Maximum Entropy Estimation of Neuroelectric Source Covariance Statistics

M. E. Pflieger
Source Signal Imaging, Inc., San Diego, CA, USA

Abstract—**A maximum unscaled entropy solution to the spatial covariance inverse problem is presented, and the theory is applied to task-related EEG in four frequency bands. Whereas the obtained second order solutions are of interest in their own right, this new method also may be used, in principle, to improve underlying source covariance and forward models that are used to obtain first order inverse solutions.**

Keywords—**EEG, MEG, source estimation, correlation matrix, maximum entropy covariance estimation**

I. Introduction

Distributed solutions to the bioelectromagnetic inverse problem for m electromagnetic (EEG, MEG, or combined EMEG) channels typically estimate a first order physical quantity, i.e., current density, at n oriented locations in the brain, where $n \gg m$. The "underdeterminacy" of the problem may be quantified as the log ratio of unknowns to knowns, i.e., $\log_{10}(n/m)$. For example, with $m \approx 200$ and $n \approx 10000$, the underdeterminacy is 1.7 orders of magnitude. Because second order quantities such as variance or power may be estimated using essentially the same inverse methods (e.g., by squaring the first order solution), the underdeterminacy of the problem is not thereby increased. However, it *doubles* (e.g., from 1.7 to 3.4 orders of magnitude) for the *full covariance* inverse problem, i.e., estimation of $n(n+1)/2$ source covariances given $m(m+1)/2$ measurement covariances. If a solution to the full second order problem is sought via first order methods — i.e., source covariance matrix $\mathbf{S} \approx \mathbf{HCH}^{\mathrm{T}}$, where $\mathbf{C}$ is a measurement covariance matrix and $\mathbf{j} \approx \mathbf{Hv}$ estimates current density vector $\mathbf{j}$ given measurement vector $\mathbf{v}$ — it is highly likely that errors in the first order solution will be magnified in the second order solution. In addition, first order solutions are designed to use only m measurements rather than $m(m+1)/2$. These considerations indicate the need for an intrinsic solution to the second order problem.

A deeper reason for seeking a second order solution that is independent of particular first order solutions is that the latter explicitly or implicitly make assumptions regarding second order statistics. For example, regional activity estimation (REGAE), an estimator of activity in brain regions of interest, explicitly assumes a Gaussian statistical process in the brain, characterized by a source covariance matrix ([1]). Other local estimators, such as linearly constrained minimum variance beamformers, may be classified, in part, by their source covariance assumptions ([2]). Turning to global estimators: As noted in [3], Bayesian estimators naturally incorporate the source covariance matrix $\mathbf{S}$ as prior information, and minimum ℓ_2-norm estimators may be construed as adopting the norm $\|\mathbf{j}\| = \sqrt{\mathbf{j}^{\mathrm{T}}\mathbf{S}^{-1}\mathbf{j}}$. Thus, a second order estimation procedure that stands on its own avoids circular assumptions, and might be used to improve subsequent first order estimation.

As noted, the full covariance inverse problem is doubly underdetermined. However, if the source correlation matrix is known, the underdeterminacy of the remaining problem of estimating source variances from measurement covariances is $\log_{10}(2n/(m(m+1)))$, which, for $m \approx 200$ and $n \approx 10000$, is -0.3 orders of magnitude. Thus, the covariance-to-variance problem, assuming known source correlations, can be overdetermined. Although this is not the full problem, it suggests that source correlation modeling may provide a viable solution strategy. The spatial correlation structure of neuroelectric activity in the brain is exceedingly complex, and detailed prior knowledge is deficient; but even a crude model, such as the following, may help. Given two brain regions, it is plausible to expect zero-lag correlations between them approximately to the extent that their inputs come from common brain regions. If neighboring regions tend to share inputs, this principle entails (at minimum) a local correlation structure that decays with some distance measure, such as cortical surface distance.

