independent of previous history).

The radial component of a dipole produces no external magnetic field in the spherical head model [30], so the last column of the MEG response matrix is zero. Thus,

$$A(\boldsymbol{\theta}) = [[A_B(\boldsymbol{\theta}), \mathbf{0}_{m_B \times 1}]^T, A_E(\boldsymbol{\theta})^T]^T, \quad (1.2.5)$$

where  $A_B(\boldsymbol{\theta})$  and  $A_E(\boldsymbol{\theta})$  are the MEG and EEG response matrices with dimensions  $m_B \times 2$  and  $m_E \times 3$ , respectively. The symbol  $\mathbf{0}_{m_B \times 1}$  denotes the  $m_B \times 1$  vector with zero entries. Define  $r = \text{rank}(A(\boldsymbol{\theta}))$ ; usually  $r = 3$  except when only MEG sensors are employed ( $r = 2$ ).

The following model is used for  $n$  distinct dipoles:

$$\mathbf{y} = [A(\boldsymbol{\theta}_1) \cdots A(\boldsymbol{\theta}_n)][\mathbf{s}_1^T \cdots \mathbf{s}_n^T]^T + \mathbf{w}, \quad (1.2.6)$$

**Figure 1.1.** Dipole in a sphere.

where  $A(\boldsymbol{\theta}_i)$ ,  $\boldsymbol{\theta}_i$ , and  $\mathbf{s}_i$ ,  $i = 1, 2, \dots, n$ , are the array response matrix, vector of location parameters, and moment vector of the  $i$ th dipole, respectively. Observe that the above equation is a special case of (1.2.4) with  $A(\boldsymbol{\theta})$ ,  $\boldsymbol{\theta}$ , and  $\mathbf{s}$  substituted with  $[A(\boldsymbol{\theta}_1) \cdots A(\boldsymbol{\theta}_n)]$ ,  $\boldsymbol{\theta} = [\boldsymbol{\theta}_1^T \cdots \boldsymbol{\theta}_n^T]^T$ , and  $\mathbf{s} = [\mathbf{s}_1^T \cdots \mathbf{s}_n^T]^T$ , respectively. Note that in this case,  $\boldsymbol{\theta}$  and  $\mathbf{s}$  are  $3n$ - and  $nr$ -dimensional vectors, respectively. Since the dipoles are at distinct locations, we assume that  $A(\boldsymbol{\theta})$  has full rank equal to  $nr$ .

The noise vector  $\mathbf{w}$  is assumed to be zero-mean with unknown spatial covariance  $\Sigma$ , whereas the source moment signal is deterministic. Thus, the mean and covariance matrix of the snapshot  $\mathbf{y}$  are  $A(\boldsymbol{\theta})\mathbf{s}$  and  $\Sigma$ , respectively. The noise is predominantly due to background activity in neurons. An assumption of Gaussianity, 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 [25]. Tests for the normality of background EEG signals have also been developed in [43].

### 1.3 Maximum Likelihood Estimation

#### 1.3.1 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 which may need to be validated in practice: homogeneity tests for the evoked responses have been derived in [44]]. In the  $k$ th trial ( $k = 1, \dots, K$ ),  $N$  temporal data vectors (snapshots)  $\mathbf{y}_k(1), \mathbf{y}_k(2), \dots, \mathbf{y}_k(N)$  are collected. We refer to the matrix

$$Y_k = [\mathbf{y}_k(1) \cdots \mathbf{y}_k(N)] \quad (1.3.1)$$

as the *spatio-temporal 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

$$\mathbf{s}(t) = X\boldsymbol{\phi}(t, \boldsymbol{\eta}), \quad (1.3.2)$$

where  $X$  is a full-rank matrix of unknown coefficients with dimensions  $nr \times l$  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. In the above model, the number of basis functions  $l$  should be smaller than the number of snapshots  $N$ ; otherwise, we could simply choose  $\Phi = I_N$ , see also Section 1.4. The

measurement model is then

$$\mathbf{y}_k(t) = A(\boldsymbol{\theta})X\boldsymbol{\phi}(t, \boldsymbol{\eta}) + \mathbf{w}_k(t), \quad t = 1, \dots, N, \quad k = 1, \dots, K, \quad (1.3.3)$$

where the noise vectors  $\mathbf{w}_k(t)$  are assumed to be zero mean with unknown spatial covariance  $\Sigma$ , 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  $\Sigma$  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 [50] (see also [36], [58, chapter 6.4], and [72, chapter 5]). In statistics, it is usually applied to fitting growth curves and thus is also called the growth-curve model [50]–[58].

Equations (1.3.3) can be written compactly as

$$Y_k = A(\boldsymbol{\theta})X\Phi(\boldsymbol{\eta}) + W_k, \quad k = 1, \dots, K, \quad (1.3.4)$$

where  $W_k = [\mathbf{w}_k(1) \ \mathbf{w}_k(2) \ \cdots \ \mathbf{w}_k(N)]$  is the *noise matrix* and

$$\Phi(\boldsymbol{\eta}) = [\boldsymbol{\phi}(1, \boldsymbol{\eta}) \ \boldsymbol{\phi}(2, \boldsymbol{\eta}) \ \cdots \ \boldsymbol{\phi}(N, \boldsymbol{\eta})] \quad (1.3.5)$$

is the *basis-function matrix*. Define the projection matrix onto the row space of  $\Phi(\boldsymbol{\eta})$  as

$$\Pi_\phi(\boldsymbol{\eta}) = \Phi(\boldsymbol{\eta})^T [\Phi(\boldsymbol{\eta})\Phi(\boldsymbol{\eta})^T]^{-1} \Phi(\boldsymbol{\eta}). \quad (1.3.6)$$

In Appendix 1.A, we extend the GMANOVA equations to multiple trials. We show that, for known  $\boldsymbol{\theta}$  and  $\boldsymbol{\eta}$  and Gaussian noise  $\mathbf{w}_k(t)$ , the ML estimates of  $X$  and  $\Sigma$  are

$$\hat{X} = [A(\boldsymbol{\theta})^T S^{-1} A(\boldsymbol{\theta})]^{-1} A(\boldsymbol{\theta})^T S^{-1} \bar{Y} \Phi(\boldsymbol{\eta})^T [\Phi(\boldsymbol{\eta})\Phi(\boldsymbol{\eta})^T]^{-1}, \quad (1.3.7a)$$

$$\hat{\Sigma}(\boldsymbol{\theta}, \boldsymbol{\eta}) = S + (1/N) \cdot (I_m - TS^{-1}) \bar{Y} \Pi_\phi(\boldsymbol{\eta}) \bar{Y}^T (I_m - TS^{-1})^T, \quad (1.3.7b)$$

where

$$\bar{Y} = \frac{1}{K} \sum_{k=1}^K Y_k, \quad (1.3.8a)$$

$$S = \hat{R} - \frac{1}{N} \cdot \bar{Y} \Pi_\phi(\boldsymbol{\eta}) \bar{Y}^T, \quad (1.3.8b)$$

$$\hat{R} = \frac{1}{NK} \sum_{k=1}^K Y_k Y_k^T, \quad (1.3.8c)$$

$$T = A(\boldsymbol{\theta}) [A(\boldsymbol{\theta})^T S^{-1} A(\boldsymbol{\theta})]^{-1} A(\boldsymbol{\theta})^T, \quad (1.3.8d)$$

and  $I_m$  denotes the identity matrix of size  $m$ . Note that  $S$  is a function 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 chapter. For the above model (and under the

Gaussianity assumption), the sufficient statistics are  $\bar{Y}$  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  $\hat{\Sigma} = S = 1/K \sum_{k=1}^K (y_k - \bar{y})^2$  and  $\hat{X} = \bar{y}/a$ .

If  $\Sigma$  is known, the ML estimate of  $X$  is simply

$$\hat{X} = [A(\boldsymbol{\theta})^T \Sigma^{-1} A(\boldsymbol{\theta})]^{-1} A(\boldsymbol{\theta})^T \Sigma^{-1} \bar{Y} \Phi(\boldsymbol{\eta})^T [\Phi(\boldsymbol{\eta}) \Phi(\boldsymbol{\eta})^T]^{-1}, \quad (1.3.9)$$

as can easily be shown by differentiating the log-likelihood function (see Appendix 1.A) using the identity  $\frac{\partial}{\partial X} \text{tr}(P^T X Q) = P Q^T$  [52, p. 72].

If  $\boldsymbol{\theta}$  and  $\boldsymbol{\eta}$  are not known (in addition to  $X$  and  $\Sigma$ ), their ML estimates  $\hat{\boldsymbol{\theta}}$  and  $\hat{\boldsymbol{\eta}}$  are obtained by maximizing the concentrated likelihood function (see Appendix 1.A)

$$l_{\text{ML}}(\boldsymbol{\theta}, \boldsymbol{\eta}) = \frac{\hat{R}}{|\hat{\Sigma}(\boldsymbol{\theta}, \boldsymbol{\eta})|} \quad (1.3.10a)$$

$$= \frac{|\Phi(\boldsymbol{\eta})[I_N - (1/N)\bar{Y}^T Q(\boldsymbol{\theta})\bar{Y}]\Phi(\boldsymbol{\eta})^T|}{|\Phi(\boldsymbol{\eta})[I_N - (1/N)\bar{Y}^T \hat{R}^{-1}\bar{Y}]\Phi(\boldsymbol{\eta})^T|} \quad (1.3.10b)$$

$$= \frac{|A(\boldsymbol{\theta})^T S^{-1} A(\boldsymbol{\theta})|}{|A(\boldsymbol{\theta})^T \hat{R}^{-1} A(\boldsymbol{\theta})|}, \quad (1.3.10c)$$

where

$$Q(\boldsymbol{\theta}) = \hat{R}^{-1} - \hat{R}^{-1} A(\boldsymbol{\theta}) [A(\boldsymbol{\theta})^T \hat{R}^{-1} A(\boldsymbol{\theta})]^{-1} A(\boldsymbol{\theta})^T \hat{R}^{-1} \quad (1.3.11)$$

and  $|\cdot|$  denotes the determinant. The above concentrated likelihood function is written in the form of a generalized likelihood ratio (GLR) test statistic (see e.g. [69] and [52, p. 418] for the definition of GLR) for testing  $H_0 : X = 0$  versus  $H_1 : X \neq 0$ . To find the ML estimates of  $X$  and  $\Sigma$ , substitute  $\boldsymbol{\theta}$  and  $\boldsymbol{\eta}$  in (1.3.7a) and (1.3.7b) by  $\hat{\boldsymbol{\theta}}$  and  $\hat{\boldsymbol{\eta}}$ . Efficient methods for computing (1.3.10) are derived in [19, sec. VII].

If  $\mathbf{s}(t)$  is modeled as a linear combination of  $l$  basis functions without any prior on their shape (i.e. nonparametric basis functions, where  $\Phi(\boldsymbol{\eta}) = \Phi$  is completely unknown), we can concentrate (1.3.10) with respect to  $\Phi$  as well, as shown in Section 1.4. 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 [70], where a concentrated likelihood function of a similar form was obtained and applied to direction-of-arrival (DOA) estimation. Also, a signal subspace fitting (SSF) criterion approximating this likelihood function was proposed. Here (see also [19], [16]), we consider a more general model with multiple trials (suitable for analyzing evoked responses), vector source signals, and parametric and nonparametric basis functions. This formulation also includes as special cases the ML radar array processing methods in [59] and [18]. For a discussion on identifiability of the unknown parameters, see Appendix 1.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 non-linear 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 1.4.

The above estimators have good asymptotic properties even when the noise is not Gaussian: Theorem 7.1 in Subsection 1.3.4 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 referring to these estimators as the ML estimators. Having a similar terminology problem, some authors refer to these methods as extended least squares (ELS).

### 1.3.2 Ordinary and Generalized Least Squares

The (nonlinear) ordinary least squares method [72, pp. 447–448] applied to the above model gives the residual sum of squares

$$l_{\text{OLS}}(\boldsymbol{\theta}, \boldsymbol{\eta}) = \text{tr} \left\{ \widehat{R} - \frac{1}{N} A(\boldsymbol{\theta}) [A(\boldsymbol{\theta})^T A(\boldsymbol{\theta})]^{-1} A(\boldsymbol{\theta})^T \bar{Y} \Pi_{\Phi}(\boldsymbol{\eta}) \bar{Y}^T \right\} \quad (1.3.12)$$

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 (1.3.9), (1.A.4a), and (1.A.4b) into the likelihood function in (1.A.1); 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  $\widehat{R}$  does not affect the above minimization. If  $K = 1$  and  $\Phi = I_N$ , i.e.  $\mathbf{s}(t)$  is an arbitrary vector at each time point  $t = 1, \dots, N$ , this method coincides with the deterministic maximum likelihood in e.g. [60].

The generalized least squares indexgeneralized least squares (GLS) (GLS) method [72, pp. 448–449] is the ML method for spatially correlated Gaussian noise with known spatial covariance  $\Sigma$ . It is a simple extension of OLS, since it reduces to applying OLS to the spatially pre-whitened data, yielding the following cost function:

$$l_{\text{GLS}}(\boldsymbol{\theta}, \boldsymbol{\eta}) = \text{tr} \left[ \Sigma^{-1} \widehat{R} - \frac{1}{N} \Sigma^{-1} A(\boldsymbol{\theta}) [A(\boldsymbol{\theta})^T \Sigma^{-1} A(\boldsymbol{\theta})]^{-1} A(\boldsymbol{\theta})^T \Sigma^{-1} \bar{Y} \Pi_{\Phi}(\boldsymbol{\eta}) \bar{Y}^T \right]. \quad (1.3.13)$$

Detection methods for this case with  $\Phi = I_N$  are derived in [15].

### 1.3.3 Estimated Generalized Least Squares

The estimated generalized least squares (EGLS) method [72, pp. 449–450] (which is an approximation of the ML method) is obtained by substituting the noise covariance  $\Sigma$  in the GLS cost function by its strongly and  $\sqrt{K}$ -consistent estimator  $\tilde{\Sigma}$ , i.e.  $\tilde{\Sigma}$  needs to satisfy

$$\tilde{\Sigma} \xrightarrow{\text{a.s.}} \Sigma, \quad (1.3.14\text{a})$$

$$\tilde{\Sigma} = \Sigma + O_p(K^{-1/2}), \quad (1.3.14\text{b})$$

where  $\xrightarrow{\text{a.s.}}$  indicates almost sure convergence and  $O_p(K^{-1/2})$  denotes a sequence of random variables that is bounded in probability by  $K^{-1/2}$ , see also the discussion in Section 1.3.4. Therefore, the EGLS cost function is

$$l_{\text{EGLS}}(\boldsymbol{\theta}, \boldsymbol{\eta}) = \text{tr} \left[ \tilde{\Sigma}^{-1} \hat{R} - \frac{1}{N} \tilde{\Sigma}^{-1} A(\boldsymbol{\theta}) [A(\boldsymbol{\theta})^T \tilde{\Sigma}^{-1} A(\boldsymbol{\theta})]^{-1} A(\boldsymbol{\theta})^T \tilde{\Sigma}^{-1} \bar{Y} \Pi_{\Phi}(\boldsymbol{\eta}) \bar{Y}^T \right]. \quad (1.3.15)$$

For example, a good choice for an estimator of  $\Sigma$  is

$$\tilde{\Sigma} = \hat{R} - \frac{1}{N} \bar{Y} \bar{Y}^T, \quad (1.3.16)$$

which is strongly and  $\sqrt{K}$ -consistent, provided that the noise  $\mathbf{w}_k(t)$  has finite fourth-order moments. Based on [24, chapter 5.6], it follows that ML and EGLS are asymptotically equivalent, provided that the measurement model (1.3.3) is correct. Differences between ML and EGLS approaches arise for small sample sizes and when modeling inaccuracies occur: ML incorporates these inaccuracies into the estimated spatial noise covariance whereas EGLS does not. Consequently, we adopt the ML approach in this chapter. (It is particularly important to use the ML approach for deriving the scanning methods in Section 1.5, since these methods are based on an inaccurate, single-dipole model.)

