The Wald test and Cramér-Rao Bound for misspecified models in electromagnetic source analysis

Lourens J. Waldorp, Hilde M. Huizenga, Raoul P.P.P. Grasman

Abstract— **By using signal processing techniques an estimate can be obtained of activity in the brain from the electro- or magneto-encephalogram (EEG or MEG). For a proper analysis, a test is required to indicate whether the model for brain activity fits. A problem in using such tests is that often not all assumptions are satisfied, like the assumption of the number of shells in EEG. In such a case a test on the number of sources (model order) might still be of interest. A detailed analysis is presented of the Wald test for these cases. One of the advantages of the Wald test is that it can be used when not all assumptions are satisfied. Two different, previously suggested Wald tests in electromagnetic source analysis (EMSA) are examined: a test on source amplitudes and a test on the closeness of source pairs. The Wald test is analytically studied in terms of alternative hypotheses that are close to the null hypothesis (local alternatives). It is shown that the Wald test is asymptotically unbiased, that is has the correct level and power, which makes it appropriate to use in EMSA. An accurate estimate of the Cramér-Rao bound (CRB) is required for the use of the Wald test when not all assumptions are satisfied. The sandwich CRB is used for this purpose. It is defined for non-separable least squares with constraints required for the Wald test on amplitudes and MEG. Simulations with EEG show that when the sensor positions are incorrect, or the number of shells is incorrect, or the conductivity parameter is incorrect, then the CRB and Wald test are still good, with a moderate number of trials. Additionally, the CRB and Wald test appear robust against an incorrect assumption on the noise covariance. A combination of incorrect sensor positions and noise covariance does affect the possibility of detecting a source with small amplitude.**

Keywords— **Source localization, separable least squares, approximate model, model checking, parameter covariance, constrained optimization, Fisher-information with constraints**

I. Introduction

IN electromagnetic source analysis (EMSA) an estimate is obtained of the activity in the brain from the electro- or magneto-encephaolgram (EEG or MEG) measured at or near the scalp [1]. The inverse problem in which an estimate is obtained relies on many assumptions [2], [3]. Some examples in EEG are the number of shells [4], their conductivities [5], [6], their shapes, the type of source (i.e. a dipole or quadrupole etc., see e.g., [7], [8]), and the number of sources (see [9] when dipoles are used).

A model comprised of several such assumptions is only an approximation of what is really going on. In [7] it is shown that dipoles are a good approximation of extended sources, but in [9] and [10, Ch. 3] it is shown that using the incorrect number of dipoles can have serious effects on localization. Also in [11] it is shown that the distance between the brain and skull has a large effect on sensitivity of both EEG and MEG. In addition, statistical assumptions, like the distribution of the noise, can affect the accuracy of the dipole parameters greatly [12].

Even if some of the assumptions like the number of shells in EEG may be incorrect, it can still be of interest to know how many sources are present (model order). Many methods to detect sources and methods to determine the model order exist. These methods vary from purely statistical based techniques like the residual variance [9], or sequential tests based on singular value decompositions [13], or multivariate testing [14] to signal processing techniques like MUSIC [15], [16], and to biophysical model based approaches [17], [18]. However, these methods do not explicitly incorporate the fact that some of the assumptions could be incorrect.

In the present paper, the notion of an approximate (incorrect) model used in the statistical analysis is made precise and is used to determine a proper goodness-of-fit (GOF) and model selection test. A GOF test determines whether the model fits the data or not, and a model selection test compares different models, for instance, two models with different number of sources. The Wald test seems well suited for both these purposes because it tests parameters directly and has asymptotically relatively large power for hypotheses that are very 'similar' (local alternatives). The Wald test has been used before in similar settings [19], [20]. Two Wald tests are used in the context of EMSA, which have been suggested in [21] for a single time sample, and in [22] for a spatio-temporal model. The first is a test on source amplitudes. The idea is that if a source has (nearly) zero amplitude it should be considered as inactive. The other is a test on source 'closeness'. If the (Euclidean) distance between two sources is small, then it might be too difficult to distinguish them and they could be better modeled as one.

Both Wald tests proved to be useful when used as a model selection test to determine model order (i.e. the number of sources). In [22] the Wald tests were compared for the spatio-temporal model to the Akaike and Bayesian information criteria (AIC and BIC), and the residual variance. Both the AIC and the Wald tests performed well, but the Wald tests have several advantages over the AIC: (1) the Wald tests indicate whether a model fits (GOF), whereas the AIC can only compare models, (2) each dipole can be tested, with the Wald test on amplitudes, to determine at which time samples it is active, and finally (3) the Wald tests can be used even if not all assumptions are satisfied. For example, if the sensor positions are incorrectly measured in EEG, then model order can still be determined by the Wald test.

A prerequisite for using the Wald test in the way described above is that a good estimate of the Cramér-Rao bound (CRB) of the parameters is available, even for approximate models. It has been shown before (e.g., [23], [24]) that using the Hessian

*Lourens J. Waldorp is both with the Department of Psychology, University of Amsterdam, Roetersstraat 15, 1018 WB, Amsterdam, and with the Department of Neurocognition, Maastricht University, Maastricht, The Netherlands, email: waldorp@psy.uva.nl. The Netherlands Organization for Scientific Research (NWO) is gratefully acknowledged for funding this project.

H.M. Huizenga, and R.P.P.P. Grasman are with the Department of Psychology, University of Amsterdam, Amsterdam, The Netherlands

matrix in least squares (LS), or the Fisher information matrix in maximum likelihood (ML), is inappropriate if the model is approximate. This means that the CRB as described in [25] is not appropriate if, for example, the assumed number of shells in EEG is incorrect. Therefore, a different estimator for the CRB should be used, the so-called sandwich estimator. The sandwich estimator is not new (e.g., [26], [27], [28]). The essential difference between the traditional CRB and the sandwich is that the latter involves no assumption on the model being correct. The expression is consequently slightly more involved than the traditional CRB, but still only contains first-order partial derivatives of the model. In a study on categorical time series it was found that the sandwich estimator gives very precise estimates of the parameter variance in spite of possible modeling errors [29]. In [30] it was shown that the Wald test using the sandwich estimator is a level α test (i.e. it rejects the null hypothesis when it is true with probability not grater than α, typically 0.05), even when the model is approximate. This means that some of the assumptions about the model can be incorrect but that the Wald test still is a level α test.

Since the biophysical function in EEG and MEG has both a nonlinear and a linear part, nonlinear separable least squares (NSLS) is used to estimate the parameters [31], [32]. In [28] the asymptotic CRB of all estimated parameters in NSLS is derived, although constraints are not incorporated. Using the Wald test on amplitudes requires a reparameterization which in turn requires constraints to render the problem identifiable [12]. For MEG in a spherically symmetric head model only the tangential part is measured [2], and so a constraint is required to account for this. In the present paper, constraints in NSLS are incorporated, together with the fact that models are approximations. This problem is defined for M-estimators, which refers to maximizing a general function of parameters. An example is Huber's function which is designed to minimize the influence of extreme data points [33].

The paper is organized as follows. First the NSLS method is briefly described, followed by the reparameterization and incorporating the ensuing constraints in the CRB. Then the Wald test is described, together with some performance properties. Finally, a numerical example is given, in which EEG data are simulated and analyzed in terms of the Wald test and standard errors for the parameters.

II. Parameter and CRB estimation

Data from p sensors on independent trials $j = 1, \ldots, n$ are collected in the p vector Y_j with average $\bar{Y} = \frac{1}{n}\sum_{j=1}^{n} Y_j$. A model for the mean $E\{Y_j\} = \mu$ is constructed such that hopefully $\mu = f_\theta(X)$, where θ is a q vector of parameters, and X denotes the (fixed) sensor locations (orientations). Then

$$Y_j = f_\theta(X) + r_j. \tag{1}$$

The residual r_j contains both modeling error and noise e_j. The noise e_j refers to all irrelevant activation (see e.g., [34]) and has mean zero and covariance matrix Σ. In that case the noise is said to be colored (i.e. correlated and heteroscedastic). If, on the other hand, $\Sigma = \sigma^2 I_p$, where I_p is the $p \times p$ identity matrix, then the noise is white (i.e. uncorrelated and homoscedastic). The model can be extended to capture more than a single time sample, where either just the dipole locations are fixed over time [35] or both the locations and orientations are fixed over time [36], [37].