In [4], Sahani and Nagarajan describe a variational Bayesian approach to estimate the source correlation matrix, which introduces sparsity via hyperparameter optimization. Whereas sparse solutions have low entropy, the present paper describes an approach that, by contrast, aims for maximum entropy solutions, i.e., those that are maximally distributed throughout the brain and minimally constrained. The rationale is that a solution represents a spontaneous process in the brain, and it is desirable to retain as many statistical degrees of freedom as possible, consistent with our knowledge of data and models. Undoubtedly, there will be some situations in which sparse solutions are correct, and others in which distributed solutions are correct. If a source covariance matrix will be used for subsequent first order estimation — as in the case of REGAE — a maximum entropy solution may be desirable in either case.

II. Methods

A. Theory

The assumed relationship between n current dipole elements in the source domain and m measurement channels derived from $m' \geq m$ sensors is

$$\mathbf{v}(t) = \mathbf{DFj}(t), \tag{1}$$

where $\mathbf{j}(t)$ is an n-dimensional current density vector at time t; $\mathbf{F}$ is an $m' \times n$ forward matrix; $\mathbf{D}$ is an $m \times m'$ derivation matrix from sensors to channels; and $\mathbf{v}(t)$ is an m-

dimensional derived measurement vector at t. For EEG, $\mathbf{D}$ includes at minimum the referencing scheme; more generally, it may incorporate an arbitrary linear derivation. The $m \times n$ gain matrix is defined as $\mathbf{G} \equiv \mathbf{DF}$.

Assuming a zero-mean process, one typical way to estimate an $m \times m$ observed covariance matrix $\mathbf{C}$ from a set of derived measurements is

$$\mathbf{C} = \left\langle \mathbf{v}\mathbf{v}^{\mathrm{T}} \right\rangle \approx \frac{1}{p} \sum_{t=1}^{p} \mathbf{v}(t)\mathbf{v}(t)^{\mathrm{T}}, \tag{2}$$

where the number of samples p exceeds the number of derived channels several fold (e.g., $p \approx 4m$). Given $\mathbf{C}$, an $n \times n$ source covariance matrix $\mathbf{S}$ is desired such that

$$\mathbf{C} = \mathbf{GSG}^{\mathrm{T}}. \tag{3}$$

In the underdetermined case considered here, this constraint does not uniquely determine $\mathbf{S}$. However, given some prior expectations embodied by a matrix $\mathbf{\Omega}$, called a *source covariance model*, $\mathbf{S}$ may be uniquely specified as the source covariance satisfying (3) that in some sense is "closest" to the expected model. One direct approach to formalize this problem is to minimize the Frobenius norm [5] between $\mathbf{S}$ and $\mathbf{\Omega}$. Another is to minimize the Kullback-Leibler divergence [6] between the zero-mean Gaussian probability density functions corresponding to $\mathbf{S}$ and $\mathbf{\Omega}$.

A third approach, taken here, is to maximize the entropy of a *constrained spontaneity matrix* $\mathbf{\Psi}$ constructed so that

$$\mathbf{S} = \mathbf{\Omega}^{1/2}\mathbf{\Psi}\mathbf{\Omega}^{1/2} \tag{4}$$

satisfies (3). Consider first the ideal case of a source covariance model $\mathbf{\Omega}$ that accounts perfectly for the observed covariance $\mathbf{C}$ up to a scale factor. It follows that $\mathbf{\Psi} = \sigma^2\mathbf{I}_n$, which corresponds to the independent and identically distributed n-variate Gaussian distribution with standard deviation σ. This generator of random activity ("spontaneity") has maximum entropy in an *unscaled* sense, i.e., disregarding the scale σ. In real cases, the source covariance model insufficiently explains the data, so that $\mathbf{\Psi}$ must deviate from the scaled identity to satisfy (3). Nevertheless, according to the maximum entropy principle used here, the added constraints should be minimized so that $\mathbf{\Psi}$ approximates $\sigma^2\mathbf{I}_n$ as nearly as possible. Thus, if

$$\mathbf{\Psi}^{1/2} = \sigma\mathbf{I}_n + \mathbf{\Xi}, \tag{5}$$

then the scale of the constraints matrix $\mathbf{\Xi}$ should be minimized relative to σ. Propositions 1 and 2 provide a way to realize this approach.