If  $\boldsymbol{\eta}$  is known, it is possible to exploit information contained in  $\Phi$  and choose  $\tilde{\Sigma} = S$  [see (1.3.8b)] as an estimator of  $\Sigma$  [assuming finite fourth-order moments of  $\mathbf{w}_k(t)$ ]. For a single trial ( $K = 1$ ),  $N \rightarrow \infty$  is needed to study the asymptotic properties, and it is easy to show that  $S$  is a strongly and  $\sqrt{N}$ -consistent estimator of  $\Sigma$ . This EGLS estimator results in the SSF method in [70] (where its asymptotic equivalence with the ML method was also shown).

### 1.3.4 ML versus OLS

In this section we show consistency and asymptotic normality of the ML estimates, as well as consistency of the OLS estimates. In Appendix 1.D, we show that the ML estimates are asymptotically more efficient than the OLS estimates.

Define  $\boldsymbol{\rho} = [\text{vec}(X)^T, \boldsymbol{\theta}^T, \boldsymbol{\eta}^T]^T$  and  $\boldsymbol{\psi} = \text{vech}(\Sigma)$ . Here, the  $\text{vech}$  operator creates a single column vector by stacking the elements below the main diagonal columnwise. Then, the vector of all unknown parameters is  $\boldsymbol{\gamma} = [\boldsymbol{\rho}^T, \boldsymbol{\psi}^T]^T$ . Also, let us assume that the true values of the parameter vectors  $\boldsymbol{\rho}$  and  $\boldsymbol{\gamma}$  are  $\boldsymbol{\rho}_0 = [\text{vec}(X_0)^T, \boldsymbol{\theta}_0^T, \boldsymbol{\eta}_0^T]^T$  and  $\boldsymbol{\gamma}_0 = [\boldsymbol{\rho}_0^T, \boldsymbol{\psi}_0^T]^T$ , respectively.

Let  $\mathcal{I}_{\text{signal}}(\boldsymbol{\gamma})$  be the Fisher information matrix of the signal parameters  $\boldsymbol{\rho}$ . The exact expression will be given in Section 1.6, see equations (1.6.2) and (1.6.3). 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  $\mathbf{w}_k(t)$ ,  $k = 1, \dots, K$ ,  $t = 1, \dots, N$  are independent, identically distributed (i.i.d.) with zero mean and arbitrary positive definite covariance  $\Sigma$ ,
- R3)  $A(\boldsymbol{\theta})$  and  $\Phi(\boldsymbol{\eta})$  are continuous and have continuous first and second partial derivatives with respect to  $\boldsymbol{\theta}$  and  $\boldsymbol{\eta}$ ,
- R4) The matrix  $\mathcal{I}_{\text{signal}}(\boldsymbol{\gamma})$  is non-singular,
- R5)  $\text{tr} \left\{ \Sigma^{-1} [A(\boldsymbol{\theta})X\Phi(\boldsymbol{\eta}) - A(\boldsymbol{\theta}_0)X_0\Phi(\boldsymbol{\eta}_0)] [A(\boldsymbol{\theta})X\Phi(\boldsymbol{\eta}) - A(\boldsymbol{\theta}_0)X_0\Phi(\boldsymbol{\eta}_0)]^T \right\} = 0$  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 Appendix 1.B). Observe that the above conditions do not require specific distributional assumptions on the noise vectors  $\mathbf{w}_k(t)$ .

Before proceeding with the asymptotic results, let us introduce some notation:

$$z_n = O_p(a_n) \quad (1.3.17)$$

denotes that a sequence of random variables  $\{z_n\}$  is bounded in probability by a sequence of positive real numbers  $\{a_n\}$  (see e.g. [72, p. 34]). Also,

$$z_n = o_p(a_n) \quad (1.3.18)$$

implies that  $z_n/a_n$  converges to zero in probability.

**Theorem 1.3.1:** *Under the regularity conditions R1)–R5), the ML estimate of  $\boldsymbol{\rho}$  satisfies (as  $K \rightarrow \infty$ )*

$$\hat{\boldsymbol{\rho}} \xrightarrow{\text{a.s.}} \boldsymbol{\rho}_0, \quad (1.3.19a)$$

$$\hat{\boldsymbol{\rho}} = \boldsymbol{\rho}_0 + O_p(K^{-1/2}), \quad (1.3.19b)$$

and

$$\sqrt{NK} (\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}_0) \stackrel{d}{\rightarrow} \mathcal{N}(\mathbf{0}, NK\mathcal{I}_{\text{signal}}(\boldsymbol{\gamma}_0)^{-1}), \quad (1.3.20)$$

where  $\stackrel{d}{\rightarrow}$  indicates convergence in distribution.

*Proof:* The proof follows from [24, chapter 5.6] and [27], where it is shown for a more general case, see also [70], [72, pp. 300–301], [23].

**Theorem 1.3.2:** *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  $K \rightarrow \infty$ )*

$$\widehat{\boldsymbol{\rho}}_{\text{OLS}} \xrightarrow{\text{a.s.}} \boldsymbol{\rho}_0, \quad (1.3.21\text{a})$$

$$\begin{aligned} \widehat{\boldsymbol{\rho}}_{\text{OLS}} &= \boldsymbol{\rho}_0 + \left[ K \sum_{t=1}^N D(t, \boldsymbol{\rho}_0)^T D(t, \boldsymbol{\rho}_0) \right]^{-1} \\ &\quad \times \sum_{k=1}^K \sum_{t=1}^N D(t, \boldsymbol{\rho}_0)^T \mathbf{w}_k(t) + o_p(K^{-1/2}), \end{aligned} \quad (1.3.21\text{b})$$

$$= \boldsymbol{\rho}_0 + O_p(K^{-1/2}), \quad (1.3.21\text{c})$$

where

$$D(t, \boldsymbol{\rho}) = \frac{\partial(A(\boldsymbol{\theta})X\phi(t, \boldsymbol{\eta}))}{\partial \boldsymbol{\rho}^T}. \quad (1.3.22)$$

Here  $K \sum_{t=1}^N D(t, \boldsymbol{\rho})^T D(t, \boldsymbol{\rho})$  equals  $\mathcal{I}_{\text{signal}}(\boldsymbol{\gamma})$  in equations (1.6.2) and (1.6.3) when  $\Sigma = I_m$ .

*Proof:* See Appendix 1.C.

If  $\mathbf{w}_k(t)$  are i.i.d. normal, it can be shown that the ML estimate  $\widehat{\boldsymbol{\rho}}$  is asymptotically efficient in the sense of first-order efficiency [52, Sections 5c.2 and 5f.2].

In Appendix 1.D 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  $\widehat{R}$ , see also Subsection 1.3.2.

## 1.4 Nonparametric Basis Functions

In this section, we present ML estimation for nonparametric basis function model

$$Y_k = A(\boldsymbol{\theta})X\Phi + W_k, \quad k = 1, \dots, K, \quad (1.4.1)$$

where *both*  $X$  and  $\Phi$  are unknown matrices of full rank  $l$ . This is equivalent to assuming that the spatio-temporal dipole moment matrix  $[\mathbf{s}(1)\mathbf{s}(2)\cdots\mathbf{s}(N)] = X\Phi$  is unknown with rank  $l$ . Hence,  $l$  is a measure of the level of correlation between the moment components, see below. We exploit 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 unknown  $\Phi$  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 1.6. This result is also confirmed by simulation results, see Section 1.9, Example 7.1.

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 1.E.

In Appendix 1.E, we derive the concentrated likelihood function for the nonparametric basis functions by maximizing (1.3.10b) with respect to  $\Phi$  using the Poincaré separation theorem [52, pp. 64–65]. The resulting concentrated likelihood function is given by the product of the  $l$  largest generalized eigenvalues of the matrices  $I_N - (1/N) \cdot \bar{Y}^T Q(\boldsymbol{\theta}) \bar{Y}$  and  $I_N - (1/N) \cdot \bar{Y}^T \hat{R}^{-1} \bar{Y}$ , where  $Q(\boldsymbol{\theta})$  is defined in (1.3.11). The rows of an ML estimate  $\hat{\Phi}$  are the corresponding generalized eigenvectors of the above two matrices. Assuming  $nr \leq N$  (which holds in most practical applications), there can be only  $\text{rank}[A(\boldsymbol{\theta})] = nr$  generalized eigenvalues greater than one (and the rest are equal to one), thus  $1 \leq l \leq nr$ , see also (1.E.10) in Appendix 1.E. If  $l = 1$  all the components of the dipole moments have the same temporal evolution (up to a scaling factor), 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

$$l_{\text{ML}}(\boldsymbol{\theta}) = \frac{|I_N - (1/N) \cdot \bar{Y}^T Q(\boldsymbol{\theta}) \bar{Y}|}{|I_N - (1/N) \cdot \bar{Y}^T \hat{R}^{-1} \bar{Y}|}, \quad (1.4.2)$$

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 concentrated likelihood function for known basis functions in the form of Dirac pulses, i.e.  $\Phi = I_N$  [see (1.3.10b)]. 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  $\hat{\Phi}$  are not unique. However, in Appendix 1.E, we show that, regardless of which ML estimate of  $\Phi$  is chosen, the concentrated likelihood function and the estimated dipole moment temporal evolution  $\hat{X}\hat{\Phi} = [\hat{\mathbf{s}}(1)\hat{\mathbf{s}}(2)\cdots\hat{\mathbf{s}}(N)]$  are unique. The orthonormal set of the ML estimates of the basis functions can be constructed from the above ML estimates as  $\hat{\Phi}_{\text{orth}} = [\hat{\Phi}\hat{\Phi}^T]^{-1/2}\hat{\Phi}$ . 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 this chapter.

Consider now the case where the data set of each trial contains a part with baseline data. Thus,  $Y_k = [Y_{1k}, Y_{2k}]$ ,  $k = 1, \dots, K$ , where  $Y_{1k}$  is a spatio-temporal

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  $A(\boldsymbol{\theta})X\Phi_2$  corrupted by noise. The statistical properties of the noise, described in Section 1.3, are assumed to be the same for both  $Y_{1k}$  and  $Y_{2k}$ . Thus,  $\Phi = [0, \Phi_2]$ ,  $N = N_1 + N_2$ , and  $\bar{Y} = [\bar{Y}_1, \bar{Y}_2]$ . A simple extension of the above results shows that the concentrated likelihood function  $l_{\text{ML}}(\boldsymbol{\theta})$  is the product of the  $l$  largest generalized eigenvalues of  $I_{N_2} - (1/N) \cdot \bar{Y}_2^T Q(\boldsymbol{\theta}) \bar{Y}_2$  and  $I_{N_2} - (1/N) \cdot \bar{Y}_2^T \hat{R}^{-1} \bar{Y}_2$  (see Appendix 1.E). This concentrated likelihood function can be viewed as an extension of the detectors in [34] and [35].

It is also possible to estimate a nonparametric array response matrix  $A$  if the basis functions  $\Phi$  are known or parametric, see [35, p. 25] and [61]. 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 [59] and [18], and in wireless communications for channel estimation and synchronization, see [21]. In this chapter, we apply the nonparametric array model to derive a MUSIC-like scanning method, which is the first application of this model to EEG/MEG, see the following section and Appendix 1.F.

## 1.5 Scanning Methods

Using the ML estimation results in Sections 1.3 and 1.4, we derive two scanning schemes, based on maximizing suitably chosen functions of the data and the single-dipole array response, thus reducing the dimensionality of the problem compared with the multiple-dipole location algorithms. In the EEG/MEG literature, scanning has often been performed using the MUSIC algorithm [62]. However, MUSIC does not perform well when the sources are correlated [60], [63] or the noise is spatially correlated [64], or both. The scanning algorithms proposed here take into account spatially correlated noise with unknown covariance. The first scheme consists of maximizing the concentrated likelihood function for a single dipole. The second scheme is based on matching the estimated column space of the array response matrix with a single-dipole array response using a MUSIC-like function.

If the dipole moments are assumed to be uncorrelated ( $\Phi = I_N$ ), the first scanning scheme reduces to computing

$$l_{\text{scan}}(\boldsymbol{\theta}) = \frac{|I_N - (1/N) \cdot \bar{Y}^T Q(\boldsymbol{\theta}) \bar{Y}|}{|I_N - (1/N) \cdot \bar{Y}^T \hat{R}^{-1} \bar{Y}|} \quad (1.5.1a)$$

$$= \frac{|A(\boldsymbol{\theta})^T [\hat{R} - (1/N) \cdot \bar{Y} \bar{Y}^T]^{-1} A(\boldsymbol{\theta})|}{|A(\boldsymbol{\theta})^T \hat{R}^{-1} A(\boldsymbol{\theta})|} \quad (1.5.1b)$$

using the array response  $A(\boldsymbol{\theta})$  for a single dipole [see (1.3.10) and (1.4.2)]. Whether (1.5.1a) or (1.5.1b) should be used for scanning depends on a particular application. For example, if  $r < N$  (which is typically the case), (1.5.1b) is computationally more

efficient. Expression (1.5.1b) can be viewed as the ratio of the Capon spectral estimate for location  $\boldsymbol{\theta}$  using the data  $Y$ , and the Capon spectral estimate in the direction  $\boldsymbol{\theta}$  using the projection of  $Y$  onto the space orthogonal to the row space of  $\Phi(\boldsymbol{\eta})$ . For a source with fixed orientation in time, the dipole moments are fully correlated, and the rank of the moment matrix is  $l = 1$ . Then, the concentrated likelihood is

$$l_{\text{scan}}(\boldsymbol{\theta}) = \lambda_{\text{MAX}}[I_N - (1/N) \cdot \bar{Y}^T Q(\boldsymbol{\theta}) \bar{Y}, I_N - (1/N) \cdot \bar{Y}^T \hat{R}^{-1} \bar{Y}], \quad (1.5.2)$$

where  $\lambda_{\text{MAX}}[\cdot, \cdot]$  denotes the largest generalized eigenvalue of the two matrices given in the parenthesis. In the special case where only one time snapshot is used [i.e.  $N = 1$ ,  $Y_k = \mathbf{y}_k$ , and  $\bar{Y} = \bar{\mathbf{y}} = (1/K) \sum_{k=1}^K \mathbf{y}_k$ ], and after a linear transformation, the concentrated likelihood (1.5.1a) simplifies to

$$l_{\text{scan}}(\boldsymbol{\theta}) = \bar{\mathbf{y}}^T \hat{R}^{-1} A(\boldsymbol{\theta}) [A(\boldsymbol{\theta})^T \hat{R}^{-1} A(\boldsymbol{\theta})]^{-1} A(\boldsymbol{\theta})^T \hat{R}^{-1} \bar{\mathbf{y}}. \quad (1.5.3)$$

The scanning procedures in (1.5.1)–(1.5.3) require only a 3-D search over the space of dipole location parameters. They are intuitively appealing, since they evaluate the likelihood of a dipole at a particular location, while simultaneously estimating the unknown noise covariance, which accounts for the sources of brain activity at other locations.