Modeling error arises if the true mean μ cannot be attained with the function f_θ used in the regression for any θ. A model for the mean is then said to be approximate or incorrect. This is, for example, the case if the incorrect number of shells are used in EEG or if the sensor positions used in the inverse problem are incorrect. Then $\mu \neq f_\theta$ for any θ considered. Consequently, the residual $r_j = \mu - f_\theta + e_j$ contains both modeling error $\mu - f_\theta$ and noise e_j. Note that a rotation or translation of the sensors does not fall into the category of an approximate model, since then there exists a θ with a rotated coordinate system or translated origin such that $\mu = f_\theta$. In addition to incorrect assumptions about the model of the mean, assumptions about the noise characteristics could be incorrect. For instance, the noise could be assumed white whereas in fact it is colored. In this example the standard errors derived from the CRB are then usually too small [12]. In general such incorrect assumptions have a deleterious effect on the estimate of the CRB and on tests dependent on an estimate of the CRB, like the Wald test [24].

The source parameters in θ can be obtained by minimizing

$$Q_\theta(Y) = (\bar{Y} - f_\theta)'\Sigma^{-1}(\bar{Y} - f_\theta). \tag{2}$$

For EEG and MEG the biophysical function f_θ can be simplified by explicitly using the partial linearity of the function, such that $f_\theta = G_\tau \beta$, where τ are the $3d$ location and β the $3d$ moment parameters of d sources, and $\theta' = (\tau', \beta')$ with $q = 6d$. In NSLS, for a given τ the linear estimate $\hat{\beta} = (G_\tau'\Sigma^{-1}G_\tau)^{-1}G_\tau'\Sigma^{-1}\bar{Y}$ is used to construct the minimization function [31]

$$Q_\tau(Y) = \bar{Y}'\Sigma^{-1}(I_p - P_G)\bar{Y}, \tag{3}$$

where $P_G = G_\tau(G_\tau'\Sigma^{-1}G_\tau)^{-1}G_\tau'\Sigma^{-1}$. Minimizing Q_τ with respect to τ yields the NSLS estimate $\hat{\tau}$. If the model f_θ is incorrect then the estimate $\hat{\theta}$ will be biased with respect to the true parameter, that is $E\{\hat{\theta} - \theta_0\} \neq 0$. However, there does exist θ_* which minimizes the prediction mean squared error $E\{Q_\tau\}$. In this case θ_* can be interpreted as the LS approximation f_{θ_*} of the true but unknown function μ [23].

Let the first-order partial derivative $p \times q$ matrix be $\dot{f}_\theta = F_\theta$. The asymptotic CRB of $\hat{\theta}$ is derived in [28]. If $\hat{\theta}$ converges in probability to θ_*, then $\sqrt{n}(\hat{\theta} - \theta_*)$ has asymptotic covariance

$$J^{-1}IJ^{-1} = \\ \big(F_\theta'\Sigma^{-1}F_\theta\big)^{-1}\big(F_\theta'\Sigma^{-1}R\Sigma^{-1}F_\theta\big)\big(F_\theta'\Sigma^{-1}F_\theta\big)^{-1}, \tag{4}$$

where $R = E\{r_j r_k'\}$, which is zero if the trials j and k are unequal since it was assumed that trials are independent. It is assumed that f_θ has continuous first-order partial derivatives F_θ of full rank at θ_* and that the linear approximation of f_θ is sufficient (see for this last point [38, Ch. 4]). The square root of the diagonal elements of (4) are the standard errors, which are used to construct confidence intervals.

By noting that $R = (\mu - f_\theta)(\mu - f_\theta)' + \Sigma$, it can be seen that if $f_\theta = \mu$, then $R = \Sigma$. The assumption of a correct model is then seen to lead to the familiar, less involved expression for the CRB, which is $2J^{-1} = (F_\theta'\Sigma^{-1}F_\theta)^{-1}$. If, in addition, the noise

is white, $\Sigma = \sigma^2 I_p$, then $2J^{-1} = \sigma^2(F_\theta' F_\theta)^{-1}$. The CRB in (4) is valid asymptotically even if an approximate model is used [30], where for example, the sensor positions are incorrect.

In practice neither Σ nor R are known and have to be estimated in order to obtain an estimate of the CRB. For Σ an unbiased and asymptotically consistent estimate is (e.g., [12], [39], [40], [41])

$$S = \frac{1}{n-1}\sum_{j=1}^{n}(Y_j - \bar{Y})(Y_j - \bar{Y})', \tag{5}$$

when $n > p$. This is a common and practical choice if the interest is not in noise characteristics. Other estimates of the noise covariance matrix based on parametric models for the noise covariance (e.g., [34], [42]) could also be used, provided they are consistent for Σ which requires that the model of the noise covariance is correct. Using a consistent estimate like S may pose a problem in finite samples, because then it may not be very close to Σ. This is especially problematic in testing hypotheses, as in the Wald test, since the test will not have the expected distribution. As will be shown, this problem in the Wald test can be avoided by using the sandwich estimate. An estimate of R is given by

$$\hat{R} = \frac{1}{n}\sum_{j=1}^{n}(Y_j - f_{\hat{\theta}})(Y_j - f_{\hat{\theta}})', \tag{6}$$

when $n > p$. This estimate is an extension of the commonly used version for the case of white noise, i.e. $\Sigma = \sigma^2 I_p$ (see e.g., [23], [43]).

From the estimates of the noise covariance above, three estimators of the CRB can be constructed: two from the familiar one $2J^{-1}$, and one from the sandwich $J^{-1}IJ^{-1}$. First of all, to compute the CRB the estimate $\hat{\theta}$ is substituted for θ_*. Let $\hat{\sigma}^2 = Q_{\hat{\tau}}/(p-q)$. Then the familiar estimate of the CRB is $2\hat{\sigma}^2 J^{-1}$, where S is used instead of Σ. If either an assumption for the mean or covariance is incorrect, then this estimate is expected to perform poorly. This is because either $\hat{\sigma}^2$ is biased or S is not close to Σ. Let $s^2 = \frac{1}{p(n-1)}\sum_j (Y_j - \bar{Y})'(Y_j - \bar{Y})$ which is the average of the diagonal of S. Then an alternative CRB estimate, based on the familiar one, is $2s^2 J^{-1}$, where again S is used. This estimate is expected to be better than the familiar one because there is no modeling error in s^2. There is still, however, the possibility that S is not close to Σ. The third one is the sandwich estimator $J^{-1}IJ^{-1}$ with $\hat{R}$ and S. The sandwich estimator is expected to perform well even if both the model for the mean and noise covariance are incorrect. Asymptotically, for the true model, these three estimators are equivalent, and so the finite sample behavior in different situations is of interest.