Proposition 1. Let: $\mathbf{C}$ be a $m \times m$ symmetric positive definite matrix (e.g., a covariance matrix); $\mathbf{C}^{-1/2}$ be its inverse square root; and $\mathbf{B}$ be $m \times n$, with $\mathbf{B}^{\mathrm{T}}\mathbf{C}^{-1/2} = \mathbf{UWV}^{\mathrm{T}}$ a singular value decomposition (SVD; [5]) with all non-zero singular values. (The SVD form implies that $\mathbf{W}$ is non-negative diagonal, and that $\mathbf{U}^{\mathrm{T}}\mathbf{U} = \mathbf{V}^{\mathrm{T}}\mathbf{V} = \mathbf{VV}^{\mathrm{T}} = \mathbf{I}_m$; though in general $\mathbf{UU}^{\mathrm{T}} \neq \mathbf{I}_n$.) Then for any scalar σ

$$\mathbf{C} = \mathbf{B}(\sigma\mathbf{I}_n + \mathbf{U}(\mathbf{W}^{-1} - \sigma\mathbf{I}_m)\mathbf{U}^{\mathrm{T}})^2\mathbf{B}^{\mathrm{T}}. \tag{6}$$

Proof. See Appendix A.

In particular, if $\mathbf{B} = \mathbf{G}\mathbf{\Omega}^{1/2}$, then (3) is satisfied via (4) and (5) when

$$\mathbf{\Xi} = \mathbf{U}(\mathbf{W}^{-1} - \sigma\mathbf{I}_m)\mathbf{U}^{\mathrm{T}}. \tag{7}$$

Because the columns of $\mathbf{U}$ are orthonormal, $\mathbf{\Xi}$ incorporates m independent constraints, the ith constraint weighted as $|w_i^{-1} - \sigma|$, where $\mathrm{diag}(w_1 \cdots w_m) = \mathbf{W}$. Proposition 2 may be used to find σ that minimizes the contribution of $\mathbf{\Xi}$ relative to $\sigma\mathbf{I}_n$, thereby maximizing the unscaled entropy of $\mathbf{\Psi}$.

Proposition 2. Define

$$\phi(\sigma) \equiv \frac{1}{m\sigma^2} \sum_{i=1}^{m} (w_i^{-1} - \sigma)^2, \tag{8}$$

where σ and all $\{w_i \mid i = 1, m\}$ are positive. Then ϕ has a unique minimum at

$$\sigma_* = \left(\sum_{i=1}^{m} w_i^{-2} \right) \left(\sum_{i=1}^{m} w_i^{-1} \right)^{-1}, \tag{9}$$

and $0 \leq \phi(\sigma_*) < 1$.

Proof. See Appendix B.

Therefore, the maximum (unscaled) entropy solution is

$$\begin{aligned} \mathrm{svd}(\mathbf{\Omega}^{1/2}\mathbf{G}^{\mathrm{T}}\mathbf{C}^{-1/2}) &= \mathbf{UWV}^{\mathrm{T}} \\ \mathbf{\Psi}_*^{1/2} &= \sigma_*\mathbf{I}_n + \mathbf{U}(\mathbf{W}^{-1} - \sigma_*\mathbf{I}_m)\mathbf{U}^{\mathrm{T}}. \\ \mathbf{S}_* &= \mathbf{\Omega}^{1/2}\mathbf{\Psi}_*\mathbf{\Omega}^{1/2} \end{aligned} \tag{10}$$