If a realistic head model is used, we can impose anatomic constraints [12], i.e. assume that sources can lie only on the surface of the cortex with moments orthogonal to the cortex. Then, we can recast the array response matrix by incorporating the known orientation, which gives a response vector  $\mathbf{a}(\boldsymbol{\theta})$ , where  $\boldsymbol{\theta}$  is a  $2 \times 1$  parameter vector describing a location on the surface of the cortex. In this case, the dimensionality of the search is only 2-D, and the scanning function simplifies to

$$\begin{aligned} l_{\text{scan}}(\boldsymbol{\theta}) &= 1 + \frac{1}{N \mathbf{a}(\boldsymbol{\theta})^T \hat{R}^{-1} \mathbf{a}(\boldsymbol{\theta})} \cdot \mathbf{a}(\boldsymbol{\theta})^T \hat{R}^{-1} \bar{Y} [I_N - \frac{1}{N} \bar{Y}^T \hat{R}^{-1} \bar{Y}]^{-1} \bar{Y}^T \hat{R}^{-1} \mathbf{a}(\boldsymbol{\theta}) \\ &= \frac{\mathbf{a}(\boldsymbol{\theta})^T [\hat{R} - \frac{1}{N} \cdot \bar{Y} \bar{Y}^T]^{-1} \mathbf{a}(\boldsymbol{\theta})}{\mathbf{a}(\boldsymbol{\theta})^T \hat{R}^{-1} \mathbf{a}(\boldsymbol{\theta})}, \end{aligned} \quad (1.5.4)$$

which follows by substituting  $A(\boldsymbol{\theta}) = \mathbf{a}(\boldsymbol{\theta})$  into (1.5.1) and using the formula for the determinant of a partitioned matrix [72, result v at p. 8]. The performance of the above scanning method relies on the validity of the above constraints and would require using patient-specific MRI images to extract the necessary information (e.g. the surface of the cortex).

We now propose the second scanning scheme based on matching the estimated subspace of the array response matrix with a single-dipole array response. The matching is performed using a MUSIC-like function (see Appendix 1.F):

$$l_{\text{MUSIC}}(\boldsymbol{\theta}) = \frac{1}{\lambda_{\text{MIN}}[A(\boldsymbol{\theta})^T [\hat{R}^{-1} - U_{nr} U_{nr}^T] A(\boldsymbol{\theta}), A(\boldsymbol{\theta})^T \hat{R}^{-1} A(\boldsymbol{\theta})]}, \quad (1.5.5)$$

where

- $\lambda_{\min}[\cdot, \cdot]$  denotes the smallest generalized eigenvalue of the two matrices given in the parenthesis,
- $A(\boldsymbol{\theta})$  is a single-dipole array response matrix, and
- $U_{nr}$  is the matrix whose columns are the generalized eigenvectors of  $(1/N) \cdot \overline{Y Y^T}$  and  $\widehat{R}$ , corresponding to their largest  $nr$  generalized eigenvalues, normalized such that  $U_{nr}^T \widehat{R} U_{nr} = I_{nr}$ .

We can compute  $U_{nr}$  as follows:  $U_{nr} = \widehat{R}^{-1/2} V_{nr}$ , where  $V_{nr}$  is the matrix whose columns are the orthonormal eigenvectors of  $\widehat{R}^{-1/2} \overline{Y Y^T} \widehat{R}^{-1/2}$  that correspond to its  $nr$  largest eigenvalues, see Appendix 1.F. Unlike the first scanning scheme, this method requires specifying the number of sources  $n$ . The second scanning scheme [in (1.5.5)] generally outperforms the first scheme [in (1.5.1)] if the sources are uncorrelated and if their number is correctly specified.

If the data set contains a part with baseline data (see also Section 1.4),  $\overline{Y}$  in the above expressions would simply need to be substituted by  $\overline{Y_2}$ .

## 1.6 Fisher Information Matrix and Cramér-Rao Bound

The FIM can be viewed as a measure of the intrinsic accuracy of a distribution [52]. Its inverse is the Cramér-Rao bound (CRB), which is a lower bound on the covariance matrix of any unbiased estimator. It is achieved asymptotically by the ML estimator, see [19, theorem 1].

Denote the Kronecker product between two matrices by  $\otimes$ , see [72, p. 11] for the definition and some properties. In Appendix 1.G, we derive the FIM for the above model as

$$\mathcal{I}(\boldsymbol{\gamma}) = \begin{bmatrix} \mathcal{I}_{\text{signal}}(\boldsymbol{\gamma}) & 0 \\ 0 & \mathcal{I}_{\text{noise}}(\boldsymbol{\psi}) \end{bmatrix}, \quad (1.6.1)$$

where

$$\mathcal{I}_{\text{signal}}(\boldsymbol{\gamma}) = \begin{bmatrix} \mathcal{I}_{xx} & \mathcal{I}_{\theta x}^T & \mathcal{I}_{\eta x}^T \\ \mathcal{I}_{\theta x} & \mathcal{I}_{\theta\theta} & \mathcal{I}_{\eta\theta}^T \\ \mathcal{I}_{\eta x} & \mathcal{I}_{\eta\theta} & \mathcal{I}_{\eta\eta} \end{bmatrix}, \quad (1.6.2)$$

and

$$\mathcal{I}_{xx} = K \Phi(\boldsymbol{\eta}) \Phi(\boldsymbol{\eta})^T \otimes A(\boldsymbol{\theta})^T \Sigma^{-1} A(\boldsymbol{\theta}), \quad (1.6.3a)$$

$$\mathcal{I}_{\theta x} = K D_A(\boldsymbol{\theta})^T [X \Phi(\boldsymbol{\eta}) \Phi(\boldsymbol{\eta})^T \otimes \Sigma^{-1} A(\boldsymbol{\theta})], \quad (1.6.3b)$$

$$\mathcal{I}_{\eta x} = K D_\phi(\boldsymbol{\theta})^T [\Phi(\boldsymbol{\eta})^T \otimes X^T A(\boldsymbol{\theta})^T \Sigma^{-1} A(\boldsymbol{\theta})], \quad (1.6.3c)$$

$$\mathcal{I}_{\theta\theta} = K D_A(\boldsymbol{\theta})^T [X \Phi(\boldsymbol{\eta}) \Phi(\boldsymbol{\eta})^T X^T \otimes \Sigma^{-1}] D_A(\boldsymbol{\theta}), \quad (1.6.3d)$$

$$\mathcal{I}_{\eta\theta} = K D_\phi(\boldsymbol{\theta})^T [\Phi(\boldsymbol{\eta})^T X^T \otimes X^T A(\boldsymbol{\theta})^T \Sigma^{-1}] D_A(\boldsymbol{\theta}), \quad (1.6.3e)$$

$$\mathcal{I}_{\eta\eta} = K D_\phi(\boldsymbol{\theta})^T [I_N \otimes X^T A(\boldsymbol{\theta})^T \Sigma^{-1} A(\boldsymbol{\theta}) X] D_\phi(\boldsymbol{\theta}), \quad (1.6.3f)$$

$$D_A(\boldsymbol{\theta}) = \frac{\partial \text{vec}(A(\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}^T}, \quad D_\phi(\boldsymbol{\eta}) = \frac{\partial \text{vec}(\Phi(\boldsymbol{\eta}))}{\partial \boldsymbol{\eta}^T}, \quad (1.6.3g)$$

whereas the  $(i, j)$ th entry of  $\mathcal{I}_{\text{noise}}(\boldsymbol{\psi})$  is [49]

$$\left[\mathcal{I}_{\text{noise}}(\boldsymbol{\psi})\right]_{ij} = \frac{NK}{2} \text{tr} \left[ \Sigma^{-1} \frac{\partial \Sigma}{\partial \psi_i} \Sigma^{-1} \frac{\partial \Sigma}{\partial \psi_j} \right]. \quad (1.6.4)$$

Further, let  $\Sigma^{-1} = [\sigma^{ij}]$  and  $\Sigma = [\sigma_{ij}]$ , then a simple formula solves (1.6.4):

$$\text{tr} \left[ \Sigma^{-1} \frac{\partial \Sigma}{\partial \sigma_{pq}} \Sigma^{-1} \frac{\partial \Sigma}{\partial \sigma_{rs}} \right] = \begin{cases} 2(\sigma^{qr} \sigma^{ps} + \sigma^{pr} \sigma^{qs}) & p \neq q, r \neq s \\ 2\sigma^{pr} \sigma^{qr} & p \neq q, r = s, \\ (\sigma^{pr})^2 & p = q, r = s \end{cases} \quad (1.6.5)$$

where  $p, q, r, s \in \{1, \dots, m\}$ . As expected, the information increases linearly with the number of trials  $K$ . The information on noise  $\mathcal{I}_{\text{noise}}(\boldsymbol{\psi})$  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  $\mathcal{I}(\boldsymbol{\gamma})$  that separates the signal and noise parts, its inverse is computed by simply inverting the two diagonal blocks. Thus,  $\text{CRB}_{\text{signal}}(\boldsymbol{\gamma})$  for the unknown noise covariance is equal to the corresponding CRB for known noise covariance. Therefore, the ML method and GLS with correctly specified  $\Sigma$  (see Subsection 1.3.2) have the same asymptotic covariance.

Using the formula for the inverse of a partitioned matrix (see e.g. [29, theorem 8.5.11] and [72, result vi, p. 8]), the CRB for the dipole location and basis-function parameters is

$$\text{CRB}_1 = \begin{bmatrix} \text{CRB}_{\theta\theta} & \text{CRB}_{\eta\theta}^T \\ \text{CRB}_{\eta\theta} & \text{CRB}_{\eta\eta} \end{bmatrix} = \left\{ \begin{bmatrix} \mathcal{I}_{\theta\theta} & \mathcal{I}_{\eta\theta}^T \\ \mathcal{I}_{\eta\theta} & \mathcal{I}_{\eta\eta} \end{bmatrix} - \begin{bmatrix} \mathcal{I}_{\theta x} \\ \mathcal{I}_{\eta x} \end{bmatrix} \cdot \mathcal{I}_{xx}^{-1} \cdot [\mathcal{I}_{\theta x}^T, \mathcal{I}_{\eta x}^T] \right\}^{-1}. \quad (1.6.6)$$

From (1.6.3) it follows that

$$\mathcal{I}_{\eta\theta} - \mathcal{I}_{\eta x} \mathcal{I}_{xx}^{-1} \mathcal{I}_{\theta x}^T = 0, \quad (1.6.7)$$

implying that  $\text{CRB}_1$  is block-diagonal; therefore  $\text{CRB}_{\theta\theta} = [\mathcal{I}_{\theta\theta} - \mathcal{I}_{\theta x} \mathcal{I}_{xx}^{-1} \mathcal{I}_{\theta x}^T]^{-1}$  and  $\text{CRB}_{\eta\eta} = [\mathcal{I}_{\eta\eta} - \mathcal{I}_{\eta x} \mathcal{I}_{xx}^{-1} \mathcal{I}_{\eta x}^T]^{-1}$ , yielding

$$\text{CRB}_{\theta\theta}(\boldsymbol{\gamma}) = \frac{1}{NK} \left[ D_A(\boldsymbol{\theta})^T \left( \widehat{R}_s \otimes \{\Sigma^{-1} - \Sigma^{-1} A(\boldsymbol{\theta}) [A(\boldsymbol{\theta})^T \Sigma^{-1} A(\boldsymbol{\theta})]^{-1} A(\boldsymbol{\theta})^T \Sigma^{-1}\} \right) D_A(\boldsymbol{\theta}) \right]^{-1}, \quad (1.6.8a)$$

$$\text{CRB}_{\eta\eta}(\boldsymbol{\gamma}) = \frac{1}{K} \left[ D_\Phi(\boldsymbol{\eta})^T \left\{ [I_N - \Pi_\Phi(\boldsymbol{\eta})] \otimes X^T A(\boldsymbol{\theta})^T \Sigma^{-1} A(\boldsymbol{\theta}) X \right\} D_\Phi(\boldsymbol{\eta}) \right]^{-1}, \quad (1.6.8b)$$

$$\text{CRB}_{\theta\eta}(\boldsymbol{\gamma}) = 0, \quad (1.6.8c)$$

where

$$\begin{aligned} \widehat{R}_s &= (1/N) \cdot [\mathbf{s}(1) \mathbf{s}(2) \cdots \mathbf{s}(N)] \cdot [\mathbf{s}(1) \mathbf{s}(2) \cdots \mathbf{s}(N)]^T \\ &= (1/N) \cdot X \Phi(\boldsymbol{\eta}) \Phi(\boldsymbol{\eta})^T X^T \end{aligned} \quad (1.6.9)$$

is the estimated covariance matrix of the dipole moment components. Here, to simplify the notation, we omit the dependence of  $\widehat{R}_s$  on  $X$  and  $\boldsymbol{\eta}$ . The expression in (1.6.8a) implies that  $\text{CRB}_{\theta\theta}$  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.  $[\mathbf{s}(1) \mathbf{s}(2) \cdots \mathbf{s}(N)] = X\Phi(\boldsymbol{\eta})$  (Note that this condition is satisfied for a trivial choice of basis functions in the form of Dirac pulses, i.e.  $\Phi = I_N$ ; then  $[\mathbf{s}(1) \mathbf{s}(2) \cdots \mathbf{s}(N)] = X$ .) For  $K = 1$ , (1.6.8a) is equal to the deterministic CRB for known  $\Sigma$  in [60] and [45].

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 (1.6.8c)] is a generalization of similar results in [59] and [18], where it was shown for a particular choice of a basis function, suitable for radar array processing. We have shown in [21] that the above CRB decoupling holds for complex data and noise models as well. As a consequence of the decoupling, (1.6.8a) remains valid when a nonparametric basis-function model is used.

## 1.7 Goodness-of-fit Measures

Goodness-of-fit measures are used to show the degree to which the model-fitted data agrees with the observed data. Most existing measures in the EEG/MEG literature do not account for spatial correlation of the noise and are often computed for only one snapshot [30], [65]. Following [72, chapter 8.3], we consider a multivariate extension of the usual  $R^2$  statistic from univariate linear regression which accounts for multiple snapshots and spatially correlated noise. A similar measure for a single snapshot has been recently introduced in [39].

We consider goodness-of-fit based on the averaged brain responses  $\bar{Y}$ , since utilizing the individual trials would result in very low values of  $R^2$ . Our hypothesized and “null” models are:

$$\text{Hypothesized Model: } \bar{\mathbf{y}}(t) = \mathbf{f}(t, \boldsymbol{\rho}) + \bar{\mathbf{w}}(t), \quad \text{E}[\bar{\mathbf{w}}(t)] = \mathbf{0}, \quad \text{cov}(\bar{\mathbf{w}}(t)) = \bar{\Sigma}, \quad (1.7.1a)$$

$$\text{Null Model: } \bar{\mathbf{y}}_0(t) = \bar{\mathbf{w}}_0(t), \quad \text{E}[\bar{\mathbf{w}}_0(t)] = \mathbf{0}, \quad \text{cov}(\bar{\mathbf{w}}_0(t)) = \bar{\Sigma}_0, \quad (1.7.1b)$$

where  $\mathbf{f}(t, \boldsymbol{\rho})$  denotes an arbitrary model used to fit the data and  $\boldsymbol{\rho}$  is the unknown source parameter to be estimated (in our case  $\mathbf{f}(t, \boldsymbol{\rho}) = A(\boldsymbol{\theta})X\phi(t, \boldsymbol{\eta})$ ).

Denote the fitted value of  $\mathbf{f}(t, \boldsymbol{\rho})$  by  $\hat{\mathbf{y}}(t)$ ,  $t = 1, \dots, N$ . By forming the squared Mahalanobis distances  $d(t, V)^2 = [\bar{\mathbf{y}}(t) - \hat{\mathbf{y}}(t)]^T V^{-1} [\bar{\mathbf{y}}(t) - \hat{\mathbf{y}}(t)]$  and  $d_0(t, V)^2 = \bar{\mathbf{y}}(t)^T V^{-1} \bar{\mathbf{y}}(t)$  for any positive definite matrix  $V$ , we define the explained residual variation of  $\bar{\mathbf{y}}(t)$  relative to the null model and  $V$  as

$$R^2(t, V) = 1 - \frac{d(t, V)^2}{d_0(t, V)^2}, \quad (1.7.2)$$

and the overall explained residual variation as

$$R^2(V) = 1 - \frac{\sum_{t=1}^N d(t, V)^2}{\sum_{t=1}^N d_0(t, V)^2}. \quad (1.7.3)$$

Perfect fit is achieved if  $R^2(V) = 1$ , while a complete lack of fit is indicated by  $R^2(V) \leq 0$ , since the equality is achieved for  $\hat{\mathbf{y}}(t) = \mathbf{0}$ .