III. CRB FOR CONSTRAINED MODELS

If constraints are required, because the model f_θ requires these or because of identifiability, then the CRB in (4) has to be modified. If θ is not identifiable then J is singular [44], and so the CRB in (4) cannot be estimated. In [45] a method to incorporate constraints into the CRB is described for maximum likelihood. For more general estimation methods (M-estimators), a slightly different version is required. This version is still based on finding simultaneously the Lagrange multipliers λ and the parameters θ, but assumes continuity of the function Q_θ and of a constraints function c_θ around θ_* in the form of a Lipschitz condition [33, Th. 5.23]. Let c_θ denote a vector with r constraints, and the $r \times q$ matrix $\dot{c}_\theta = C_\theta$ its first-order derivative, and assume that C_θ has full row-rank r. Furthermore, let $J_+ = J + C_\theta' C_\theta$ which is non-singular if $\mathrm{im}(F_\theta) \cap \mathrm{im}(C_\theta') = \{0\}$ [44]. Then estimating the parameters θ under constraints c_θ with Lagrange multipliers is done by finding a solution for the set of equations [45]

$$\begin{pmatrix} J_+ & C_\theta' \\ C_\theta & O \end{pmatrix}\begin{pmatrix} \hat{\theta} - \theta_* \\ \hat{\lambda} - \lambda_* \end{pmatrix} = \begin{pmatrix} \dot{Q}_\theta \\ O \end{pmatrix},$$

where O is a matrix (or vector) of zeros with appropriate dimensions. Under the regularity conditions of theorem 5.23 in [33], and the assumption that the augmented matrix on the left, denoted by J_a, is non-singular, the result on M-estimators can be used. The result is summarized in lemma 1.

Lemma 1. Assume that C_θ has full row-rank, and that J_a is non-singular. If the regularity conditions of Theorem 5.23 in [33, p. 53] are satisfied, then

$$\sqrt{n}\begin{pmatrix} \hat{\theta} - \theta_* \\ \hat{\lambda} - \lambda_* \end{pmatrix} = \begin{pmatrix} J_+ & C_\theta' \\ C_\theta & O \end{pmatrix}^{-1}\begin{pmatrix} \sqrt{n}\dot{Q}_\theta \\ O \end{pmatrix} + o_p(1), \tag{7}$$

where $o_p(1)$ denotes a sequence that converges to zero in probability [33]. The vector $\sqrt{n}\dot{Q}_\theta = -2F_\theta'\Sigma^{-1}\frac{1}{\sqrt{n}}\sum_{j=1}^{n} r_j$, is assumed to be multivariate normal with mean zero and covariance matrix $I = F_\theta'\Sigma^{-1}R\Sigma^{-1}F_\theta$. Then, the CRB of $\sqrt{n}(\hat{\theta} - \theta_*)$ is $J_c^{-1}IJ_c^{-1}$ with $J_c^{-1} = J_+^{-1}(I_q - C_\theta'(C_\theta J_+^{-1} C_\theta')^{-1} C_\theta J_+^{-1})$. If the true model is used, then the CRB becomes $2J_c^{-1}$ [45]. The number of free parameters has to be adjusted accordingly, and becomes $q_f = q - r$.

A. Application to EMSA

The CRB $J_c^{-1}IJ_c^{-1}$ can be used for the source parameters in EMSA. The reparameterization of the model with amplitudes separated from the moments requires constraints for both EEG and MEG. Additionally, for MEG in a homogeneous spherical head model, identifiability constraints need to be imposed to ensure that the location is orthogonal to the moment of each source [2]. The spatio-temporal model with fixed locations [35] or fixed locations and orientations of the dipoles over time samples [36], [37] induces no additional constraints.

For both EEG and MEG the reparameterization consists of the simple transformation such that the moment parameters in β are split-up into orientation parameters and an amplitude parameter for each source. This is done by the transformation $\beta = B\alpha$, where B is $3d \times d$ and block diagonal with in each block the normalized orientation vector $\beta_i^n = \beta_i/\alpha_i$ for $i = 1, \ldots, d$, and α contains the amplitudes [15]. The corresponding constraint is that the norm of the orientation vector of each dipole equals 1, that is $||\beta_i^n|| - 1 = 0$. If MEG is used then only the tangential part of the dipole is measured and so it is also required that the location and orientation vector are orthogonal, that is $\tau_i'\beta_i^n = 0$ for each dipole.

All constraints can be put in the vector c_θ. The Jacobian of the constraints vector is the matrix C_θ. The expressions for the matrices J, I, and C_θ required for the CRB are given in the Appendix for EEG and MEG.

IV. Wald test

The Wald test determines whether $h_\theta = 0$ for a given k vector valued function of the parameters. The null hypothesis for testing is $H_0 : h_\theta = 0$, and the alternative is denoted by $H_1 : h_\theta \neq 0$. By using the delta method (e.g., [33, Ch. 3] it can be seen that if h_θ is continuous and if the first-order partial derivative $k \times q$ matrix $\dot{h}_\theta = H_\theta$ has full row-rank k, then the covariance matrix of h_θ is $H_\theta D_\theta H'_\theta$, where D_θ is the covariance matrix of $\hat{\theta}$. The Wald test is then defined as [38]

$$W_{\hat{\theta}} = n h'_{\hat{\theta}} (H_{\hat{\theta}} D_{\hat{\theta}} H'_{\hat{\theta}})^{-1} h_{\hat{\theta}}. \quad (8)$$

If Y_j are normal, then under H_0 the Wald test is asymptotically χ^2_k distributed [38]. Only linear hypotheses of the form $h_\theta = U\theta = 0$ are considered.

Informally, a test is asymptotically consistent if it can distinguish between the null and alternative distribution with sufficient data [33]. However, nearly all statistical tests are asymptotically consistent, and so it is a necessary but not a sufficient condition to use a test. To evaluate the performance of a statistical test, it is more informative to look at alternative hypotheses that are close to the null hypothesis, since this makes it harder to distinguish the two [33], [46]. An example is misspecifying an amplitude parameter of a source by a small amount. The alternative hypothesis is then referred to as a local alternative. A local alternative for the parameter $\theta = \theta_0 + \phi_0/\sqrt{n}$, or equivalently, $\phi_0 = \sqrt{n}(\theta - \theta_0)$, for the true value θ_0 under H_0, is defined as $H_1 : h_\theta = h_{\theta_0 + \phi_0/\sqrt{n}} \neq 0$, written as h_{ϕ_0}. These local alternatives are much more difficult to distinguish from the null hypothesis than other alternatives.

The linear hypothesis is of the form $U\theta = 0$ and so the first-order derivative is $\dot{h}_\theta = H = U$. The local alternative for an estimate $\hat{\phi}_0 = \sqrt{n}(\hat{\theta} - \theta_0)$ is then $h_{\hat{\phi}_0} = U\theta_0 + U\hat{\phi}_0/\sqrt{n}$. Since $U\theta_0 = 0$ by hypothesis, the Wald test can be rewritten as

$$W_{\hat{\phi}_0} = \hat{\phi}'_0 U' (U D_{\hat{\phi}_0} U')^{-1} U \hat{\phi}_0. \quad (9)$$

The distribution of $W_{\hat{\phi}_0}$ is asymptotically distributed as χ^2_k for local alternatives [46]. This test is in fact on the noncentrality parameter $\delta_{\phi_0} = \phi'_0 U' (U D_{\phi_0} U')^{-1} U \phi_0$ of the noncentral distribution $\chi^2_k(\delta)$.

The power function of the test is defined as the probability that H_0 is rejected at some specified level α [33]. Let Z be $N(0, I_p)$ distributed, $\Omega^{1/2}$ the Cholesky factor of $U D_\phi U'$, and K_α the quantile of the appropriate distribution determined at level α. Then the power function is

$$P_{\hat{\phi}}(W_{\hat{\phi}} > K_\alpha) \to 1 - P_\phi(||Z + \Omega^{-1/2} U \phi||^2 \leq K_\alpha), \quad (10)$$

as $n \to \infty$. A test is asymptotically a level α test if the probability that H_0 is rejected given that H_0 is true is at most α. If, in addition, the test rejects H_0 when H_1 is true with probability 1, called the power of the test, then it is asymptotically consistent [47]. The Wald test can be seen to be asymptotically consistent, because for ϕ_0 under H_0, $1 - P_{\phi_0}(||Z + \Omega^{-1/2} U \phi_0||^2 \leq K_\alpha) \leq \alpha$, and for alternatives $\phi_* = \sqrt{n}(\theta - \theta_*)$ under H_1, $1 - P_{\phi_*}(||Z + \Omega^{-1/2} U \phi_*||^2 \leq K_\alpha) \to 1$. As said before, most statistical tests are asymptotically consistent, and so local alternatives should be considered.