Further, the quantity $\phi(\sigma_*)$ indicates the degree to which the source covariance model $\mathbf{\Omega}$ can explain the observed covariance $\mathbf{C}$. If $\mathbf{\Omega}$ is ideal, so that no added constraints are needed to satisfy (3), then $\phi(\sigma_*) = 0$. This condition is equivalent to $w_1 = w_2 = \cdots = w_m = w$, for in this case $\sigma_* = w^{-1}$. On the other hand, $\phi(\sigma_*) < 1$ is guaranteed for *any* source covariance model that satisfies the mild condition $\mathrm{rank}(\mathbf{\Omega}^{1/2}\mathbf{G}^{\mathrm{T}}\mathbf{C}^{-1/2}) = m$. Consequently, as $\phi(\sigma_*) \to 1$, the constraints term $\mathbf{\Xi}$ becomes balanced with the "spontaneity term" $\sigma\mathbf{I}_n$; and this worst case situation indicates that the source covariance model is not helping to satisfy (3). Note that the case $\sigma = 0$, which relies exclusively on $\mathbf{\Xi}$, produces a solution to (3), namely, $\mathbf{\Psi} = \mathbf{UW}^{-2}\mathbf{U}^{\mathrm{T}}$, which has *minimum* entropy. It therefore is never a candidate for the maximum entropy solution.

B. EEG Data

The theory was applied to a sample EEG dataset provided by Richard Clark and Kathryn Moores of Flinders University. Although the experimental details are of intrinsic interest, the aims of this paper are methodological, and so the description here is brief. A healthy, right-handed, male subject participated in a visual verbal working memory experiment, of which a low memory load condition ("fixed target") is analyzed here. A whole-head, T1-weighted structural MRI was obtained (Siemens VISION, 1.5 Tesla); 124-channel EEG was recorded continuously during the task (Neuroscan ESI-128, 400 Hz); and electrode locations were digitized (Polhemus Fastrak). Further details are in [7].

C. Analysis

EMSE Suite software (Source Signal Imaging, version 5.1) was used to: (i) segment the structural MRI into regions, including scalp, skull, CSF, white matter, and gray matter; (ii) produce 2D meshes for cortical, inner skull, outer skull, and scalp surfaces; (iii) co-register digitized electrode locations to the structural MRI; (iv) calculate 3-shell spherical and 3-compartment boundary element method (BEM) head models; (v) inspect, common average reference, bandpass filter (2-pole, zero-phase, Butterworth squared), and select the task-related EEG segments; and (vi) estimate measurement covariances in four frequency bands: 8-13 Hz (alpha), 14-20 Hz (beta-1), 21-32 Hz (beta-2), and 33-55 Hz (gamma). About 150 event-related epochs (0 s to 1 s post-stimulus) were concatenated for this purpose (for a total of about 60,000 time samples).

Spatial SVD was applied to each resulting covariance matrix, and a number of principal components was retained so as to account for 95% of the total variance (13 components for alpha, 21 for beta-1, 22 for beta-2, and 27 for gamma). The principal vectors were used to construct the derivation matrix **D** of equation (1). The forward matrix **F** was computed in four (2×2) ways: {head model = spherical or BEM} × {cortical mesh vertices = 16,027 or 24,904}. A dipole element was placed at each mesh vertex using a surface-normal orientation ([8]). The source covariance model **Ω** was assumed to be the identity matrix.

III. Results

Cortical maps for the diagonal (standard deviations) of the estimated maximum entropy source covariance matrix square root are shown in Fig. 1 for the BEM using the 24,904-vertex mesh. Values are mapped relative to σ_*.