Different metrics can be chosen by using various values of  $V$ : a good choice is  $\bar{\Sigma}_0$ , since it is associated with the fixed “null” model. This allows comparisons across different hypothesized models, which is one of the desirable properties of  $R^2$  [37]. However,  $\bar{\Sigma}_0$  is not known, in general. It can be estimated from the baseline data (see also [15]), or computed through an analytical approximation, using a random dipole field model of spontaneous brain activity [14], [38]. Observe that  $R^2(t, \sigma^2 I_m)$  is the “proportion of variance explained” for the averaged snapshot at time  $t$  (see e.g. [30], [65]). Using  $R^2(\sigma^2 I_m)$  and  $R^2(t, \sigma^2 I_m)$ , which do not take into account noise spatial correlation, can yield misleading results, as observed in Example 7.2, Section 1.9, and [37].

## 1.8 EEG/MEG Sensor Array Design

The goal of sensor array design is to determine the *design parameters*  $\xi$  (i.e. sensor array locations, number and type of sensors etc.) to minimize a given cost function. In the following, we propose a sensor array design criterion based on minimizing the volume of a linearized confidence region for dipole location parameters.

We first show how to construct a linearized confidence region for the dipole parameters and compute its volume using Wald tests. The linearized confidence region (in the form of an ellipsoid) for testing  $H_0: \mathbf{h}(\boldsymbol{\rho}) = \mathbf{0}$ , where  $\mathbf{h}$  is a once continuously-differentiable function, is defined as

$$T^2(\boldsymbol{\gamma}) = \mathbf{h}(\boldsymbol{\rho})^T [H(\boldsymbol{\rho}) \cdot \text{CRB}_{\text{signal}}(\boldsymbol{\gamma}) \cdot H(\boldsymbol{\rho})^T]^{-1} \mathbf{h}(\boldsymbol{\rho}) \leq g, \quad (1.8.1)$$

where

$$H(\boldsymbol{\rho}) = \frac{\partial \mathbf{h}(\boldsymbol{\rho})}{\partial \boldsymbol{\rho}^T}, \quad (1.8.2a)$$

$$\text{CRB}_{\text{signal}}(\boldsymbol{\gamma}) = \mathcal{I}_{\text{signal}}(\boldsymbol{\gamma})^{-1}, \quad (1.8.2b)$$

[see also (1.6.2) and (1.6.3)], and  $g$  is the threshold computed to satisfy a desired probability of false alarm [73], [52, chapter 6e.3], [72, chapter 7.3.3]. From (1.8.1), testing  $H_0: \boldsymbol{\rho} - \boldsymbol{\rho}_0 = 0$  yields the confidence ellipsoid of the following form:

$$(\boldsymbol{\rho} - \boldsymbol{\rho}_0)^T \text{CRB}_{\text{signal}}(\boldsymbol{\gamma})^{-1} (\boldsymbol{\rho} - \boldsymbol{\rho}_0) \leq g. \quad (1.8.3)$$

The squared volume of this ellipsoid is proportional to  $|\text{CRB}_{\text{signal}}(\boldsymbol{\gamma})|$  evaluated at  $\boldsymbol{\rho} = \boldsymbol{\rho}_0$ . Similarly, testing  $H_0: \boldsymbol{\theta} - \boldsymbol{\theta}_0 = 0$  yields the confidence ellipsoid whose

squared volume is proportional to  $|\text{CRB}_{\theta\theta}(\boldsymbol{\gamma})|$  evaluated at  $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ . Wald tests have been criticized in [7] for yielding too small (i.e. optimistic) confidence regions. Nevertheless, they provide an idea about the shapes of confidence regions and their relative sizes at different locations, and are thus applicable to EEG/MEG array design.

In the theory of optimal experimental designs, minimizing the determinant of the CRB for *all* signal parameters  $|\text{CRB}_{\text{signal}}(\boldsymbol{\gamma})|$  with respect to the design variables  $\boldsymbol{\xi}$  is referred to as the  $D$ -optimal design, whereas minimizing the determinant of the CRB of a subset of parameters of interest (for example,  $|\text{CRB}_{\theta\theta}(\boldsymbol{\gamma})|$  when  $\boldsymbol{\theta}$  is of interest) is referred to as the  $D_s$ -optimal design, see [1].

For simplicity, let us concentrate on the model (1.2.4), where only one dipole, one time snapshot and one trial are used ( $n = N = K = 1$ ). Then,  $\boldsymbol{\rho} = [\mathbf{s}^T, \boldsymbol{\theta}^T]^T$ , and (1.6.2)–(1.6.3) simplify to

$$\mathcal{I}_{\text{signal}}(\boldsymbol{\gamma}) = \begin{bmatrix} \mathcal{I}_{ss} & \mathcal{I}_{\theta s}^T \\ \mathcal{I}_{\theta s} & \mathcal{I}_{\theta\theta} \end{bmatrix} \quad (1.8.4)$$

and

$$\mathcal{I}_{ss} = A(\boldsymbol{\theta})^T \Sigma^{-1} A(\boldsymbol{\theta}), \quad (1.8.5a)$$

$$\mathcal{I}_{\theta s} = D_A(\boldsymbol{\theta})^T [\mathbf{s} \otimes \Sigma^{-1} A(\boldsymbol{\theta})], \quad (1.8.5b)$$

$$\mathcal{I}_{\theta\theta} = D_A(\boldsymbol{\theta})^T [\mathbf{ss}^T \otimes \Sigma^{-1}] D_A(\boldsymbol{\theta}). \quad (1.8.5c)$$

In this case,  $\boldsymbol{\rho}$  is a  $d \times 1$  vector, where  $d = 3 + r$ . Now,  $\text{CRB}_{\theta\theta}(\boldsymbol{\gamma})$  can be readily computed by applying the matrix inversion formula (see [29, theorem 8.5.11] and [72, result vi, p. 8]) to (1.8.4):

$$\begin{aligned} \text{CRB}_{\theta\theta}(\boldsymbol{\gamma}) &= [\mathcal{I}_{\theta\theta} - \mathcal{I}_{\theta s} \mathcal{I}_{ss}^{-1} \mathcal{I}_{\theta s}^T]^{-1} \\ &= \left\{ D_A(\boldsymbol{\theta})^T \left[ \mathbf{ss}^T \otimes [\Sigma^{-1} - \Sigma^{-1} A(\boldsymbol{\theta}) (A(\boldsymbol{\theta})^T \Sigma^{-1} A(\boldsymbol{\theta}))^{-1} A(\boldsymbol{\theta})^T \Sigma^{-1}] \right] D_A(\boldsymbol{\theta}) \right\}^{-1}, \end{aligned} \quad (1.8.6)$$

see also (1.6.8a). Applying the formula for the determinant of a partitioned matrix in [72, result v, p. 8] to (1.8.4), we obtain the following relationship between the  $D$ -optimality criterion (for the vector of all signal parameters  $\boldsymbol{\rho}$ ) and  $D_s$ -optimality criterion for  $\boldsymbol{\theta}$ , under the measurement model in (1.2.4):

$$|\text{CRB}_{\text{signal}}(\boldsymbol{\gamma})| = |\text{CRB}_{ss|\theta}(\boldsymbol{\theta}, \Sigma)| \cdot |\text{CRB}_{\theta\theta}(\boldsymbol{\gamma})| = \frac{|\text{CRB}_{\theta\theta}(\boldsymbol{\gamma})|}{|A(\boldsymbol{\theta})^T \Sigma^{-1} A(\boldsymbol{\theta})|}, \quad (1.8.7)$$

where

$$\text{CRB}_{ss|\theta}(\boldsymbol{\theta}, \Sigma) = \mathcal{I}_{ss}(\boldsymbol{\gamma})^{-1} = [A(\boldsymbol{\theta})^T \Sigma^{-1} A(\boldsymbol{\theta})]^{-1} \quad (1.8.8)$$

is the CRB for  $\mathbf{s}$  assuming that the dipole location  $\boldsymbol{\theta}$  is *known*. Therefore, the squared volume of the confidence ellipsoid for all signal parameters  $\boldsymbol{\rho}$  is equal to the product between the

- squared volume of the confidence ellipsoid for the dipole location  $\boldsymbol{\theta}$  and
- squared volume of the confidence ellipsoid for the dipole moment vector  $\mathbf{s}$  assuming known dipole location  $\boldsymbol{\theta}$ .

### 1.8.1 Reparametrization Invariance

We now show that  $D$ - and  $D_s$ -optimality criteria are invariant to reparametrization. If  $\boldsymbol{\rho} \mapsto \mathbf{h}(\boldsymbol{\rho})$  is a smooth non-singular reparametrization of the signal parameters, the CRB for this model is equal to

$$\text{CRB}_h(\boldsymbol{\gamma}) = H(\boldsymbol{\rho}) \cdot \text{CRB}_{\text{signal}}(\boldsymbol{\gamma}) \cdot H(\boldsymbol{\rho})^T, \quad (1.8.9)$$

where  $H(\boldsymbol{\rho})$  is defined in (1.8.2a). Therefore,

$$|\text{CRB}_h(\boldsymbol{\gamma})| = |H(\boldsymbol{\rho})|^2 \cdot |\text{CRB}_{\text{signal}}(\boldsymbol{\gamma})|. \quad (1.8.10)$$

Provided that the reparametrization  $\mathbf{h}(\boldsymbol{\rho})$  is independent of  $\boldsymbol{\xi}$ ,  $|\text{CRB}_h(\boldsymbol{\gamma})|$  and  $|\text{CRB}_{\text{signal}}(\boldsymbol{\gamma})|$  are minimized for the same choice of the design parameters  $\boldsymbol{\xi}$ . A similar argument implies that the  $D_s$ -optimality criterion is invariant to reparametrization as well. Therefore, if  $D$ - and  $D_s$ -optimal design criteria are used, it becomes irrelevant whether the dipole location  $\boldsymbol{\theta}$  is expressed in the Cartesian or spherical coordinate system, which is a very desirable property.

### 1.8.2 Relationship between Optimal Array Design and Information Theory

We show that the  $D$ -optimal designs have information-theoretic justification. First, recall the definitions of relative entropy and mutual information, see [11, chapter 9.5]. The *relative entropy* (also called the Kullback-Leibler distance or information divergence) between probability densities  $p(\mathbf{x})$  and  $q(\mathbf{x})$  is defined as

$$D(p \| q) = \int p(\mathbf{x}) \log \left( \frac{p(\mathbf{x})}{q(\mathbf{x})} \right) d\mathbf{x}, \quad (1.8.11)$$

Then, the *mutual information* between two random vectors  $\mathbf{y}$  and  $\boldsymbol{\rho}$  with joint density  $f_{\mathbf{y},\boldsymbol{\rho}}(\mathbf{y}, \boldsymbol{\rho})$  is

$$\begin{aligned} I(\mathbf{y}, \boldsymbol{\rho}) &= D(f_{\mathbf{y},\boldsymbol{\rho}}(\mathbf{y}, \boldsymbol{\rho}) \| f_{\mathbf{y}}(\mathbf{y})f_{\boldsymbol{\rho}}(\boldsymbol{\rho})) \\ &= \int f_{\boldsymbol{\rho}}(\boldsymbol{\rho}) d\boldsymbol{\rho} \int f_{\mathbf{y}|\boldsymbol{\rho}}(\mathbf{y}|\boldsymbol{\rho}) \log \left( \frac{f_{\mathbf{y}|\boldsymbol{\rho}}(\mathbf{y}|\boldsymbol{\rho})}{f_{\mathbf{y}}(\mathbf{y})} \right) d\mathbf{y} \end{aligned} \quad (1.8.12)$$

where  $f_{\mathbf{y}}(\mathbf{y})$  and  $f_{\boldsymbol{\rho}}(\boldsymbol{\rho})$  are the marginal densities of  $\mathbf{y}$  and  $\boldsymbol{\rho}$ , respectively.

To simplify the following discussion, we assume that the noise covariance  $\Sigma$  is known; hence  $\text{CRB}_{\text{signal}} = \text{CRB}_{\text{signal}}(\boldsymbol{\rho}) = \mathcal{I}_{\text{signal}}(\boldsymbol{\rho})^{-1}$ . It was shown in [10] that, under certain regularity conditions, the relative entropy between the conditional

and the marginal distributions of the measurement vector  $\mathbf{y}$  [denoted by  $p(\mathbf{y}|\boldsymbol{\rho})$  and  $p(\mathbf{y})$ , respectively] exhibits the following asymptotic behavior:

$$D(p(\mathbf{y}|\boldsymbol{\rho}) \| p(\mathbf{y})) = \frac{d}{2} \log\left(\frac{m}{2\pi e}\right) + \frac{1}{2} \log |\mathcal{I}_{\text{signal}}(\boldsymbol{\rho})| - \log(p(\boldsymbol{\rho})) + o(1), \quad (1.8.13)$$

where  $p(\boldsymbol{\rho})$  is the prior density of  $\boldsymbol{\rho}$  and  $o(1) \rightarrow 0$  as  $m \rightarrow \infty$ . [Recall that  $m$  is the number of sensors in the array and  $d$  is the size of the signal parameter vector  $\boldsymbol{\rho}$ .] Obviously, maximizing the above expression with respect to the design parameters  $\boldsymbol{\xi}$  is asymptotically equivalent to minimizing  $|\text{CRB}_{\text{signal}}(\boldsymbol{\rho})|$ , provided that the prior distribution  $p(\boldsymbol{\rho})$  is not functionally dependent on  $\boldsymbol{\xi}$ .

**Relationship with Mutual Information and Bayesian Array Design:** From (1.8.12) and (1.8.13), it follows that

$$I(\mathbf{y}, \boldsymbol{\rho}) = \frac{d}{2} \log\left(\frac{m}{2\pi e}\right) + \frac{1}{2} \int f_{\rho}(\boldsymbol{\rho}) \log |\mathcal{I}_{\text{signal}}(\boldsymbol{\rho})| d\boldsymbol{\rho} + h(\boldsymbol{\rho}) + o(1), \quad (1.8.14)$$

where

$$h(\boldsymbol{\rho}) = - \int f_{\rho}(\boldsymbol{\rho}) \log[f_{\rho}(\boldsymbol{\rho})] d\boldsymbol{\rho} \quad (1.8.15)$$

is the differential entropy of  $\boldsymbol{\rho}$ , see [11, chapter 9]). Hence, asymptotically (as  $m \rightarrow \infty$ ), the mutual information between the data vector  $\mathbf{y}$  and the vector of signal parameters  $\boldsymbol{\rho}$  is maximized by minimizing

$$E_{\rho}[\log |\text{CRB}_{\text{signal}}(\boldsymbol{\rho})|] = \int f_{\rho}(\boldsymbol{\rho}) \log |\text{CRB}_{\text{signal}}(\boldsymbol{\rho})| d\boldsymbol{\rho}, \quad (1.8.16)$$

provided that the prior distribution  $f_{\rho}(\boldsymbol{\rho})$  does not depend on the design parameter vector  $\boldsymbol{\xi}$ . The criterion (1.8.16) belongs to the class of Bayesian optimal experimental designs, recently proposed in [2] (see also references therein). Now, using (1.8.7), we can decompose (1.8.16) as

$$\begin{aligned} E_{\rho}[\log |\text{CRB}_{\text{signal}}(\boldsymbol{\rho})|] &= E_{\rho}[\log |\text{CRB}_{\theta\theta}(\boldsymbol{\rho})|] - E_{\theta}[\log |A(\boldsymbol{\theta})^T \Sigma^{-1} A(\boldsymbol{\theta})|] \\ &= \int f_{\rho}(\boldsymbol{\rho}) \log |\text{CRB}_{\theta\theta}(\boldsymbol{\rho})| d\boldsymbol{\rho} - \int f_{\theta}(\boldsymbol{\theta}) \log |A(\boldsymbol{\theta})^T \Sigma^{-1} A(\boldsymbol{\theta})| d\boldsymbol{\theta}, \end{aligned} \quad (1.8.17)$$

where the first term in the above expression can be viewed as a Bayesian  $D_s$ -optimal design criterion for the vector of dipole location parameters  $\boldsymbol{\theta}$ .