From the power function it can be seen that if ϕ is close to 0, then the test is a level α test. On the other hand if ϕ is not close to 0, then the power tends to 1. The aim is then to find out for which convergence rates of $\hat{\phi} = \sqrt{n}(\hat{\theta} - \theta_0)$ the power function does not go to either 0 or 1. If $\hat{\phi}$ has the asymptotic representation as in (7) from lemma 1, then $\hat{\theta} - \theta_0$ converges to zero at rate $O(n^{-1/2})$. For estimators that converge to zero at this rate, the power is strictly between α and 1 [33]. Such a test is called asymptotically unbiased if it is also asymptotically a level α test. This means that the Wald test has higher power than several other level α tests, which do not share this property, for hypotheses that are difficult to distinguish. It is clearly seen from in the power function, that $\Omega^{-1/2}$, which includes both the parameter and noise covariance matrix, directly influences the power of a test. So the power for local alternatives depends both on the estimate of the CRB and of the noise covariance matrix.

In practice only finite samples are available and so the Wald test might be adjusted to incorporate the distribution of the estimates involved. As described in the previous section, if all assumptions on the model are correct and H_0 is true, the matrix D_θ can be either $2\hat{\sigma}^2 J_c^{-1}$ or $2s^2 J_c^{-1}$. The distributions of both $\hat{\sigma}^2$ and s^2 are known to be $\chi^2_{n-q_f}$ and χ^2_{n-p} respectively, when H_0 is true [38]. Then the Wald test can be written as a ratio of two independent chi-square distributions, which is F-distributed if the degrees of freedom are incorporated. The result is summarized in the following proposition, which is a small extension of the results in [38, p. 231].

Proposition 2. Let Y_j be $N(\mu, \Sigma)$ and $H_0 : U\theta = 0$ is true. Then for known Σ

$$(i) \quad W_{\hat{\theta}}(2\hat{\sigma}^2 J_c^{-1}) = \frac{\hat{\theta}' U' (U J_c^{-1} U')^{-1} U \hat{\theta}}{2k\hat{\sigma}^2} \sim F_{k, p-q_f}$$

$$(ii) \quad W_{\hat{\theta}}(2s^2 J_c^{-1}) = \frac{\hat{\theta}' U' (U J_c^{-1} U')^{-1} U \hat{\theta}}{2ks^2} \sim F_{k, n-p}.$$

This still is not optimal, since S is not taken into account, but should be more accurate than the χ^2_k distribution. If not all assumptions are satisfied but $H_0 : U\theta = 0$ is true, then D_θ should be $J_c^{-1} I J_c^{-1}$. In that case White has shown [30] that only the Wald test with the sandwich $W_{\hat{\theta}}(J_c^{-1} I J_c^{-1})$ is a level α test. In section V Numerical Example, the robustness of these tests is investigated for several incorrect assumptions.

A. Application to EMSA

The above results are used in the following to define two different Wald tests for EMSA [21]. The first test is on source amplitudes. If a source has small amplitude, then it can be considered inactive. The second is a test on the (Euclidean) distance between paired sources. If the distance between two sources is very small, then they can be better modeled by one source.

For the Wald test on source amplitudes (denoted by W_α), the linear hypothesis is $U = (O_{d,6d}, I_d)$, where $O_{d,6d}$ denotes a matrix with zeros of dimensions $d \times 6d$. This gives $U\theta = \alpha$ the source amplitudes and the covariance $U D_\theta U' = D_\alpha$ for the

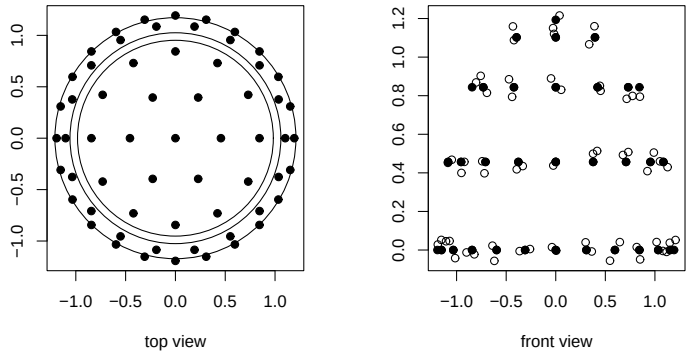

Fig. 1. Left: The sensors and schematic representation of the radii. Right: The 61 sensor configuration in its original form (solid dots) and the perturbed sensor configuration (open circles).

amplitudes. If (i) is used from proposition 2, then W_α is distributed as $F_{d,p-q_f}$. The test W_α is an omnibus test in the sense that it determines whether any of the sources is active. Subsequent univariate tests can be performed to see whether each of the sources individually is active. A univariate test with estimator (i) is distributed as $F_{1,p-q_f}$.

For the Wald test on the distance between the sources (denoted by W_τ), it has to be considered which pairs are tested. If H_0 is true then the distances between the sources is zero. And so only all comparisons of one source with each of the others is required. For example, if three sources are tested, and H_0 is true, then $\tau_1 - \tau_2 = 0$ and $\tau_1 - \tau_3 = 0$, and so $\tau_2 - \tau_3 = \tau_1 - \tau_3$ and it was seen that this is zero. For the $d-1$ comparisons with the first source, the linear hypothesis is $U = (1_{d-1} \otimes I_3, -I_{3(d-1)})$. With estimator (i) from proposition 2, W_τ is distributed as $F_{3(d-1),p-q_f}$. As before, W_τ is an omnibus test, and determines whether any or a linear combination of source pairs has large distance. Subsequent semi-univariate tests on source pairs individually can be performed to see which source pairs are distant. A semi-univariate test on a source pair with estimator (i) is distributed as $F_{3,p-q_f}$.

Both tests can be used for GOF and model order selection. The GOF is given by the omnibus test, which indicates by a non-significant test result that the model fits. Model order selection can be performed by repeatedly using the univariate Wald tests. If the univariate Wald test is significant for each of the sources, then a model with an additional source can be tested for significance. The model with the fewest sources for which the univariate Wald tests still indicate significance, is finally selected.

V. Numerical example

In the numerical example several of the above results are illustrated. First, the differences of the three estimators of the CRB is examined for different sample sizes and for correct and incorrect model assumptions. Next, it is shown for the three different CRB estimators whether the null distribution of the Wald test in porposition 2 is accurate for different sample sizes. Subsequently, it is determined whether the Wald test is a level α test. And lastly, the power of the Wald test is investigated by using local alternatives.

EEG data are generated by two dipoles in a spherical head model with three isotropic and concentric shells on 61 sensors (see Fig. 1). The radii are for the brain, skull, and skin are respectively 0.87, 0.92, and 1 [48], and the corresponding conductivities are 1, 0.80, and 1 $(\Omega m)^{-1}$. So-called Berg-parameters were estimated to approximate the three shell EEG truncated series (to 70 terms) with the above mentioned radii and conductivities [49], [50]. The dipoles have locations $(0, \pm 0.4, 0.7)$ with eccentricity 0.81, both with orientation $(1, 0, 0)$ and amplitude 1.

Noise distributed as $N(0, \Sigma)$ is added to the sensors. The matrix Σ is defined as $\sigma_{ij} = \exp(-d_{ij}/a)$ where d_{ij} is the Euclidean distance between sensors and a is a correlation parameter set to 0.8. The noise variance for the mean was set at 10% of the maximum output of the dipole model f_θ. Three interesting noise conditions are possible: noise is generated as $e_j \sim N(0, \sigma^2 I_p)$ and estimation with $\hat{\Sigma} = s^2 I_p$, noise is generated as $e_j \sim N(0, \Sigma)$ and estimation with $\hat{\Sigma} = S$, and finally, noise is generated as $e_j \sim N(0, \Sigma)$ and estimation with $\hat{\Sigma} = s^2 I_p$, as if the noise were white.