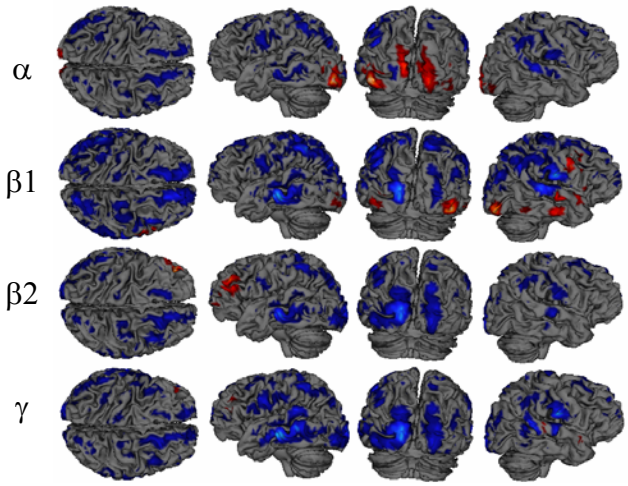

Fig. 1. Cortical maps of the diagonal of the square root of the estimated source covariance matrix for alpha (8-13 Hz), beta-1 (14-20 Hz), beta-2 (21-32 Hz), and gamma (33-55 Hz) bands. Gray corresponds to σ_*; red means "greater than σ_*" and blue means "less than σ_*". The source domain was a cortical mesh with 24,904 vertices; the source covariance model was the identity; the forward model was computed using a 3-compartment BEM; and data were reduced so as to conserve 95% of the variance.

Table 1 summarizes the $\phi_* \equiv \phi(\sigma_*)$ measure obtained for the four forward matrices: {Spherical head model vs. BEM head model} × {16,027 vertices vs. 24,904 vertices}.

Table 1	Spherical		BEM	
ϕ_*	16,027	24,904	16,027	24,904
8-13 Hz	*0.242*	*0.255*	*0.245*	*0.253*
14-20 Hz	*0.191*	*0.197*	*0.254*	*0.258*
21-32 Hz	*0.313*	*0.356*	*0.398*	*0.434*
33-55 Hz	*0.408*	*0.424*	*0.480*	*0.498*

IV. Discussion

A. Interpretation

As Fig. 1 illustrates, the maximum unscaled entropy method for estimating source standard deviations via measurement covariances produces plausible results. The values mapped are deviations from the constant σ_*. Positive deviations are those in excess of σ_*, and negative deviations are those below σ_*. Thus, σ_* is providing a kind of brain-wide "reference". According to the goal of minimizing imposed constraints, these deviations are the smallest necessary to reconstruct the observed covariances. Although the overall solution (with the addition of σ_*) has very high spatial uniformity, the deviations themselves apparently are structured. For the alpha band, positive deviations are near V1 occipital areas, and in lateral occipital areas (mostly on the left); and there are some bilateral negative deviations (e.g., superior frontal). The beta-1 band shows enhanced lateral occipital positive deviations that coexist with negative V1 deviations, hinting that the underlying dynamics differ; also, there are some notable right-left asymmetries. For the beta-2 band, all is negative in the occipital regions; and a left-frontal positive deviation appears: possibly near Broca's area (recalling the verbal working memory task). The gamma band shows a hint of the same left-frontal region.

The results of Table 1 were initially surprising: Considering that ϕ_* is a goodness-of-fit measure for $\mathbf{\Omega}^{1/2}\mathbf{F}^{\mathrm{T}}$, it seems natural to expect that improvement of the forward matrix **F** via a refined source domain and a more realistic volume conductor model should diminish ϕ_*; but the opposite was observed. An hypothesis to explain this is that (a) the source covariance model $\mathbf{\Omega} = \mathbf{I}$ is too simple—in particular, there are local correlations of brain activity—and (b) the simplified source domain and volume conductor models provided extra smoothing to compensate somewhat for this deficiency. For the less-refined mesh, each dipole element represents a larger cortical surface area: a kind of local spatial averaging. Likewise, the spherical head model "blurs" the forward solution, thus providing a different form of smoothing. If true, then the following theoretical extension should provide a way to address the problem.