Following the discussion in Section 1.8.1, it is easy to show that the above Bayesian  $D$ - and  $D_s$ -optimal are invariant to smooth non-singular reparametrizations  $\boldsymbol{\rho} \mapsto \mathbf{h}(\boldsymbol{\rho})$ .

Another interesting array design criterion is the mean-square error of dipole location estimates, which we examined in [31].

## 1.9 Numerical Examples

### Example 7.1: Simulated Data

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 1.3 and 1.6.

The simulation was performed for an MEG configuration of 37 radial magnetometers located on a spherical helmet of radius 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 4-D Neuroimaging Inc., San Diego, California.

We generated two coherent dipole sources. The components  $s_\theta$  and  $s_\varphi$  of the first dipole change in time

$$s_\theta = 15 \exp(-(t-60)^2/8^2) - 5 \exp(-(t-40)^2/17^2) \quad [\text{nA} \cdot \text{m}], \quad (1.9.1a)$$

$$s_\varphi = 13 \exp(-(t-60)^2/12^2) - 3 \exp(-(t-40)^2/17^2) \quad [\text{nA} \cdot \text{m}], \quad (1.9.1b)$$

and the corresponding components of the second dipole as  $s_\theta(t)$  and  $-s_\varphi(t)$ , i.e. the sources are correlated. The dipoles are symmetric relative to the mid-sagittal 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 of radius 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  $\mathcal{N}(0, \sigma_m^2)$ . For  $\sigma_m = 1 \text{ nA} \cdot \text{m}$  the total noise standard deviation at the sensors was approximately 110 fT, consistent with  $25 \text{ fT}/\sqrt{\text{Hz}}$  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 [30]. The peak value of the signal at the sensor with the largest response was around 270 fT, consistent with typical values measured in practical applications.

In EEG/MEG literature, several parametric models have been used to model temporal evolution of the evoked responses: decaying sinusoids (see [68]), double Gaussian (see [67]), or Hermite wavelets (see [26]). 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.

In Figure 1.2, we compare the localization accuracies of the ML, GLS, OLS, and scanning methods by showing the mean localization errors per dipole  $(1/2) \cdot$

**Figure 1.2.** Average location error per dipole as a function of the noise level  $\sigma_m$  for (a) ML method with parametric basis functions, (b) ML method with  $l = 2n$  nonparametric basis functions, (c) GLS with  $\Phi = I_N$ , (d) OLS with parametric basis functions, (e) OLS with  $\Phi = I_N$ . (f) scanning with unknown dipole orientation that is fixed in time.

$(\|\widehat{\mathbf{p}}_1 - \mathbf{p}_1\|^2 + \|\widehat{\mathbf{p}}_2 - \mathbf{p}_2\|^2)^{1/2}$  (averaged over the 50 runs) as functions of  $\sigma_m$ . Here,  $\|\cdot\|$  denotes the Euclidean norm, and  $\mathbf{p}_1$ ,  $\mathbf{p}_2$  and  $\widehat{\mathbf{p}}_1$ ,  $\widehat{\mathbf{p}}_2$  are the location vectors of the two dipoles and the corresponding ML estimates, see also (1.2.1). 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  $\sigma_m$  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  $\sigma_m$  (shown in Figure 1.2), whereas the ML estimation errors remain very small, showing the robustness of the ML method. A similar trend was also observed in [39], [40].

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 1.6, 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  $\text{CRB}_{\theta\theta}$  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 for unknown fixed (time-invariant) dipole orientation in (1.5.2), see Section 1.5. As shown in Figure 1.2, this algorithm is robust to the increase in the noise level  $\sigma_m$  because it accounts for spatially correlated noise. Further, for larger values of  $\sigma_m$ , 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  $\Phi = I_N$  requires a 6-D search for the two-dipole fit). Note that, for small values of  $\sigma_m$ , 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  $K \rightarrow \infty$ , both the ML and OLS estimates converge to the true parameters, as shown in [19, 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 [41].

**Figure 1.3.** Left: Auditory evoked responses averaged over 100 trials. Right: N100 peak response (view from above).

### Example 7.2: Real Data

We demonstrate the performance of the proposed method in application to real auditory evoked response data. The dipole moments were estimated using the ML method in Section 1.4. The optimization algorithms were initialized using the scanning technique described in Section 1.5, assuming known orientation, see discussion below. We also apply the goodness-of-fit statistic suggested in Section 1.7.

The measurements were obtained by CTF’s whole head MEG system with 143 first-order gradiometers in an unshielded environment. An auditory stimulus at 1 kHz was applied to the subject and repeated  $K = 100$  times. After being sampled at 1250 Hz, the data was low-pass filtered with a 30 Hz cut-off frequency, and the DC offset was removed. The number of snapshots per trial was  $N = 400$ , corresponding to an observation time of 320 ms. The prestimulus interval contains the first 120 ms (150 snapshots) of each trial, and is used as baseline. The peak activity is referred to as N100m [51] because it occurs approximately 100 ms after stimulus onset. It has been hypothesized that the location of the N100m sources (which are often modeled by current dipoles) depend on many parameters: latency, stimulus frequency, stimulus intensity [see e.g. [48]]. Accurate source locations and moments are necessary for studying the above dependencies, which could help understanding of the functioning of the human auditory cortex.

Figure 1.3 shows averaged (over trials, timelocked to the instant of stimulus application at  $t = 0$ ) temporal evolutions of all 143 channel measurements, and the spatial distribution of N100m peak response over the helmet (interpolated between sensor locations), viewed from above. Figure 1.4 shows side views of the N100m (averaged) peak responses.

Figure 1.5 shows the estimated temporal evolutions of the dipole moment components  $s_\theta(t)$  and  $s_\varphi(t)$ , using the ML method with  $l = 3$  nonparametric basis functions (see Section 1.4). These estimates capture a small delay between the

**Figure 1.4.** N100 peak response in  $x$ - $z$  plane (left and right ear view).

**Figure 1.5.** Estimated dipole moment components of the auditory evoked response as a function of time.

peak responses of the right and left ear. Fitting two dipoles yielded the following location estimates:  $\hat{\boldsymbol{\theta}}_1 = [1.57, 1.59, 5.15 \text{ cm}]^T$  and  $\hat{\boldsymbol{\theta}}_2 = [1.61, -1.36, 5.4 \text{ cm}]^T$ . These locations are shown in Figures 1.6, 1.7, and 1.8, overlaid on MRI scans of the brain (the circles show projections of the sphere used to model the head).

Using the OLS estimator with  $\mathbf{s}(t)$  allowed to vary arbitrarily in time (see Subsection 1.3.2) yields  $\hat{\boldsymbol{\theta}}_{1\text{OLS}} = [1.63, 1.57, 3.81 \text{ cm}]^T$  and  $\hat{\boldsymbol{\theta}}_{2\text{OLS}} = [1.70, -1.49, 3.75 \text{ cm}]^T$ . Observe that the OLS estimates of the dipoles' azimuth and elevation are similar to those obtained by the ML, whereas their distances from the center of the head are significantly smaller. This is consistent with the results of Example 7.1, where the OLS estimates moved deeper as the noise level increased.

In the above examples, the optimization algorithms were initialized using the scanning procedure suggested in Section 1.5 with dipoles having known orientations. Scanning is performed for elevational and azimuthal orientations separately, yielding two sets of initial values. The best initial estimates obtained by scanning were only 6.5 and 8 mm away from the ML location estimates of the two dipoles. In Figures

**Figure 1.6.** Left and right dipole location estimates overlaid on coronal MRI brain scans.

**Figure 1.7.** Dipole location estimates overlaid on axial MRI brain scans.

**Figure 1.8.** Dipole location estimates overlaid on sagittal MRI brain scans.

**Figure 1.9.** Concentrated likelihood function for a single dipole, used for scanning in  $\mathbf{u}_\varphi$  direction over a sphere of radius 5 cm.

**Figure 1.10.** Single dipole concentrated likelihood function, used for scanning in  $\mathbf{u}_\vartheta$  direction over a sphere of radius 5 cm.

	1 dipole	2 dipoles	3 dipoles
$R^2(\sigma^2 I_m)$	0.2912	0.5839	0.5854
$R^2(\bar{\Sigma}_0)$	-0.0054	0.7031	0.7374

**Table 1.1.** Goodness-of-fit for the ML method.

	1 dipole	2 dipoles	3 dipoles
$R^2(\sigma^2 I_m)$	0.3845	0.7574	0.8362
$R^2(\bar{\Sigma}_0)$	-0.3200	0.4952	0.4160

**Table 1.2.** Goodness-of-fit for the OLS method.

1.9 and 1.10, we demonstrate how the scanning function for known fixed dipole orientation depends on the choice of orientation. We show the scanning function in (1.5.4) for azimuthal and elevational orientations, respectively. In particular, we plot this function on the sphere of radius 5 cm (around the center of the head) for assumed dipole orientations described by  $\mathbf{u}_\varphi$  (azimuthal) and  $\mathbf{u}_\theta$  (elevational). The ML dipole estimates (at the N100m peak, for  $n = 2$ ) are shown by black arrows. Both scanning functions show exactly two distinguished peaks close to the ML estimates. The peaks in the scanning function are quite close to these estimates, showing the feasibility of the scanning technique. Interestingly, in this case, scanning in the azimuthal direction has peaks closer to the actual ML estimates, compared to scanning in the elevational direction.

Next, we computed goodness-of-fit measures  $R^2(V)$  as a function of the number of dipoles for  $V = \sigma^2 I_m$  and  $V = \bar{\Sigma}_0$ , estimated from the baseline data (containing only the background noise). Table 1.1 shows the goodness-of-fit of the ML method as a function of the number of fitted dipoles. Both values of  $R^2$  saturate when the data is fitted with 3 dipoles. Table 1.2 shows the goodness-of-fit of the OLS method (with  $\Phi = I_N$ ). Since  $1 - R^2(\sigma^2 I_m)$  equals the OLS cost function,  $R^2(\sigma^2 I_m)$  has higher values than in Table 1.1. However, the dipole locations of a 3-dipole fit are clearly not admissible (all three dipoles fall within a 2 cm distance from the center of the head, which is far from the cortex). This unacceptable solution causes a drop in the value of  $R^2(\bar{\Sigma}_0)$  compared with a 2-dipole fit, whereas  $R^2(\sigma^2 I_m)$  continues to improve. A drawback of  $R^2(\bar{\Sigma}_0)$  is that it does not have a direct quantitative interpretation, since it is “proportion of variance explained” of the transformed (i.e. spatially whitened) data. Thus, *both*  $R^2(\bar{\Sigma}_0)$  and  $R^2(\sigma^2 I_m)$  should be computed and analyzed as goodness-of-fit indicators.

## 1.10 Conclusions

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. Cramér-Rao bounds for the proposed model were derived. We also showed that the proposed estimators are asymptotically more efficient than the nonlinear OLS estimators. Finally, we proposed optimal array design criteria and discussed their properties.

We presented numerical examples with simulated and real MEG data, demonstrating the performance of the ML methods. The ML and OLS methods were compared; the ML was more accurate and robust, confirming the theoretical results in Section 1.3.

In [17], 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. There are several interesting topics for further research:

- 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 [47], [9];
- simulation analysis and comparison of the proposed scanning methods;
- classifying evoked responses for diagnostic purposes;
- 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 [46]);
- more extensive applications to real data.

## Acknowledgment

This work was supported by the Air Force Office of Scientific Research under Grants F49620-00-1-0083 and F49620-02-1-0339, the National Science Foundation under Grant CCR-0105334, and the Office of Naval Research under Grant N00014-01-1-0681. We are grateful to Drs J. Vrba, T. Cheung, and D. Cheyne from CTF Systems Inc. for providing the real data in Example 7.2.

---



---

# APPENDICES

## 1.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  $\Phi$  into a matrix  $\tilde{\Phi} = [\Phi \cdots \Phi]$  of size  $l \times NK$ . [We use  $A$  instead of  $A(\boldsymbol{\theta})$  and  $\Phi$  instead of  $\Phi(\boldsymbol{\eta})$ , since  $\boldsymbol{\theta}$  and  $\boldsymbol{\eta}$  are assumed to be known.] Using the above notation and the measurement model from Section 1.2.2, the likelihood function can be written as

$$f(\tilde{Y}; X, \Sigma) = \frac{1}{(2\pi)^{mNK/2} |\Sigma|^{NK/2}} \exp \left\{ -\frac{1}{2} \text{tr} [\Sigma^{-1} (\tilde{Y} - AX\tilde{\Phi})(\tilde{Y} - AX\tilde{\Phi})^T] \right\}. \quad (1.A.1)$$

Then, according to [36], [58, chapter 6.4], the ML estimates of  $X$  and  $\Sigma$  are:

$$\hat{X} = (A^T S^{-1} A)^{-1} A^T S^{-1} \tilde{Y} \tilde{\Phi}^T (\tilde{\Phi} \tilde{\Phi}^T)^{-1}, \quad (1.A.2a)$$

$$\hat{\Sigma} = S + \frac{1}{NK} \cdot (I_m - TS^{-1}) \tilde{Y} \tilde{\Phi}^T (\tilde{\Phi} \tilde{\Phi}^T)^{-1} \tilde{\Phi} \tilde{Y}^T (I_m - TS^{-1})^T, \quad (1.A.2b)$$

where

$$\hat{R} = \frac{1}{NK} \tilde{Y} \tilde{Y}^T = \frac{1}{NK} \sum_{k=1}^K Y_k Y_k^T, \quad (1.A.3a)$$

$$S = \hat{R} - \frac{1}{NK} \cdot \tilde{Y} \tilde{\Phi}^T (\tilde{\Phi} \tilde{\Phi}^T)^{-1} \tilde{\Phi} \tilde{Y}^T, \quad (1.A.3b)$$

$$T = A(\boldsymbol{\theta}) [A(\boldsymbol{\theta})^T S^{-1} A(\boldsymbol{\theta})]^{-1} A(\boldsymbol{\theta})^T. \quad (1.A.3c)$$

Observing that

$$\tilde{\Phi} \tilde{\Phi}^T = K \Phi \Phi^T \quad (1.A.4a)$$

$$\tilde{\Phi} \tilde{Y}^T = \sum_{k=1}^K \Phi Y_k^T = K \Phi \bar{Y}^T \quad (1.A.4b)$$

directly yields (1.3.8a)–(1.3.8d). Substituting the above estimates into the likelihood function (1.A.1) we obtain the concentrated likelihood function

$$f(\tilde{Y}; \hat{X}, \hat{\Sigma}) = \frac{1}{(2\pi)^{mNK/2} |\hat{\Sigma}|^{NK/2}} \exp[-(1/2) \cdot mNK], \quad (1.A.5)$$

and the concentrated likelihood function  $l_{\text{ML}}(\boldsymbol{\theta}, \boldsymbol{\eta}) = \hat{R}/|\hat{\Sigma}(\boldsymbol{\theta}, \boldsymbol{\eta})|$  in (1.3.10) follows. Using [21, eq. (3.3) and App. A] yields an alternative expression for the concentrated likelihood:

$$l_{\text{ML}}(\boldsymbol{\theta}, \boldsymbol{\eta}) = \frac{|\tilde{\Phi}(I_{NK} - (1/NK) \cdot \tilde{Y}^T Q(\boldsymbol{\theta}) \tilde{Y}) \tilde{\Phi}^T|}{|\tilde{\Phi}(I_{NK} - (1/NK) \cdot \tilde{Y}^T \hat{R}^{-1} \tilde{Y}) \tilde{\Phi}^T|}, \quad (1.A.6)$$

and (1.3.10b) follows by applying the identities (1.A.4) to (1.A.6). Using similar arguments, the third expression in (1.3.10) easily follows from [21, eq. (3.5) and App. A].