To show what happens when an assumption in the biophysical model is incorrect, either the sensors positions are perturbed or a one shell head model is used instead of the three-shell head model. For the sensor perturbation a small amount of uniformly distributed noise (between -0.02 and 0.02) is added to the sensor positions. The data are generated with the perturbed sensor configuration and are estimated with the original version (Fig. 1, right). Both assuming incorrect sensor positions and an incorrect number of shells introduce a small bias in the source parameter estimates, and it is investigated to what extent it affects the CRB estimators and Wald tests.

A. CRB estimators

To examine the difference between the estimators of the CRB, several situations are considered. The model f_θ can be correct or incorrect (correct or perturbed sensor configuration, or three or one shell head model) and the statistical assumptions on the noise covariance matrix can be correct or incorrect (estimated noise covariance is the same as generated or not). To show the performance of the estimators, the ratio of estimated to the true standard errors (se) averaged over all source parameters is used, which should be 1. The true standard errors are defined as the standard deviation of the estimates of 100 simulations.

In Fig. 2 it can be seen in the left panel that if all model assumptions are correct then the CRB estimates of all three different estimators are approximately equal and asymptotically correct (close to 1). To investigate the effect of the size of the modeling error on the CRB estimate, the skull conductivity parameter is varied such that a difference between the true en incorrect one is between 0.05 and 0.20. This is shown in the right panel of Fig. 2. It can be seen that the sandwich estimate is closest to 1 and that the other two decrease more rapidly as the size of the difference increases. If the number of trials is increased, then the sandwich CRB estimate gets closer to 1. The left panel of Fig. 3 shows that the noise covariance estimate S determines the accuracy of the CRB estimates. As S converges to Σ with increasing n, the CRB estimates slowly tend to 1. If $e_j \sim N(0, \Sigma), \hat{\Sigma} = s^2 I_p$, right panel, then only the sandwich estimator is close to 1, and is not deteriorated by S. Since

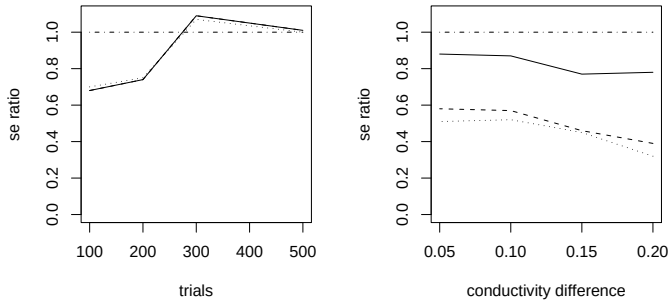

Fig. 2. The ratio of se averaged over all parameters for $J_c^{-1}IJ_c^{-1}$ $(-)$, $2\hat{\sigma}^2 J_c^{-1}$ $(\cdots)$, and $2s^2 J_c^{-1}$ $(--)$ and with $n = 100$. The line $(\cdot - \cdot)$ is at the correct ratio 1. Left: correct head model and $e_j \sim N(0, \sigma^2 I_p), \hat{\Sigma} = s^2 I_p$. Right: ratio se as a function of conductivity difference and $e_j \sim N(0, \Sigma), \hat{\Sigma} = s^2 I_p$.

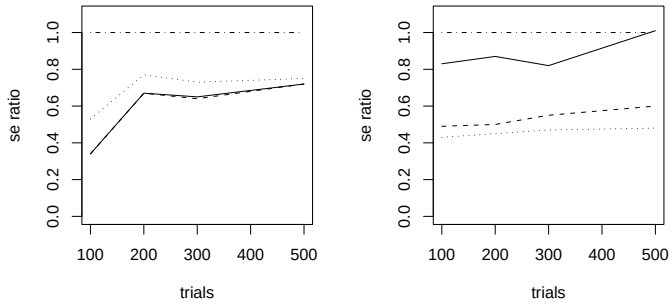

Fig. 3. Similar as in Fig. 2 but for perturbed sensor positions. Left: $e_j \sim N(0, \Sigma), \hat{\Sigma} = S$. Right: $e_j \sim N(0, \Sigma), \hat{\Sigma} = s^2 I_p$.

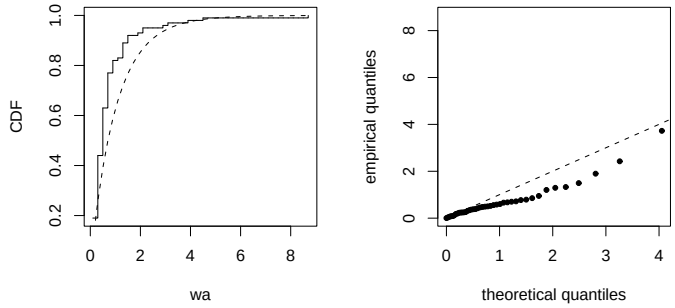

Fig. 4. Left: The empirical CDF $(-)$ of $W_\alpha(J_c^{-1}IJ_c^{-1})$ and the theoretical CDF $(--)$ for $n = 100$, $e_j \sim N(0, \Sigma), \hat{\Sigma} = s^2 I$, and the perturbed sensor configuration. Right: A q-q plot of the empirical $(\cdots)$ and theoretical CDF $(--)$.

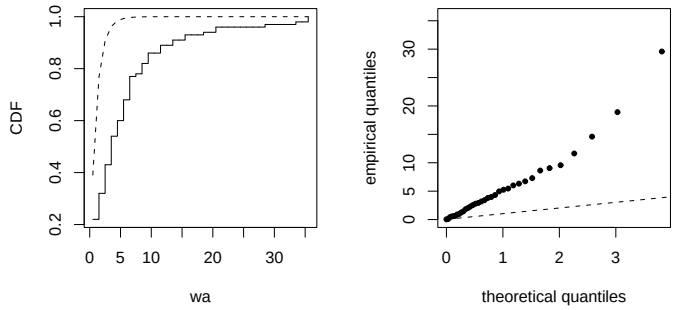

Fig. 5. Same as in Fig. 4 for the test $W_\alpha(2\hat{\sigma}^2 J_c^{-1})$, $n = 100$, and $e_j \sim N(0, \Sigma), \hat{\Sigma} = \sigma^2 I_p$.

the sandwich CRB estimate remains fairly stable for a range of modeling error of the conductivity and is similar to the error in the CRB obtained with the perturbation of the sensor positions, only the latter misspecification is shown in the remainder of the paper.

B. Wald test

To examine whether the theoretically assumed distribution of the Wald test corresponds to the empirical distribution, both a cumulative distribution function (CDF) and a quantile-quantile (q-q) plot are used. The first gives the overall correspondence of the distribution functions and the latter gives more precise information on correspondence at the specific quantiles. Information on large quantiles (tail behavior) is important for the level of the test. This is shown for the Wald test on source amplitudes, but similar results have been obtained with the test on source location.

When all assumptions are satisfied the correspondence between the theoretical and empirical distribution is high, which confirms proposition 2. In Fig. 4 it can be seen that when the senor positions are incorrect and the noise covariance is incorrectly assumed white, the empirical distribution follows the theoretical one reasonably well with $W_\alpha(J_c^{-1}IJ_c^{-1})$ for $n = 100$, but not well in the tails. This is confirmed by the q-q plot. However, Fig. 5 shows that for the same situation $W_\alpha(2\hat{\sigma}^2 J_c^{-1})$ does not come close to the assumed distribution, especially in the tails (see the q-q plot).

No significant difference was found in any of the conditions between using the χ^2 and F distribution for the Wald tests. This is because S is not accounted for, and with the current parameters $n = 100$ trials are sufficient for accurate estimation of s^2 and $\hat{\sigma}^2$.

To determine whether the W_α is a level α test, the probability of rejecting the null hypothesis when it is true, which should be at most α, is computed as described in the previous section. In this example $\alpha = 0.05$, as is commonly done [47].