B. Extension of Theory

Considering that $\phi_* \equiv \phi(\sigma_*)$ measures the "goodness of fit" of the source covariance model to the data, a *parameterized* model $\mathbf{\Omega}[\boldsymbol{\omega}]$ may be optimized by finding a parameter vector $\boldsymbol{\omega}$ that minimizes ϕ_*. In general, a covariance matrix $\mathbf{\Omega}$ may be factored as

$$\mathbf{\Omega} = \mathbf{A}\mathbf{\Gamma}\mathbf{A}, \quad (11)$$

where $\mathbf{A}$ is a diagonal amplitude (standard deviation) matrix, and $\mathbf{\Gamma}$ is a correlation matrix ($-1 \le \Gamma_{ij} = \Gamma_{ji} \le 1$, and $\Gamma_{ii} = 1$). Thus, the model parameters may be classified as those which apply to the amplitude matrix ($\mathbf{A}[\boldsymbol{\alpha}]$) and those which apply to the correlation matrix ($\mathbf{\Gamma}[\boldsymbol{\gamma}]$).

For example, a simple one-parameter covariance model may be implemented that assumes a spatially flat amplitude distribution; i.e., $\mathbf{A} = \mathbf{I}_n$. In accordance with the "common input principle" suggested in the Introduction, local spatial correlations in cortex are plausible. To formalize, if i and j are dipole elements within the same hemisphere, then

$$\Gamma_{ij} = \exp\{-l_{ij} / \lambda\}, \quad (12)$$

where l_{ij} is the cortical surface distance between i and j, and λ is a local coupling length parameter. This primitive model neglects many other kinds of correlations, such as those between homologous inter-hemispheric cortical areas.

In summary, given $\mathbf{C}$ and $\mathbf{G}$, an extended maximum entropy algorithm iterates over various values of λ in search of an optimal λ_* such that, using the source covariance model $\mathbf{\Omega} = \mathbf{\Gamma}[\lambda_*]$, $\phi(\sigma_*)$ is minimized. The results are $\mathbf{S}_*$ (the maximum entropy source covariance estimate), σ_* (standard deviation of spontaneous activity), λ_* (local coupling length), and ϕ_* (a goodness-of-fit measure).

V. Conclusion

A novel solution to the second-order inverse problem of estimating source variances and covariances from sensor covariances has been presented. In principle, estimation of $\mathbf{S}$ may be used to improve first-order inverse solutions. The ϕ_* measure may help to improve underlying parameterized models, in particular, source covariance models. This work remains to be compared with that of [9].

Appendix

A. Proof of Proposition 1

The plan is to reduce the proposed identity to $\mathbf{I}_m = \mathbf{I}_m$ using SVD properties $\mathbf{U}^{\mathrm{T}}\mathbf{U} = \mathbf{I}_m$ and $\mathbf{V}^{\mathrm{T}}\mathbf{V} = \mathbf{I}_m$.

Because $\mathbf{C}$ is nonsingular, (6) is equivalent to

$$\mathbf{I}_m = \mathbf{C}^{-1/2}\mathbf{B}(\sigma\mathbf{I}_n + \mathbf{U}(\mathbf{W}^{-1} - \sigma\mathbf{I}_m)\mathbf{U}^{\mathrm{T}})^2\mathbf{B}^{\mathrm{T}}\mathbf{C}^{-1/2}. \quad (A1)$$

Symbol manipulation produces the following identity:

$$(\sigma\mathbf{I}_n + \mathbf{U}(\mathbf{W}^{-1} - \sigma\mathbf{I}_m)\mathbf{U}^{\mathrm{T}})^2 = \sigma^2\mathbf{I}_n + \mathbf{U}(\mathbf{W}^{-2} - \sigma^2\mathbf{I}_m)\mathbf{U}^{\mathrm{T}}.$$

Substituting $\mathbf{B}^{\mathrm{T}}\mathbf{C}^{-1/2} = \mathbf{U}\mathbf{W}\mathbf{V}^{\mathrm{T}}$, the r.h.s. of (A1) becomes

$$\mathbf{V}\mathbf{W}\mathbf{U}^{\mathrm{T}}(\sigma^2\mathbf{I}_n + \mathbf{U}(\mathbf{W}^{-2} - \sigma^2\mathbf{I}_m)\mathbf{U}^{\mathrm{T}})\mathbf{U}\mathbf{W}\mathbf{V}^{\mathrm{T}}.$$