## 1.B Parameter Identifiability

As observed in Subsection 1.3.4, the regularity condition R5) is essentially an identifiability condition for the unknown signal parameters  $\boldsymbol{\rho}$ . Here, we formally introduce the concept of identifiability by distribution and apply it to the signal and noise parameters in the measurement model (1.3.4). We then derive several necessary identifiability conditions for the unknown parameters.

Let  $F(\mathbf{y}; \boldsymbol{\gamma})$  be the distribution of a random vector  $\mathbf{y}$  based on an unknown parameter vector  $\boldsymbol{\gamma}$ . Following [5] and [6, chapter 1.1.2], we say that the parametric function  $f(\boldsymbol{\gamma})$  is identifiable by distribution if there exist no pair of values  $\boldsymbol{\gamma}_1$  and  $\boldsymbol{\gamma}_2$  such that  $f(\boldsymbol{\gamma}_1) \neq f(\boldsymbol{\gamma}_2)$  and  $F(\mathbf{y}, \boldsymbol{\gamma}_1) = F(\mathbf{y}, \boldsymbol{\gamma}_2)$ . Since in our model the distribution of the observations is specified up to the first and second moments only, the concept of identifiability needs an appropriate modification. The following definition, adapted to our measurement model in (1.3.4), represents such a modification (see also [53, p. 74]):

**Definition.** The parameter vector  $\boldsymbol{\rho}$  is said to be identifiable if, for  $\boldsymbol{\rho}_1$  and  $\boldsymbol{\rho}_2$  (which belong to the parameter space of  $\boldsymbol{\rho}$ )

$$\boldsymbol{\rho}_1 \neq \boldsymbol{\rho}_2 \Rightarrow A(\boldsymbol{\theta}_1) X_1 \Phi(\boldsymbol{\eta}_1) \neq A(\boldsymbol{\theta}_2) X_2 \Phi(\boldsymbol{\eta}_2). \quad (1.B.1)$$

Similarly, the parameter vector  $\boldsymbol{\psi}$  is identifiable if, for  $\boldsymbol{\psi}_1$  and  $\boldsymbol{\psi}_2$  (which belong to the parameter space of  $\boldsymbol{\psi}$ )

$$\boldsymbol{\psi}_1 \neq \boldsymbol{\psi}_2 \Rightarrow \Sigma(\boldsymbol{\psi}_1) \neq \Sigma(\boldsymbol{\psi}_2). \quad (1.B.2)$$

**Remark 1.** If the conditions (1.B.1) or (1.B.2) fail, then  $\boldsymbol{\rho}$  or  $\Sigma$  are not identifiable by distribution.

**Remark 2.** Since, in our model, the noise covariance is unstructured [ $\boldsymbol{\psi} = \text{vech}(\boldsymbol{\Sigma})$ ], it is obvious that  $\boldsymbol{\psi}$  is always identifiable.

To achieve identifiability of  $\boldsymbol{\rho}$ , certain obvious conditions must be imposed on the allowable parameter values. Specifically, the parameter space of  $\boldsymbol{\theta}$  should impose ordering of dipole locations; otherwise permuting the dipole responses in the response matrix and appropriate modifications of the matrix  $X$  would violate the definition in (1.B.1). Also, the parameter space of  $X$  needs to be restricted to avoid degenerate cases, such as sets of zero rows of  $X$ , which would not allow the identification of some dipole locations. Furthermore, there is an artificial identifiability problem due to the use of spherical coordinate system: if the elevation of a dipole  $\vartheta$  is zero, its azimuth can be arbitrary and is not identifiable.

The radial dipole components are not identifiable if only MEG sensors are employed, see Section 1.2. Consequently, a dipole located in the center of the head cannot be localized by an MEG array. The parameter space of dipole locations should be restricted to avoid the above problems.

To allow unique determination of basis functions, zero columns of  $X$  should not be allowed, since the corresponding rows of  $\Phi(\boldsymbol{\eta})$  are redundant. If a nonparametric basis function model is used, the parameter space of the basis functions should be restricted to allow a unique solution; for example, it is sufficient to impose the orthonormality condition on the rows of  $\Phi$  (see also the discussion in Appendix 1.E).

We now examine the identifiability of  $X$ , assuming that  $\boldsymbol{\theta}$  and  $\boldsymbol{\eta}$  are known. [Since  $\boldsymbol{\theta}$  and  $\boldsymbol{\eta}$  are known, we use  $A$  and  $\Phi$  instead of  $A(\boldsymbol{\theta})$  and  $\Phi(\boldsymbol{\eta})$ .] A linear parametric function  $\mathbf{h}^T \text{vec}(X)$  is identifiable if, for every  $X_1, X_2 \in \mathbf{R}^{nr \times l}$

$$\mathbf{h}^T \text{vec}(X_1) \neq \mathbf{h}^T \text{vec}(X_2) \quad \Rightarrow \quad (\Phi^T \otimes A) \text{vec}(X_1) \neq (\Phi^T \otimes A) \text{vec}(X_2). \quad (1.B.3)$$

(Here, we have used the fact that  $AX_1\Phi \neq AX_2\Phi$  is equivalent to  $(\Phi^T \otimes A)\text{vec}(X_1) \neq (\Phi^T \otimes A)\text{vec}(X_2)$ , see [72, result i at p. 12].) Following [53, Theorem 4.2.1], the above condition is satisfied if and only if  $\mathbf{h}$  belongs to the column space of  $\Phi \otimes A^T$  (which has size  $nrl \times mN$ ). Since  $\text{rank}(A) = nr < m$  and  $\text{rank}(\Phi) = l \leq N$  (see the discussion in Sections 1.2–1.4), it follows that  $\Phi \otimes A^T$  has more columns than rows. Also, note that the column space of  $\Phi \otimes A^T$  is full if both  $A$  and  $\Phi$  have full ranks. Then,  $\mathbf{h}^T \text{vec}(X)$  is identifiable for an arbitrary vector  $\mathbf{h}$ , implying that  $X$  is identifiable.

We now consider joint identifiability of  $X$  and  $\boldsymbol{\theta}$ , for known  $\boldsymbol{\eta}$ . [Since  $\boldsymbol{\eta}$  is known, we use  $\Phi$  instead of  $\Phi(\boldsymbol{\eta})$ .] We also assume that all the necessary identifiability conditions discussed above hold:  $A(\boldsymbol{\theta})$  and  $\Phi$  are full-rank matrices and the above assumptions on the parameter spaces of  $\boldsymbol{\theta}$ ,  $X$ , and  $\boldsymbol{\eta}$  are satisfied. The condition for identifiability of  $\boldsymbol{\theta}$  and  $X$  is now

$$\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2 \text{ and } X_1 \neq X_2 \quad \Rightarrow \quad A(\boldsymbol{\theta}_1)X_1\Phi \neq A(\boldsymbol{\theta}_2)X_2\Phi, \quad (1.B.4)$$

and a straightforward extension of the derivation in [74] yields the following neces-

sary condition on the number of sources  $n$  that can be identified:

$$n < \frac{m + \text{rank}(X)}{2r}. \quad (1.B.5)$$

Finally, to ensure a positive definite estimate of  $\Sigma$  with probability 1, the following inequality must hold

$$NK - m - l \geq 0, \quad (1.B.6)$$

which is the usual multivariate analysis of variance (MANOVA) restriction stating that the number of observations per snapshot plus the number of basis functions must not exceed the number of snapshots (see also Appendix 1.E).

### 1.C Asymptotic Properties of the OLS Estimates

We prove the asymptotic properties of the OLS estimators stated in Theorem 7.2.

Assume that the true parameters are  $\boldsymbol{\rho} = \boldsymbol{\rho}_0$ , and define the OLS cost function as

$$\mathcal{Q}_K(\boldsymbol{\rho}) = \frac{1}{K} \sum_{k=1}^K \text{tr} \left\{ [Y_k - A(\boldsymbol{\theta})X\Phi(\boldsymbol{\eta})][Y_k - A(\boldsymbol{\theta})X\Phi(\boldsymbol{\eta})]^T \right\}. \quad (1.C.1)$$

This expression can be rewritten as

$$\begin{aligned} \mathcal{Q}_K(\boldsymbol{\rho}) &= \frac{2}{K} \sum_{k=1}^K \text{tr} \left\{ [Y_k - A(\boldsymbol{\theta}_0)X_0\Phi(\boldsymbol{\eta}_0)][A(\boldsymbol{\theta}_0)X_0\Phi(\boldsymbol{\eta}_0) - A(\boldsymbol{\theta})X\Phi(\boldsymbol{\eta})]^T \right\} \\ &\quad + \mathcal{Q}_K(\boldsymbol{\rho}_0) + \mathcal{Q}(\boldsymbol{\rho}), \end{aligned} \quad (1.C.2)$$

where

$$\mathcal{Q}(\boldsymbol{\rho}) = \text{tr} \left\{ [A(\boldsymbol{\theta})X\Phi(\boldsymbol{\eta}) - A(\boldsymbol{\theta}_0)X_0\Phi(\boldsymbol{\eta}_0)] \cdot [A(\boldsymbol{\theta})X\Phi(\boldsymbol{\eta}) - A(\boldsymbol{\theta}_0)X_0\Phi(\boldsymbol{\eta}_0)]^T \right\} \quad (1.C.3)$$

can have only non-negative values, and is exactly the expression in the regularity condition R5) with  $\Sigma = I_m$ . Then, the strong law of large numbers implies that (as  $K \rightarrow \infty$ ):

$$\mathcal{Q}_K(\boldsymbol{\rho}_0) \xrightarrow{\text{a.s.}} \text{tr}(\Sigma), \quad (1.C.4)$$

$$\frac{2}{K} \sum_{k=1}^K \text{tr} \left\{ [Y_k - A(\boldsymbol{\theta}_0)X_0\Phi(\boldsymbol{\eta}_0)][A(\boldsymbol{\theta}_0)X_0\Phi(\boldsymbol{\eta}_0) - A(\boldsymbol{\theta})X\Phi(\boldsymbol{\eta})]^T \right\} \xrightarrow{\text{a.s.}} 0. \quad (1.C.5)$$

Denote by  $\hat{\boldsymbol{\rho}}_{\text{OLS}}(K)$  the OLS estimate which minimizes  $\mathcal{Q}_K(\boldsymbol{\rho})$ . Using the above results, it follows that, as  $K \rightarrow \infty$ :

$$\mathcal{Q}_K(\hat{\boldsymbol{\rho}}_{\text{OLS}}(K)) \xrightarrow{\text{a.s.}} \text{tr}(\Sigma) + \mathcal{Q}(\boldsymbol{\rho}_{\text{OLS}}), \quad (1.C.6)$$

where  $\boldsymbol{\rho}_{\text{OLS}}$  is any limit point of the sequence of OLS estimators  $\widehat{\boldsymbol{\rho}}_{\text{OLS}}(K)$ . Moreover, since  $\mathcal{Q}_K(\widehat{\boldsymbol{\rho}}_{\text{OLS}}(K)) \leq \mathcal{Q}_K(\boldsymbol{\rho}_0)$ , we have

$$0 \leq \text{tr}(\Sigma) + \mathcal{Q}(\boldsymbol{\rho}_{\text{OLS}}) \leq \text{tr}(\Sigma), \quad (1.C.7)$$

which implies  $\mathcal{Q}(\boldsymbol{\rho}_{\text{OLS}}) = 0$ . Now, the regularity condition R5) implies that  $\boldsymbol{\rho}_{\text{OLS}} = \boldsymbol{\rho}_0$ ; hence  $\widehat{\boldsymbol{\rho}}_{\text{OLS}}(K) \xrightarrow{\text{a.s.}} \boldsymbol{\rho}_0$ , which is the result in (1.3.21a).

The proof of (1.3.21b) follows arguments similar to those in [24, chapter 4.3]. The expression

$$\left[ K \sum_{t=1}^N D(t, \boldsymbol{\rho})^T D(t, \boldsymbol{\rho}) \right]^{-1} \sum_{k=1}^K \sum_{t=1}^N D(t, \boldsymbol{\rho})^T \mathbf{w}_k(t) \quad (1.C.8)$$

has expected value  $\mathbf{0}$  and covariance  $O(K^{-1})$  because  $\mathbf{w}_k(t)$  are i.i.d. Then (1.3.21c) follows from [22, Corollary 5.1.1.2].

## 1.D ML versus OLS

We show that the ML estimates of  $\boldsymbol{\rho}$  derived in this chapter 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 7.1 implies that the asymptotic covariance matrix of  $\sqrt{NK}\widehat{\boldsymbol{\rho}}$  is (see also Appendix 1.G)

$$C_{\text{ML}}^\infty = NK [\mathcal{I}_{\text{signal}}(\boldsymbol{\theta})]^{-1} = NK \left[ \sum_{t=1}^N D(t, \boldsymbol{\rho})^T \Sigma^{-1} D(t, \boldsymbol{\rho}) \right]^{-1} = NK [D^T \mathcal{S}^{-1} D]^{-1}, \quad (1.D.1)$$

where

$$D(t, \boldsymbol{\rho}) = \frac{\partial(A(\boldsymbol{\theta})X\phi(t, \boldsymbol{\eta}))}{\partial \boldsymbol{\rho}^T}, \quad (1.D.2a)$$

$$D = [D(1, \boldsymbol{\rho})^T \cdots D(N, \boldsymbol{\rho})^T]^T, \quad (1.D.2b)$$

$$\mathcal{S} = I_N \otimes \Sigma, \quad (1.D.2c)$$

which can be further simplified [see (1.6.2) and (1.6.3)].

The asymptotic covariance matrix of  $\sqrt{NK}\widehat{\boldsymbol{\rho}}_{\text{OLS}}$  is (see Theorem 7.2)

$$C_{\text{OLS}}^\infty = NK [D^T D]^{-1} D^T \mathcal{S} D [D^T D]^{-1}. \quad (1.D.3)$$

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

$$T_d (T_d^T \mathcal{S} T_d)^{-1} T_d^T = \mathcal{S}^{-1} - \mathcal{S}^{-1} D (D^T \mathcal{S}^{-1} D)^{-1} D^T \mathcal{S}^{-1}, \quad (1.D.4)$$

which is Lemma 1 in [36] [see also [52, p. 77]]. It follows that

$$\begin{aligned} [D^T S^{-1} D]^{-1} &= [D^T D]^{-1} D^T S D [D^T D]^{-1} \\ &\quad - [D^T D]^{-1} D^T S T_d [T_d^T S T_d]^{-1} T_d^T S D [D^T D]^{-1}, \end{aligned} \quad (1.D.5)$$

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

## 1.E 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  $\Phi$ ), see (1.A.6) and (1.3.10b).

Using the Lemma in (1.D.4) we can compute  $Q(\boldsymbol{\theta})$  in the following alternative way:  $Q(\boldsymbol{\theta}) = T_a (T_a^T \hat{R} T_a)^{-1} T_a$ , 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  $I_{NK} - (1/NK) \cdot \tilde{Y}^T \hat{R}^{-1} \tilde{Y}$  in the denominator of (1.A.6) 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 non-zero with probability one. Note also that, for  $K = 1$ , the GLR in (1.A.6) would go to infinity if we choose the rows of  $\tilde{\Phi} = \Phi$  from the row space of  $\tilde{Y} = Y$ .