For white noise W_α is a level α test for all trial conditions and the correct sensor positions. However, as can be seen in Fig. 6 left panel, when the noise is colored and S is used, then all tests have elevated levels up to $n = 300$ trials. For large n (asymptotically) all three tests are level α tests. If the white noise assumption is incorrect, then the right panel of the same figure shows that only $W_\alpha(J_c^{-1}IJ_c^{-1})$ is a level α test. When the sensor positions are incorrect, the left panel of Fig. 7 shows slightly higher levels when the noise is colored compared to when the sensor positions are correct. It is clear that the inaccuracies in S overshadow the effects of modeling error from the perturbed sensor positions. The combination of perturbed sensor positions and the incorrect assumption of white noise, right panel of Fig. 7, results in highly elevated levels except for $W_\alpha(J_c^{-1}IJ_c^{-1})$, which remains around the correct level of 0.05.

To study the power of W_α with local alternatives, a hypothesis is used that is close to the null hypothesis (local alternative). In the simulations the null hypothesis corresponds to the source amplitudes of 1. A local alternative would be for example source amplitudes of 0.95 or 0.90. Then the probability is investigated that the null hypothesis is rejected when H_1 is true but quite close to H_0. All power calculations are performed with $n = 100$ trials to get realistic results. When the noise is white, the power

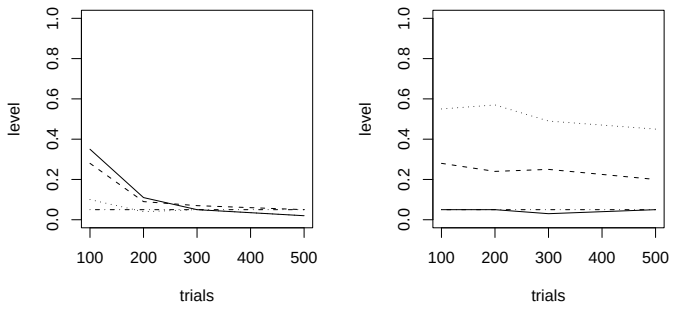

Fig. 6. Level of $W_\alpha(J_c^{-1}IJ_c^{-1})$ $(-)$, $W_\alpha(2\hat{\sigma}^2 J_c^{-1})$ $(\cdots)$, and $W_\alpha(2s^2 J_c^{-1})$ $(--)$ for the correct sensor positions. The line $(\cdot - \cdot)$ is at 0.05, the correct level of the test. Left: $e_j \sim N(0,\Sigma), \hat{\Sigma} = S$. Right: $e_j \sim N(0,\Sigma), \hat{\Sigma} = s^2 I_p$.

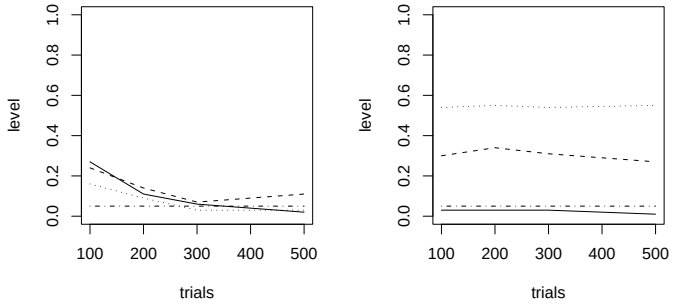

Fig. 7. Same as in Fig. 6 but with with perturbed sensor postitions.

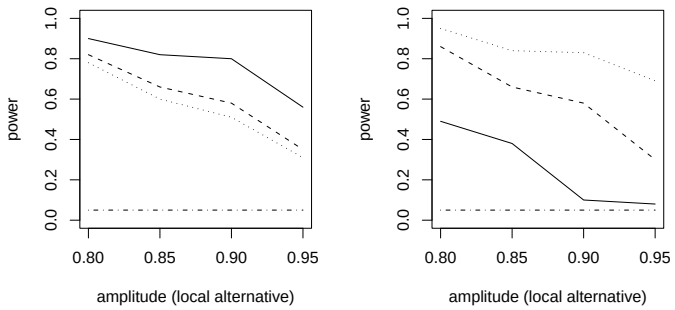

Fig. 8. Power for $W_\alpha(J_c^{-1}IJ_c^{-1})$ $(-)$, $W_\alpha(2\hat{\sigma}^2 J_c^{-1})$ $(\cdots)$, and $W_\alpha(2s^2 J_c^{-1})$ $(--)$ for the correct sensor positions and $n = 100$. The line $(\cdot - \cdot)$ is at 0.05. Left: $e_j \sim N(0,\Sigma), \hat{\Sigma} = S$. Right: $e_j \sim N(0,\Sigma), \hat{\Sigma} = s^2 I_p$.

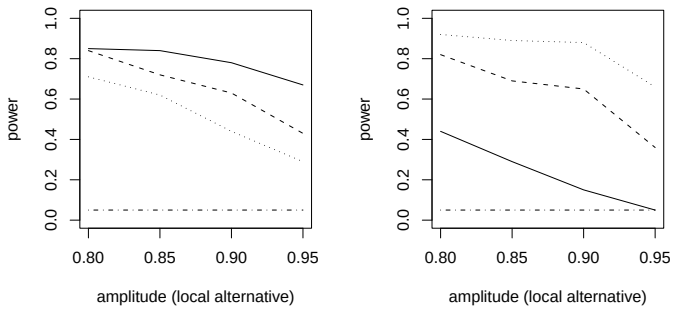

Fig. 9. Same as Fig. 8 for the perturbed sensor configuration.

of the three tests is nearly identical. Comparing the correct and incorrect sensor positions, Fig. 8 and 9 respectively, with the correct and incorrect noise assumption, left and right panel respectively, shows a similar pattern: when the noise is colored $W_\alpha(J_c^{-1}IJ_c^{-1})$ has the highest power, then $W_\alpha(2\sigma^2 J_c^{-1})$, and finally $W_\alpha(2s^2 J_c^{-1})$; the order is reversed when the noise is incorrectly assumed to be white. The power of the univariate tests in Fig. 10, show that $W_\alpha(J_c^{-1}IJ_c^{-1})$, the only level α test, has relatively little power when the noise is incorrectly assumed to be white (right panel).

VI. Conclusion

It was shown for which class of estimators the Wald test is optimal in the sense that it is asymptotically unbiased (i.e. it is a level α test and has power larger than or equal to α). Two different Wald tests were defined for GOF and model order selection purposes in EMSA: a test on source amplitudes and a test on pairwise source distance. From the simulations the Wald test appears to be a good test to perform both GOF and model order selection. The multivariate test with the sandwich estimator is asymptotically unbiased even though with modeling error for the mean and noise the distribution did not fit well for a small number of trials. The power of the multivariate test for local alternatives was quite reasonable when either the correct or incorrect model for the mean and noise was used. However, the univariate test had little power for local alternatives when both the model for the mean and noise were incorrect. This means that low amplitude sources are more difficult to detect when both mean and noise assumptions are not satisfied.

A prerequisite for using the Wald test with an incorrect model for the mean and noise, was that good CRB estimates exist. It was shown that a good estimate of the CRB can be obtained with the sandwich estimator. Correct CRB estimates were obtained even when the assumption on the noise was incorrect. When the incorrect model for the mean was assumed, the CRB estimators were reasonable for larger number of trials.

Two types of modeling error for the mean were investigated for EEG: incorrect sensor positions and incorrect number of shells. In both cases the Wald test with the sandwich CRB appeared to be robust. Additionally, the size of the error in the conductivity parameter of the skull appeared to be of little influence on the sandwich CRB. Larger modeling errors or a combination of several (interacting) modeling errors require further investigation.