Left-multiplying by $\mathbf{V}^{\mathrm{T}}$ and right-multiplying by $\mathbf{V}$ leaves the l.h.s. is intact, and the r.h.s. reduces to

$$\mathbf{W}(\sigma^2\mathbf{U}^{\mathrm{T}}\mathbf{U} + \mathbf{U}^{\mathrm{T}}\mathbf{U}(\mathbf{W}^{-2} - \sigma^2\mathbf{I}_m)\mathbf{U}^{\mathrm{T}}\mathbf{U})\mathbf{W} = \mathbf{W}\mathbf{W}^{-2}\mathbf{W} = \mathbf{I}_m.$$

B. Proof of Proposition 2

Define $a \equiv m^{-1}\sum_{i=1}^{m} w_i^{-2}$ and $b \equiv m^{-1}\sum_{i=1}^{m} w_i^{-1}$. From (8),

$$\phi(\sigma) = a\sigma^{-2} - 2b\sigma^{-1} + 1$$
$$\phi'(\sigma) = -2\sigma^{-2}(a\sigma^{-1} - b)$$
$$\phi''(\sigma) = 2\sigma^{-3}(3a\sigma^{-1} - 2b)$$

$\phi'(\sigma_*) = 0$ implies that $\sigma_* = ab^{-1}$, per (9). $\phi''(ab^{-1}) = 2a^{-3}b^4$, which is positive because both a and b are positive, considering that all singular values $\{w_i\}$ are positive. Therefore, σ_* corresponds to a unique minimum. As $\sigma \to 0$, $\phi \to \infty$. As $\sigma \to \infty$, $\phi \to 1$ from below, because $\phi(\sigma_*/2) = 1$ and $\phi(\sigma_*)$ is the global minimum. Therefore, $\phi(\sigma_*) < 1$. By construction, ϕ is non-negative; and, as noted, it can be 0 when all singular values are equal.

Acknowledgment

This work is supported by the NIMH (R44 MH064343). Thanks to Richard Clark and Kath Moores for the sample EEG dataset, and to Jandro Kirkish for helpful comments.

References

[1] M.E. Pflieger, R.E. Greenblatt, J. Kirkish, "Regional resolving power of combined MEG/EEG," *Neurol. Clin. Neurophysiol.*, in press.

[2] R.E. Greenblatt, A. Ossadtchi, M.E. Pflieger, "Local linear estimators for the bioelectromagnetic inverse problem," *IEEE Trans Signal Proc.*, in press.

[3] C. Phillips, M.D. Rugg, K.J. Friston, "Anatomically informed basis functions for EEG source localization: combining functional and anatomical constraints," *NeuroImage*, vol. 16, pp. 678-695, 2002.

[4] M. Sahani, S.S. Nagarajan, "Reconstructing MEG sources with unknown correlations," in *Advances in Neural Information Processing Systems*, vol. 16, S. Thrun, L.K. Saul, B. Schölkopf, Eds. Cambridge, MA: MIT Press, Jun. 2004.

[5] G.H. Golub, C.F. Van Loan, *Matrix Computations, Third Edition*. Baltimore: The Johns Hopkins University Press, 1996, p. 55; pp. 70-73.

[6] S. Kullback, *Information Theory and Statistics*. Mineola, NY: Dover Publications, 1997 (first published 1959), pp. 1-7, 189-190.

[7] C.R. Clark, K.A. Moores, A. Lewis, D.L. Weber, S. Fitzgibbon, R. Greenblatt, G. Brown, J. Taylor, "Cortical network dynamics during verbal working memory function," *Int. J. Psychophysiol.*, vol. 42, pp. 161-176.

[8] A.M. Dale, M.I. Sereno, "Improved localization of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: A linear approach," *J. Cogn. Neurosci.*, vol. 5, pp. 162-176.

[9] P. Patel, D. Khosla, L. Al-Dayeh, M. Singh, "Distributed source imaging of alpha activity using a maximum entropy principle," *Clin. Neurophysiol.*, vol. 110, no. 3, pp. 538-549, Mar. 1999.