Note that  $\hat{X}$  can also be computed as a function of  $\hat{R}$  instead of  $S$ :

$$\begin{aligned} \hat{X} &= \sqrt{N} \left[ A(\boldsymbol{\theta})^T \hat{R}^{-1} A(\boldsymbol{\theta}) + A(\boldsymbol{\theta})^T \hat{R}^{-1} P [I_l - P^T \hat{R}^{-1} P]^{-1} P^T \hat{R}^{-1} A(\boldsymbol{\theta}) \right]^{-1} \times \\ &\quad \times A(\boldsymbol{\theta})^T \hat{R}^{-1} P [I_l - P^T \hat{R}^{-1} P]^{-1} (\Phi \Phi^T)^{-1/2}, \end{aligned} \quad (1.E.1)$$

where  $P$  is defined as

$$P = \bar{Y} \Phi^T (\Phi \Phi^T)^{-1/2} / \sqrt{N}. \quad (1.E.2)$$

Equation (1.E.1) follows from (1.3.7) and

$$S^{-1} = \hat{R}^{-1} + \hat{R}^{-1} P (I_l - P^T \hat{R}^{-1} P)^{-1} P^T \hat{R}^{-1}, \quad (1.E.3)$$

which is obtained by using the matrix inversion lemma.

Consider the basis function matrix of the following form:

$$\Phi = C \Delta^T (I_N - \frac{1}{N} \bar{Y}^T \hat{R}^{-1} \bar{Y})^{-1/2}, \quad (1.E.4)$$

where  $C$  is an  $l \times N$  matrix of full rank  $l$ , and  $\Delta$  is the matrix whose columns are the normalized eigenvectors of

$$\Xi(\boldsymbol{\theta}) = (I_N - \frac{1}{N} \bar{Y}^T \hat{R}^{-1} \bar{Y})^{-1/2} [I_N - \frac{1}{N} \bar{Y}^T Q(\boldsymbol{\theta}) \bar{Y}] (I_N - \frac{1}{N} \bar{Y}^T \hat{R}^{-1} \bar{Y})^{-1/2}, \quad (1.E.5)$$

ordered to correspond to the eigenvalues of  $\Xi(\boldsymbol{\theta})$  (denoted by  $\lambda_j$ ,  $j = 1, \dots, N$ ) sorted in non-increasing order, i.e.  $\lambda_1 \geq \lambda_2 \cdots \geq \lambda_N$ . Thus,

$$\Xi(\boldsymbol{\theta}) = \Delta \text{diag}\{\lambda_1, \dots, \lambda_N\} \Delta^T. \quad (1.E.6)$$

Also, denote by  $\Delta_l$  the matrix containing the first  $l$  columns of  $\Delta$ , which are the eigenvectors corresponding to the largest  $l$  eigenvalues of  $\Xi(\boldsymbol{\theta})$ . Then, (1.3.10b) reduces to

$$l_{\text{ML}}(\boldsymbol{\theta}, \boldsymbol{\eta}) = \frac{|C \Delta^T \Xi(\boldsymbol{\theta}) \Delta C^T|}{|CC^T|} = \frac{|C \text{diag}\{\lambda_1, \dots, \lambda_N\} C^T|}{|CC^T|} \quad (1.E.7)$$

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,

$$\widehat{\Phi} = H \Delta_l^T (I_N - \frac{1}{N} \bar{Y}^T \widehat{R}^{-1} \bar{Y})^{-1/2}. \quad (1.E.8)$$

For  $H = I_l$ , the rows of  $\widehat{\Phi}$  are the generalized eigenvectors of the matrices  $I_N - \frac{1}{N} \bar{Y}^T Q(\boldsymbol{\theta}) \bar{Y}$  and  $I_N - \frac{1}{N} \bar{Y}^T \widehat{R}^{-1} \bar{Y}$  that correspond to the largest  $l$  generalized eigenvalues of these two matrices; the product of these eigenvalues is

$$l_{\text{ML}}(\boldsymbol{\theta}) = \prod_{j=1}^l \lambda_j. \quad (1.E.9)$$

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

$$\begin{aligned} \Xi(\boldsymbol{\theta}) = I_N + \frac{1}{N} \cdot (I_N - \frac{1}{N} \bar{Y}^T \widehat{R}^{-1} \bar{Y})^{-1/2} \bar{Y}^T \widehat{R}^{-1} A(\boldsymbol{\theta}) [A(\boldsymbol{\theta})^T \widehat{R}^{-1} A(\boldsymbol{\theta})]^{-1} \\ \cdot A(\boldsymbol{\theta})^T \widehat{R}^{-1} \bar{Y} (I_N - \frac{1}{N} \bar{Y}^T \widehat{R}^{-1} \bar{Y})^{-1/2}. \end{aligned} \quad (1.E.10)$$

The second term in (1.E.10) is a positive semidefinite symmetric matrix with rank  $\min(\text{rank}(A(\boldsymbol{\theta})), N)$ , which equals  $\text{rank}(A(\boldsymbol{\theta})) = nr$  in most practical applications.

We now show that, although  $\widehat{\Phi}$  is not unique, the moment temporal evolution  $\widehat{X} \widehat{\Phi} = [\widehat{\mathbf{s}}(1) \widehat{\mathbf{s}}(2) \cdots \widehat{\mathbf{s}}(N)]$  is. Using (1.3.7a), we get

$$[\widehat{\mathbf{s}}(1) \widehat{\mathbf{s}}(2) \cdots \widehat{\mathbf{s}}(N)] = \widehat{X} \widehat{\Phi} = [A(\boldsymbol{\theta})^T S^{-1} A(\boldsymbol{\theta})]^{-1} A(\boldsymbol{\theta})^T S^{-1} \bar{Y} \Pi_{\widehat{\Phi}}, \quad (1.E.11)$$

where  $\Pi_{\widehat{\Phi}}$ , the projection matrix on the row space of  $\widehat{\Phi}$ , is independent of  $H$  because it cancels out. Since  $S$  depends on  $\Phi$  only through  $\Pi_{\Phi}$  [see (1.3.8)],  $[\widehat{\mathbf{s}}(1) \widehat{\mathbf{s}}(2) \cdots \widehat{\mathbf{s}}(N)]$  is also independent of  $H$ .

Substituting  $H = [\Delta_l^T (I_N - \frac{1}{N} \bar{Y}^T \widehat{R}^{-1} \bar{Y})^{-1} \Delta_l]^{-1/2}$  into (1.E.8), we get orthonormal basis functions, i.e.  $\widehat{\Phi} \widehat{\Phi}^T = I_l$ .

When  $Y_k = [Y_{1k}, Y_{2k}]$ ,  $k = 1, \dots, K$ , where  $Y_{1k}$  contains baseline data and  $Y_{2k}$  contains the evoked response, (1.3.10b) becomes

$$l_{\text{ML}}(\boldsymbol{\theta}, \boldsymbol{\eta}) = \frac{|\Phi_2 (I_{N_2} - \frac{1}{N} \bar{Y}_2^T Q(\boldsymbol{\theta}) \bar{Y}_2) \Phi_2^T|}{|\Phi_2 (I_{N_2} - \frac{1}{N} \bar{Y}_2^T \widehat{R}^{-1} \bar{Y}_2) \Phi_2^T|}. \quad (1.E.12)$$

The concentrated likelihood function  $\text{GLR}(\boldsymbol{\theta})$  is the product of  $l$  largest eigenvalues of  $(I_{N_2} - \frac{1}{N} \bar{Y}_2^T \hat{R}^{-1} \bar{Y}_2)^{-1/2} [I_{N_2} - \frac{1}{N} \bar{Y}_2^T Q(\boldsymbol{\theta}) \bar{Y}_2] (I_{N_2} - \frac{1}{N} \bar{Y}_2^T \hat{R}^{-1} \bar{Y}_2)^{-1/2}$ . Thus, the result in Section 1.4 follows.

## 1.F Scanning

We derive the scanning scheme (1.5.5) based on matching the estimated array response subspace (constructed using an ML estimate of nonparametric array response) with a single-dipole array response.

Consider the nonparametric array response model, i.e. assume that  $A(\boldsymbol{\theta}) = A$  is an unknown  $m \times nr$  matrix of full rank, equal to  $nr$ . Also, assume that the dipole moments are fully uncorrelated, i.e.  $\Phi = I_N$ . In the following, we compute an ML estimate of  $A$  by maximizing the concentrated likelihood function in (1.5.1b) [where  $A(\boldsymbol{\theta})$  is replaced with  $A$ ] with respect to  $A$ .

Consider the array response matrix of the form  $A = \hat{R}^{1/2} V \mathcal{C}$ , where  $\mathcal{C}$  is an  $m \times nr$  matrix of full rank  $nr$  and  $V$  is the matrix whose columns are the (normalized) eigenvectors of

$$\Psi = \frac{1}{N} \hat{R}^{-1/2} \bar{Y} \bar{Y}^T \hat{R}^{-1/2} \quad (1.F.1)$$

that are ordered to correspond to the eigenvalues of  $\Psi$  (denoted by  $\mu_j, j = 1, \dots, m$ ) sorted in non-increasing order, i.e.  $\mu_1 \geq \mu_2 \cdots \geq \mu_m$ . Thus,

$$\Psi = V \text{diag}\{\mu_1, \dots, \mu_m\} V^T.$$

Then, the concentrated likelihood function in (1.5.1b) becomes

$$\begin{aligned} l(\boldsymbol{\theta}) &= \frac{|\mathcal{C}^T V^T \hat{R}^{1/2} (\hat{R} - \frac{1}{N} \bar{Y} \bar{Y}^T)^{-1} \hat{R}^{1/2} V \mathcal{C}|}{|\mathcal{C}^T \mathcal{C}|} \\ &= \frac{|\mathcal{C}^T [I_m - \text{diag}\{\mu_1, \dots, \mu_m\}]^{-1} \mathcal{C}|}{|\mathcal{C}^T \mathcal{C}|}, \end{aligned} \quad (1.F.2)$$

which follows from

$$\hat{R}^{1/2} \left( \hat{R} - \frac{1}{N} \bar{Y} \bar{Y}^T \right)^{-1} \hat{R}^{1/2} = V [I_m - \text{diag}\{\mu_1, \dots, \mu_m\}]^{-1} V^T \quad (1.F.3)$$

Since  $0 \leq (\bar{Y}^T \hat{R}^{-1} \bar{Y})/N \leq I_N$ , the non-zero eigenvalues of  $(\bar{Y}^T \hat{R}^{-1} \bar{Y})/N$  (equal to the non-zero eigenvalues of  $\Psi$ ) are between 0 and 1; thus,  $0 \leq \mu_j \leq 1$ , where  $j = 1, \dots, m$ . Note that  $1/(1-\mu_1) \geq 1/(1-\mu_2) \cdots \geq 1/(1-\mu_m)$ , because  $1/(1-\mu_j)$  is an increasing function of  $\mu_j$  for  $0 \leq \mu_j < 1$ . Now, it is obvious that (1.F.2) is maximized for  $\mathcal{C} = [I_{nr}, 0]^T \cdot \mathcal{H}$  where  $\mathcal{H}$  is an arbitrary  $nr \times nr$  matrix of full rank. Thus, an ML estimate of  $A$  is of the following form:

$$\hat{A} = \hat{R}^{1/2} V_{nr} \mathcal{H}, \quad (1.F.4)$$

where  $V_{nr}$  denotes the matrix containing the first  $nr$  columns of  $V$ . Further,  $I_m - V_{nr}V_{nr}^T$  is the projection matrix onto the space orthogonal to the column space of  $\widehat{R}^{-1/2}\widehat{A}$ . Therefore, for a single dipole located at  $\boldsymbol{\theta}$ ,

$$(I - V_{nr}V_{nr}^T)\widehat{R}^{-1/2}A(\boldsymbol{\theta}) \approx 0, \quad (1.F.5)$$

and a MUSIC-like scanning function easily follows (using e.g. [62]) as an inverse of the minimum generalized eigenvalue of  $A(\boldsymbol{\theta})^T\widehat{R}^{-1/2}[I - V_{nr}V_{nr}^T]\widehat{R}^{-1/2}A(\boldsymbol{\theta})$  and  $A(\boldsymbol{\theta})^T\widehat{R}^{-1}A(\boldsymbol{\theta})$ . Note also that the columns of  $U_{nr} = \widehat{R}^{-1/2}V_{nr}$  are the generalized eigenvectors of  $\frac{1}{N}\overline{Y}Y^T$  and  $\widehat{R}$  corresponding to their  $nr$  largest generalized eigenvalues (equal to  $\mu_1, \mu_2, \dots, \mu_{nr}$ ); thus

$$A(\boldsymbol{\theta})^T\widehat{R}^{-1/2}[I_m - V_{nr}V_{nr}^T]\widehat{R}^{-1/2}A(\boldsymbol{\theta}) = A(\boldsymbol{\theta})^T[\widehat{R}^{-1} - U_{nr}U_{nr}^T]A(\boldsymbol{\theta}). \quad (1.F.6)$$

## 1.G Derivation of the Fisher Information Matrix

We derive the FIM for the model in Section 1.3. Define  $\boldsymbol{\mu}(t, \boldsymbol{\rho}) = A(\boldsymbol{\theta})X\boldsymbol{\phi}(t, \boldsymbol{\eta})$ . Then, the negative log-likelihood function is

$$l(\boldsymbol{\gamma}) = \sum_{k=1}^K \sum_{t=1}^N \left\{ [\mathbf{y}_k(t) - \boldsymbol{\mu}(t, \boldsymbol{\rho})]^T \Sigma^{-1} [\mathbf{y}_k(t) - \boldsymbol{\mu}(t, \boldsymbol{\rho})] + \log |\Sigma| \right\}. \quad (1.G.1)$$

Thus, the  $(i, j)$ th entry of the FIM easily follows from [49] and [33, eq. (3.31)]

$$\begin{aligned} [\mathcal{I}(\boldsymbol{\gamma})]_{ij} &= K \sum_{t=1}^N \frac{\partial \boldsymbol{\mu}(t, \boldsymbol{\rho})^T}{\partial \gamma_i} \Sigma(\boldsymbol{\psi})^{-1} \frac{\partial \boldsymbol{\mu}(t, \boldsymbol{\rho})}{\partial \gamma_j} \\ &\quad + \frac{K}{2} \sum_{t=1}^N \text{tr} \left[ \frac{\partial \Sigma(\boldsymbol{\psi})}{\partial \gamma_i} \Sigma(\boldsymbol{\psi})^{-1} \frac{\partial \Sigma(\boldsymbol{\psi})}{\partial \gamma_j} \Sigma(\boldsymbol{\psi})^{-1} \right]. \end{aligned} \quad (1.G.2)$$

Using a well-known formula in e.g. [29, th. 16.2.2]

$$\text{tr} A^T B C D^T = [\text{vec } A]^T [D \otimes B] \text{vec } C, \quad (1.G.3)$$

we can rewrite (1.G.2) as

$$\begin{aligned} \mathcal{I}(\boldsymbol{\gamma}) &= \begin{bmatrix} K \sum_{t=1}^N D(t, \boldsymbol{\rho})^T \Sigma(\boldsymbol{\psi})^{-1} D(t, \boldsymbol{\rho}) & 0 \\ 0 & KNH(\boldsymbol{\psi})^T V(\boldsymbol{\psi})^{-1} H(\boldsymbol{\psi}) \end{bmatrix} \\ &= \begin{bmatrix} \mathcal{I}_{\text{signal}}(\boldsymbol{\gamma}) & 0 \\ 0 & \mathcal{I}_{\text{noise}}(\boldsymbol{\psi}) \end{bmatrix}, \end{aligned} \quad (1.G.4)$$

where

$$D(t, \boldsymbol{\rho}) = \frac{\partial (A(\boldsymbol{\theta})X\boldsymbol{\phi}(t, \boldsymbol{\eta}))}{\partial \boldsymbol{\rho}^T} = [D_x(t, \boldsymbol{\rho}), D_o(t, \boldsymbol{\rho}), D_n(t, \boldsymbol{\rho})], \quad (1.G.5a)$$

$$H(\boldsymbol{\psi}) = \frac{\partial \text{vec}(\Sigma(\boldsymbol{\psi}))}{\partial \boldsymbol{\psi}^T}, \quad (1.G.5b)$$

$$V(\boldsymbol{\psi}) = 2 \Sigma(\boldsymbol{\psi}) \otimes \Sigma(\boldsymbol{\psi}), \quad (1.G.5c)$$

and

$$D_x(t, \boldsymbol{\rho}) = \frac{\partial(A(\boldsymbol{\theta})X\boldsymbol{\phi}(t, \boldsymbol{\eta}))}{\partial \text{vec}(X)^T} = \boldsymbol{\phi}(t, \boldsymbol{\eta})^T \otimes A(\boldsymbol{\theta}), \quad (1.G.6a)$$

$$D_\theta(t, \boldsymbol{\rho}) = \frac{\partial(A(\boldsymbol{\theta})X\boldsymbol{\phi}(t, \boldsymbol{\eta}))}{\partial \boldsymbol{\theta}^T} = (X\boldsymbol{\phi}(t, \boldsymbol{\eta}) \otimes I_m)^T \frac{\partial \text{vec}(A(\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}^T}, \quad (1.G.6b)$$

$$D_\eta(t, \boldsymbol{\rho}) = \frac{\partial(A(\boldsymbol{\theta})X\boldsymbol{\phi}(t, \boldsymbol{\eta}))}{\partial \boldsymbol{\eta}^T} = A(\boldsymbol{\theta})X \frac{\partial \boldsymbol{\phi}(t, \boldsymbol{\eta})}{\partial \boldsymbol{\eta}^T} \quad (1.G.6c)$$

follow from [72, properties xiv. and xv. at p. 15]. Finally, (1.6.3) follows by using the above results and the identity

$$(\mathbf{a} \otimes A)B(\mathbf{c}^T \otimes C) = \mathbf{a}\mathbf{c}^T \otimes ABC. \quad (1.G.7)$$

Also, from (1.G.2), it follows:

$$I_{\text{noise}}(\boldsymbol{\psi}) = \frac{NK}{2} \text{tr}[\Sigma^{-1} \frac{\partial \Sigma}{\partial \psi_r} \Sigma^{-1} \frac{\partial \Sigma}{\partial \psi_s}], \quad r, s = 1, \dots, \frac{1}{2}m(m+1), \quad (1.G.8)$$

which can be further simplified, see (1.6.5).