Appendix

The model is $f_\theta = G_\tau B\alpha$, with $B = (\text{diag}(\alpha)^{-1} I_d \otimes 1_3')\text{diag}(\beta)$, α contains the d amplitudes, $\otimes$ is the Kronecker

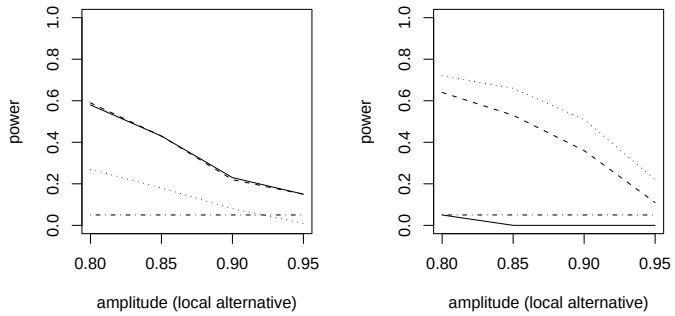

Fig. 10. Same as Fig. 8 for univariate tests and the perturbed sensor configuration.

product, $1_3 = (1,1,1)'$, and diag$(\cdot)$ constructs a diagonal matrix of a vector. The vector of the reparameterized dipole parameters is $\theta' = (\tau', (\beta^n)', \alpha')$. The matrices $J = F_\theta' \Sigma^{-1} F_\theta$, $I = F_\theta' \Sigma^{-1} R \Sigma^{-1} F_\theta$, and $C_\theta = \dot{c}_\theta$ are given in terms of the reparameterization.

Let the first-order partial derivatives $\dot{f}_\theta = F_\theta$ with respect to τ, β^n, and α be, respectively, $\dot{G}_\tau = (\alpha' B' \otimes \Sigma^{-1/2} G_\tau) \partial G_\tau$, $\dot{B} = (\alpha' \otimes \Sigma^{-1/2} G_\tau) \partial B$, and $\Sigma^{-1/2} G_\tau B$, where $\partial A(\phi) = \partial \text{vec}(A) / \partial \phi$ [44]. Then

$$J = \begin{pmatrix} J_\tau & J_{\tau\beta} & J_{\tau\alpha} \\ J_{\beta\tau} & J_\beta & J_{\beta\alpha} \\ J_{\alpha\tau} & J_{\alpha\beta} & J_\alpha \end{pmatrix}, \quad I = \begin{pmatrix} I_\tau & I_{\tau\beta} & I_{\tau\alpha} \\ I_{\beta\tau} & I_\beta & I_{\beta\alpha} \\ I_{\alpha\tau} & I_{\alpha\beta} & I_\alpha \end{pmatrix}$$

with

$$\begin{aligned} J_\tau &= \dot{G}_\tau' (B\alpha\alpha' B' \otimes \Sigma^{-1}) \dot{G}_\tau, & J_\beta &= \dot{B}' (\alpha\alpha' \otimes G_\tau' \Sigma^{-1} G_\tau) \dot{B} \\ J_\alpha &= B' G_\tau' \Sigma^{-1} G_\tau B, & J_{\tau\beta} &= \dot{G}_\tau' (B\alpha\alpha' \otimes \Sigma^{-1} G_\tau) \dot{B} \\ J_{\tau\alpha} &= \dot{G}_\tau' (B\alpha \otimes \Sigma^{-1} G_\tau) B, & J_{\beta\alpha} &= \dot{B}' (\alpha \otimes G_\tau' \Sigma^{-1} G_\tau) B \end{aligned}$$

and

$$\begin{aligned} I_\tau &= \dot{G}_\tau' (B\alpha \otimes \Sigma^{-1}) R (\alpha B' \otimes \Sigma^{-1}) \dot{G}_\tau \\ I_\alpha &= B' G_\tau' \Sigma^{-1} R \Sigma^{-1} G_\tau B, \\ I_\beta &= \dot{B}' (\alpha \otimes G_\tau' \Sigma^{-1}) R (\alpha' \otimes \Sigma^{-1} G_\tau) \dot{B} \\ I_{\tau\beta} &= \dot{G}_\tau' (B\alpha \otimes \Sigma^{-1}) R (\alpha' \otimes \Sigma^{-1} G_\tau) \dot{B} \\ I_{\tau\alpha} &= \dot{G}_\tau' (B\alpha \otimes \Sigma^{-1}) R \Sigma^{-1} G_\tau B \\ I_{\beta\alpha} &= \dot{B}' (\alpha \otimes G_\tau \Sigma^{-1}) R \Sigma^{-1} G_\tau B. \end{aligned}$$

Two cases for the constraint matrix C_θ are given: EEG and MEG. For EEG the $d \times 7d$ matrix with the partial derivatives of the constraints for the reparameterization is

$$C_\theta = \begin{pmatrix} O & B' & O \end{pmatrix},$$

For MEG the additional constraint of orthogonality between location and orientation is introduced. This leads to the $2d \times 7d$ matrix

$$C_\theta = \begin{pmatrix} B' & U(\tau) & O \\ O & B' & O \end{pmatrix},$$

where $U(\tau) = (\text{diag}(1/\psi) I_d \otimes 1_3') \text{diag}(\tau)$, where ψ contains the eccentricities of d sources.

REFERENCES

[1] H.M. Huizenga and P.C.M. Molenaar, "Estimating and testing the sources of evoked potentials in the brain," *Multivariate Behavioral Research*, vol. 29, pp. 237–267, 1994.

[2] M. Hämäläinen, R. Hari, R. Ilmoniemi, J. Knuutila, and O.V. Lounasmaa, "Magnetoencephalography–theory, instrumentation, and applications to noninvasive studies of the working human brain," *Review of Modern Physics*, vol. 65, pp. 413–497, 1993.

[3] R.D. Pascual-Marqui and R. Biscay-Lirio, "Spatial resolution of neuronal generators based on EEG and MEG measurements," *International Journal of Neuroscience*, vol. 68, pp. 93–105, 1993.

[4] J.C. de Munck and M.J. Peters, "A fast method to compute the potential in the multisphere model," *IEEE Transactions on Biomedical Engineering*, vol. 40, pp. 1166–1174, 1993.

[5] M. Peters and J. de Munck, "On the forward and the inverse problem for EEG and MEG," in *Auditory evoked magnetic fields and potentials*, F. H. Grandori, M. Hoke, and G. L. Romani, Eds. Basel: Karger, 1989.

[6] D. Gutiérrez, A. Nehorai, and C.H. Muravchik, "Estimating brain conductivities and dipole source signals with EEG arrays," .

[7] J.C. de Munck, B.W. van Dijk, and H. Spekreijse, "Mathematical dipoles are adequate to describe realistic generators of human brain activity," *IEEE Transactions on Biomedical Engineering*, vol. 35, pp. 960–966, 1988.

[8] J.C. Mosher, R.M. Leahy, D.W. Shattuck, and S. Baillet, "MEG source imaging using multipolar expansions," in *IPMI99*, A. Kuba et al., Ed. Heidelberg: Springer-Verlag, 1999.

[9] S. Supek and C.J. Aine, "Simulation studies of multiple dipole neuromagnetic source localization: model order and limits of source resolution," *IEEE Transactions on Biomedical Engineering*, vol. 40, pp. 529–540, 1993.

[10] H.M. Huizenga, *The statistical approach to electromagnetic source localization in the brain*, Ph.D. thesis, Amsterdam: University of Amsterdam, 1995.

[11] J. Malmivuo, V. Suihko, and H. Eskola, "Sensitivity distributions of EEG and MEG measurements," *IEEE Transactions on Biomedical Engineering*, vol. 44, no. 3, pp. 196–208, 1997.

[12] H.M. Huizenga and P.C.M. Molenaar, "Equivalent source estimation of scalp potential fields contaminated by heteroscedastic and correlated noise," *Brain Topography*, vol. 8, pp. 13–33, 1995.

[13] T.R. Knösche, E.M. Berends, H.R.A. Jagers, and M.J. Peters, "Determining the number of independent sources of the EEG: A simulation study on information criteria," *Brain Topography*, vol. 11, pp. 111–124, 1998.

[14] C.-G. Bénar, R.N. Gunn, C. Grova, B. Champagne, and J. Gotman, "Statistical maps for EEG dipolar source localization," *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 3, pp. 401–413, 2005.

[15] J.C. Mosher, P.S. Lewis, and R.M. Leahy, "Multiple dipole modeling and localization from spatio-temporal meg data," *IEEE Transactions on Biomedical Engineering*, vol. 39, pp. 541–557, 1992.

[16] J.C. Mosher and R.M. Leahy, "Source localization using recursively applied and projected (rap) music," *IEEE Ttransactions on Signal Processing*, vol. 47, pp. 332–340, 1999.

[17] R. Grave de Peralta Menendez, M.M. Murray, C.M. Michel, R. Martuzzi, and S.L. Gonzalez Andino, "Electrical neuroimaging based on biophysical constraints," *NeuroImage*, vol. 21, pp. 527–539, 2004.

[18] A.M. Dale, A.K. Liu, B.R. Fischi, R.L. Buckner, J.W. Belliveau, J.D. Lewine, and E. Halgren, "Dynamic statistical parametric mapping: Combining fMRI and MEG for high resolution imaging of cortical activity," *Neuron*, vol. 26, pp. 55–67, 2000.

[19] A. Sabharwal and L. Potter, "Wald statistic for model order selection in superposition models," *IEEE Transactions on Signal Processing*, vol. 50, no. 4, pp. 956–965, 2002.

[20] C.-H.J. Ying, A. Sabharwal, and R.L. Moses, "A combined order selection and parameter estimation algorithm for undamped exponentials," *IEEE Transactions on Signal Processing*, vol. 48, no. 3, pp. 693–701, 2000.

[21] L.J. Waldorp, H.M. Huizenga, R.P.P.P. Grasman, B.E. Böcker, J.C. de Munck, and P.C.M. Molenaar, "Model selection in electromagnetic source analysis with an application to VEFs," *IEEE Transactions on Biomedical Engineering*, vol. 49, pp. 1121–1129, 2002.

[22] L.J. Waldorp, H.M. Huizenga, A. Nehorai, R.P.P.P Grasman, and P.C.M. Molenaar, "Model selection in spatio-temporal electromagnetic source analysis," *IEEE Transactions on Biomedical Engineering*, to appear March 2005.

[23] H. White, "Using least squares to approximate unknown regression functions," *International Economic Review*, vol. 21, no. 1, pp. 149–170, 1980.

[24] H. White, "Consequences and detection of misspecified and nonlinear regression models," *Journal of the American Statistical Association*, vol. 76, no. 374, pp. 419–433, 1981.

[25] J.C. Mosher, M.E. Spencer, R.M. Leahy, and P.S. Lewis, "Error bounds for EEG and MEG dipole source localization," *Electroencephalography and Clinical Neurophysiology*, vol. 86, pp. 303–321, 1993.

[26] P. Stoica, M. Viberg, and B. Otterston, "Instrumental variable approach to array processing in spatially correlated noise fields," *IEEE transactions on signal processing*, vol. 43, pp. 1187–1199, 1995.

[27] M. Viberg, P. Stoica, and B. Otterston, "Array processing in correlated noise fields based on instrumental variables and subspace fitting," *IEEE transactions on signal processing*, vol. 43, pp. 1187–1199, 1995.

[28] K. Mahata and T. Söderström, "Large sample properties of separable nonlinear least squares estimators," *IEEE Transactions on Signal Processing*, vol. 52, no. 6, pp. 1650–1658, 2004.

[29] R. Golden, "Making correct statistical inferences using the wrong probability model," *Journal of Mathematical Psychology*, vol. 39, pp. 3–20, 1995.

[30] H. White, "Maximum likelihood estimation of misspecified models," *Econometrica*, vol. 50, pp. 1–25, 1982.

[31] G.H. Golub and V. Pereyra, "The differentiation of pseudo-inverses and

nolinear least squares problems whose variables separate," *SIAM Journal of Numerical Analysis*, vol. 10, no. 2, pp. 413–432, 1973.
[32] G.H. Golub and V. Pereyra, "Separable nonlinear least squares: the variable projection method and its applications," *Inverse Problems*, vol. 19, no. 2, pp. 1–26, 2003.
[33] A.W. Van der Vaart, *Asymptotic Statistics*, New York: Cambridge University Press, 1998.
[34] J.C. De Munck, P.C.M. Vijn, and F.H. Lopes da Silva, "A random dipole model for spontaneous brain activity," *IEEE Transactions on Biomedical Engineering*, vol. 39, pp. 791–804, 1992.
[35] M. Scherg, "Fundementals of dipole source potential analysis," in *Auditory evoked magnetic fields and electric potentials*, F. Grandori, M. Hoke, and G.L. Romani, Eds., pp. 40–69. Basel: Karger, 1990.
[36] J.C. de Munck, "The estimation of time varying dipoles on the basis of evoked potentials," *Electroencephalography and Clinical Neurophysiology*, vol. 77, pp. 156–160, 1990.
[37] H.M. Huizenga, J.C. de Munck, L.J. Waldorp, and R.P.P.P. Grasman, "Spatiotemporal EEG/MEG source analysis based on a parametric noise covariance model," *IEEE Transactions on Biomedical Engineering*, vol. 49, pp. 533–539, 2002.
[38] G.A.F. Seber and C.J. Wild, *Nonlinear regression*, Toronto: John Wiley and Sons, 1989.
[39] R.J. Muirhead, *Aspects of multivariate statistical theory*, New York: John Wiley & Sons, 1982.
[40] E.F. Vonesh and V.M. Chinchilli, *Linear and nonlinear models for the analysis of repeated measurements*, New York: Marcel Dekker, 1997.
[41] T. Amemiya, *Advanced econometrics*, Oxford: Basil Blackwell, 1985.
[42] L.J. Waldorp, H.M. Huizenga, C.V. Dolan, and P.C.M. Molenaar, "Estimated generalized least squares electromagnetic source analysis based on a parametric noise covariance model," *IEEE Transactions on Biomedical Engineering*, vol. 48, pp. 737–741, 2001.
[43] G. Kauermann and R.J. Carroll, "A note on the efficiency of sandwich covariance extimation," *Journal of the American Statistical Association*, vol. 96, no. 456, pp. 1387–1396, 2001.
[44] J. R. Magnus and H. Neudecker, *Matrix differential calculus with applications in statistics and economics, revised edition*, Chichester: John Wiley & Sons, 1999.
[45] S.D. Silvey, "The Lagrangian multiplier test," *Annals of Mathematical Statistics*, vol. 30, pp. 389–407, 1959.
[46] R.F. Engle, "Wald, likelihood ratio, and Lagrange multiplier tests in econometircs," in *Handbook of Econometrics, Volume I*, Z. Gilliches and M.D. Intriligator, Eds., pp. 775–826. Amsterdam: Elsevier Science Publishers BV, 1984.
[47] E.L. Lehmann, *Testing statistical hypotheses*, New York: John Wiley & Sons, 1959.
[48] S. Gonçalvez, J.C. de Munck, J.P.A. Verbunt, R.M. Heethaar, and F.H. Lopez da Silva, "*In vivo* measurent of the brain and skull resistivities using an EIT-based and combined analysis of SEF/SEP data," *IEEE Transactions on Biomedical Engineering*, vol. 50, no. 9, pp. 1124–1128, 2003.
[49] Z. Zhang, "A fast method to compute surface potentials generated by dipoles within multilayer anisotropic spheres," *Physics in Medicine and Biology*, vol. 40, pp. 335–349, 1995.
[50] J.C. Mosher, R.M. Leahy, and P.S. Lewis, "EEG and MEG: Forward solutions for inverse methods," *IEEE Transactions on Biomedical Engineering*, vol. 46, no. 3, pp. 245–259, 1999.