---

---

# BIBLIOGRAPHY

- [1] A. C. Atkinson and A. N. Donev, *Optimum Experimental Designs*. London:Oxford University Press, 1992.
- [2] A. C. Atkinson, K. Chaloner, A. M. Herzberg, and J. Juritz, “Optimum Experimental Designs for Properties of a Compartmental Model,” *Biometrics*, vol. 49, pp. 325–337, June 1993.
- [3] S. Baillet, J. C. Mosher, and R. M. Leahy, “Electromagnetic brain mapping,” *IEEE Signal Processing Mag.*, vol. 18, No. 6, pp. 14–30, Nov. 2001.
- [4] D. Bates, “The derivative of  $|X'X|$  and its use,” *Technometrics*, vol. 25, pp. 373–376, Nov. 1983.
- [5] H. Bunke and O. Bunke, “Identifiability and estimability,” *Math. Operationsforsch. Statist.*, vol. 5, pp. 223–233, 1974.
- [6] P. J. Bickel and K. A. Doksum, *Mathematical Statistics: Basic Ideas and Selected Topics*. Upper Saddle River, NJ: Prentice Hall, 2nd ed., 2000.
- [7] 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.
- [8] D. Brenner, J. Lipton, L. Kaufmann, and S. J. Williamson, “Somatically evoked magnetic fields of the human brain,” *Science*, vol. 199, pp. 81–83, 1978.
- [9] O. Bria, C. Muravchik, and A. Nehorai, “EEG/MEG error bounds for a dynamic dipole source with a realistic head model,” *Methods of Information in Medicine*, vol. 2, pp. 110–113, 2000.
- [10] B. S. Clarke and A. R. Barron, “Information-theoretic asymptotics of Bayes methods,” *IEEE Trans. Inform. Theory*, vol. 36, pp. 453–471, May 1990.
- [11] T.M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: John Wiley & Sons, 1991.

- [12] 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.
- [13] 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.
- [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, No. 8, pp. 791–804, Aug. 1992.
- [15] A. Dogandžić and A. Nehorai, "Detecting a dipole source by MEG/EEG and generalized likelihood ratio tests," *Proc. 30th Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, Nov. 1996, pp. 1196–1200.
- [16] A. Dogandžić and A. Nehorai, "Estimating evoked dipole responses by MEG/EEG for unknown noise covariance," *Proc. 19th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Chicago, IL, Oct. 1997, pp. 1224–1227.
- [17] A. Dogandžić and A. Nehorai, "Localization of evoked electric sources and design of EEG/MEG sensor arrays," *Proc. 9th IEEE SP Workshop on Statistical Signal and Array Processing*, pp. 228–231, Portland, OR, Sept. 1998.
- [18] A. Dogandžić, and A. Nehorai, "Estimating range, velocity, and direction with a radar array," *Proc. Intl. Conf. on Acoust., Speech, Signal Processing*, pp. 2773–2776, Phoenix, AZ, March 1999.
- [19] A. Dogandžić and A. Nehorai, "Estimating evoked dipole responses in unknown spatially correlated noise with EEG/MEG arrays," *IEEE Trans. Signal Processing*, vol. 48, pp. 13–25, Jan. 2000.
- [20] A. Dogandžić and A. Nehorai, "Cramér-Rao bounds for estimating range, velocity, and direction with an active array," *IEEE Trans. Signal Processing*, vol. 49, pp. 1122–1137, June 2001.
- [21] A. Dogandžić and A. Nehorai, "Space-time fading channel estimation and symbol detection in unknown spatially correlated noise," *IEEE Trans. Signal Processing*, vol. 50, pp. 457–474, March 2002.
- [22] W.A. Fuller, *Introduction to Statistical Time Series*. New York: John Wiley & Sons, 2nd ed., 1996.
- [23] A. R. Gallant, "Seemingly unrelated nonlinear regressions," *J. Econometrics*, vol. 3, pp. 35–50, 1975.
- [24] A. R. Gallant, *Nonlinear Statistical Models*. New York: John Wiley & Sons, 1987.

- [25] T. Gasser, J. Möchs, and W. KöhlerC., “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.
- [26] A. B. Geva, H. Pratt, and Y. Y. Zeevi, “Spatio-temporal multiple source localization by wavelet-type decomposition of evoked potentials,” *Electroencephalogr. Clin. Neurophysiol.*, vol. 96, pp. 278–286, 1995.
- [27] C. Gennings, V. M. Chinchilli, and W. H. Carter, “Response surface analysis with correlated data: a nonlinear model approach,” *J. Amer. Stat. Asso.*, vol. 84, pp. 805–809, 1989.
- [28] G. H. Golub and G. P. H. Styan, “Numerical computations for univariate linear models,” *J. Statist. Comput. Simul.*, vol. 2, pp. 253–274, 1973.
- [29] D. A. Harville, *Matrix Algebra From a Statistician’s Perspective*. New York:Springer-Verlag, 1997.
- [30] M. S. Hämläinen, Hari R., Ilmoniemi R., Knuutila J., and Lounasmaa O.V., “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, April 1993.
- [31] B. Hochwald and A. Nehorai, “Magnetoencephalography with diversely-oriented and multi-component sensors,” *IEEE Trans. Biomed. Eng.*, vol. 44, pp. 40–50, Jan. 1997.
- [32] 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.
- [33] S. M. Kay, *Fundamentals of Statistical Signal Processing — Estimation Theory*. Englewood Cliffs, NJ: Prentice Hall, 1993.
- [34] E. J. Kelly, “An adaptive detection algorithm,” *IEEE Trans. Aerospace and Electron. Syst.*, vol. AES-22, No. 2, pp. 115–127, Mar. 1986.
- [35] E. J. Kelly, and K. M. Forsythe, “Adaptive detection and parameter estimation for multidimensional signal models,” Technical Report 848, Lincoln Laboratory, Massachusetts Institute of Technology, April 1989.
- [36] 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.
- [37] T. O. Kvålseth, “Cautionary note about  $R^2$ ,” *The American Statistician*, vol. 39, No. 4, pp. 279–285, 1985.

- [38] B. Lütkenhöner, "Magnetic field arising from current dipoles randomly distributed in a homogeneous spherical volume conductor," *J. Appl. Phys.*, vol. 75, No. 11, pp. 7204–7210, 1994.
- [39] B. Lütkenhöner, "Dipole source localization by means of maximum likelihood estimation I. Theory and simulations," *Electroencephalogr. Clin. Neurophysiol.*, vol. 106, pp. 314–321, 1998.
- [40] B. Lütkenhöner, "Dipole source localization by means of maximum likelihood estimation II. Experimental evaluation," *Electroencephalogr. Clin. Neurophysiol.*, vol. 106, pp. 322–329, 1998.
- [41] B. Lütkenhöner and O. Steinsträter, "High-precision neuromagnetic study of the functional organization of the human auditory cortex," *Audiol. Neurootol.*, vol. 3, pp. 191–213, 1998.
- [42] C. D. McGillem, J. I. Aunon, and C. A. Pomalaza, "Improved waveform estimation procedures for event-related potentials," *IEEE Trans. Biomed. Eng.*, vol. 32, No. 6, pp. 371–379, June 1985.
- [43] 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, Sep. 1975.
- [44] J. Möchs, Pham D.T., and T. Gasser, "Testing for homogeneity of noisy signals evoked by repeated stimuli," *Annals of Statistics*, vol. 12, pp. 193–209, 1984.
- [45] J. C. Mosher, M. E. Spencer, R. M. Leahy, and P. S. Lewis, "Error bounds for MEG and EEG dipole source localization," *Electroencephalogr. Clin. Neurophysiology*, vol. 86, pp. 303–321, 1993.
- [46] C. Muravchik and A. Nehorai, "EEG/MEG error bounds for a static dipole source with a realistic head model," *IEEE Trans. Signal Processing*, vol. 49, pp. 470–484, Mar. 2001.
- [47] A. Nehorai and A. Dogandžić, "Estimation of propagating dipole sources by EEG/MEG sensor arrays," *Proc. 32nd Asilomar Conf. Signals, Syst. Comput.*, pp. 304–308, Pacific Grove, CA, Nov. 1998.
- [48] C. Pantev, M. Hoke, B. Lütkenhöner, and K. Lehnertz, "Tonotopic organization of the auditory cortex: pitch versus frequency representation," *Science*, vol. 246, pp. 486–488, 1989.
- [49] B. Porat, *Digital Processing of Random Signals: Theory and Methods*. Englewood Cliffs, NJ: Prentice Hall, 1994.
- [50] 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.

- [51] J. Raz and B. Turetsky, "Event-related potentials," *Encyclopedia of Biostatistics*, vol. 2, pp. 1407–1409. Chichester UK: John Wiley & Sons, 1998.
- [52] C. R. Rao, *Linear Statistical Inference and Its Applications*. New York: John Wiley & Sons, 2nd ed., 1973.
- [53] C.R. Rao and J. Kleffe, *Estimation of Variance Components and Applications*. New York: North-Holland, 1988.
- [54] M. Scherg and D. Von Cramon, "Two bilateral sources of the late AEP as identified by a spatio-temporal dipole model," *Electroencephalogr. Clin. Neurophysiol.*, vol. 62, pp. 32–44, 1985.
- [55] M. Scherg and D. Von Cramon, "Evoked dipole source potentials of the human auditory cortex," *Electroencephalogr. Clin. Neurophysiol.*, vol. 65, pp. 344–360, 1986.
- [56] H. P. Schwan and C. F. Kay, "Capacitive properties of body tissues," *Circ. Res.*, vol. 4, pp. 439–443, 1957.
- [57] 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, No. 6, pp. 558–562, June 1992.
- [58] M. S. Srivastava and C. G. Khatri, *An Introduction to Multivariate Statistics*. New York: North Holland, 1979.
- [59] A. L. Swindlehurst and P. Stoica, "Maximum likelihood methods in radar array signal processing," *Proc. IEEE*, vol. 86, pp. 421–441, Feb. 1998.
- [60] 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.
- [61] 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.
- [62] J. C. Mosher, P. S. Lewis, and Leahy R.M., "Multiple dipole modeling and localization from spatio-temporal MEG data," *IEEE Trans. Biomed. Eng.*, vol. 39, pp. 541–557, June 1992.
- [63] 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.
- [64] 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.

- [65] S. Supek and C.J. Aine, "Simulation studies of multiple dipole neuromagnetic source localization: model order and limits of source resolution", *IEEE Trans. Biomed. Eng.*, vol. 40, No. 6, pp. 529-540, June 1993.
- [66] P. Stoica and B. C. Ng, "On the Cramér-Rao bound under parametric constraints," *IEEE Signal Processing Letters*, vol. 5, pp. 177-179, July 1998.
- [67] M. E. Spencer, "Spatio-temporal modeling, estimation, and detection for EEG and MEG," Ph.D. Dissertation, Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, CA, 1995.
- [68] 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.
- [69] H. L. Van Trees, *Detection, Estimation and Modulation Theory*, Part I. New York: John Wiley & Sons, 1968.
- [70] 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, April 1997.
- [71] 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.
- [72] E. F. Vonesh and V. M. Chinchilli, *Linear and Nonlinear Models for the Analysis of Repeated Measurements*. New York: Marcel Dekker, 1997.
- [73] A. Wald, "Tests of statistical hypotheses concerning several parameters when the number of observations is large," *Trans. Amer. Math. Soc.*, vol. 54, pp. 426-482, 1943.
- [74] M. Wax and I. Ziskind, "On unique localization of multiple sources by passive sensor arrays," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 996-1000, July 1989.
- [75] 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.
- [76] T. Yamazaki, B. W. van Dijk, and H. Spekreijse, "The accuracy of localizing equivalent dipoles and the spatio-temporal correlations of background EEG," *IEEE Trans. Biomed. Eng.*, vol. 45, pp. 1114-1121, Sept. 1998.
- [77] 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.

---

---

# SUBJECT INDEX

- array response matrix, 396, 397, 399, 407, 408, 435
  - nonparametric, 407, 435
- basis functions, 395, 399, 406, 407, 411, 417, 418, 427, 430, 431
  - nonparametric, 395, 401, 402, 405, 406, 411, 418, 419, 427, 430, 433
  - orthonormal, 434
- Capon, 408
- covariance matrix, 399, 409
  - asymptotic, 432
  - estimated, 411
  - noise, 395, 400
- CRB, 395, 402, 409–415
  - decoupling, 411
- dipole, 393–399, 402, 403, 405–419, 421–427, 430, 435, 436
- DOA
  - estimation, 401
- EEG, 393–397, 399, 407, 411–413, 416, 427
- eigenvalue, 406, 409, 434, 435
  - generalized, 406–409, 434, 436
- eigenvector, 409, 433–436
  - generalized, 406
- FIM, 395, 402, 409, 410, 436
- GMANOVA, 400
- goodness-of-fit, 395, 411, 419, 426
- identifiability, 401, 429, 430
- imaging
  - brain, 393, 395
- information theoretic criteria, 414
- invariance
  - reparametrization, 414
- least squares
  - estimated generalized (EGLS), 403
  - extended (ELS), 402
  - ordinary (OLS), 395, 402, 403, 405, 416–418, 421, 426, 427, 431, 432
- likelihood function, 401, 402, 428, 429, 436
  - concentrated, 395, 401, 405–407, 425, 428, 429, 435
- MEG, 393–397, 399, 407, 411–413, 416, 419, 427, 430
- ML, 395, 401–403, 405, 418, 427
  - estimate(s), 395, 400, 401, 403–406, 418, 421, 426, 428, 432, 435
  - estimation, 405, 407
  - estimator, 402, 409
  - method, 395, 402–404, 410, 417–419, 426, 427
- MRI, 393, 408, 421–424, 427
- MUSIC, 407, 408

mutual information, 414

relative entropy, 414

scanning, 395, 403, 407–409, 416–418

sensor array design, 412, 413

$D$ -optimality, 413–415

$D_s$ -optimality, 413–415

    Bayesian, 415

spectral estimate

    Capon, 408