2003 IEEE Workshop on Statistical Signal Processing, Sep 28 - Oct 1, 2003, St. Louis, Missouri, pp. 294-297.

EQUIVALENCE OF LINEAR APPROACHES IN BIOELECTROMAGNETIC INVERSE SOLUTIONS

John C. Mosher, Sylvain Baillet, Richard M. Leahy

Los Alamos National Laboratory, Los Alamos, NM USA;
Cognitive Neuroscience & Brain Imaging Laboratory, CNRS-UPR 640, Paris, France;
Signal and Image Processing Institute, University of Southern California, Los Angeles, CA USA

ABSTRACT

Several different techniques have appeared in the magnetoencephalography (MEG) literature for solving the ambiguous inverse problem. We show that if the same second order statistics are uniformly applied in each case, then the resulting solutions are the same. Additionally, for the common assumption of independent sources, then the "noise significance" measure emphasises the need to select the data and noise covariance statistics, not the specific values of source statistics. Thus such algorithms as "dynamic SPM" and "synthetic aperture magnetoencephalography" are theoretically equivalent and differ in practice by practical matters of specification or estimation of data covariances.

1. THE FORWARD MODEL

The reader is referred to [1], [4], [7] as examples of reviews of magnetoencephalographic terminology and processing. We define the terminology used here,

$$\boldsymbol{d} = \boldsymbol{L}\boldsymbol{j} + \boldsymbol{v} \text{ or equivalently } \boldsymbol{d}_i = \boldsymbol{L}\boldsymbol{j}_i + \boldsymbol{v}_i \qquad (1)$$

where: $\boldsymbol{d}$, $m \times 1$, the data, measured at m external channels for 1 time sample; $\boldsymbol{L}$, $m \times p$, the *lead-field model*, relating p cortically constrained and oriented current dipoles to m external measurements; $\boldsymbol{j}$, $p \times 1$, the signed amplitudes of the current dipoles; $\boldsymbol{v}$, $m \times 1$, represents additive "noise" (anything not explicitly in the model $\boldsymbol{L}\boldsymbol{j}$).

2. THE ASSUMPTIONS HERE

The lead-field model $\boldsymbol{L}$ is known. We assume that given a current dipole's location and moment inside the brain, we can accurately calculate the corresponding measurement, including the volume currents driven by the dipole (e.g. [9]). This assumption ignores registration problems, inaccuracies in surface extractions, errors in forward model calculations, and errors in acquisition.

The forward model $\boldsymbol{L}$ is the same for all techniques shown here, i.e. the same set of cortical dipoles are used. In practice, researchers often use markedly different grids of dipoles.

We specify or estimate the first and second moments of $\boldsymbol{v}_i$ identically for all techniques, such that $E(\boldsymbol{v}_i) \equiv \boldsymbol{0}$ and $E(\boldsymbol{v}_i \boldsymbol{v}_i^T) \equiv \boldsymbol{C}_{\boldsymbol{v}}$. In practice, dramatically different techniques are used to estimate or specify the noise statistics, if at all.

We will specify or estimate either $E(\boldsymbol{j}_i \boldsymbol{j}_i^T) \equiv \boldsymbol{C}_{\boldsymbol{j}}$ or $E(\boldsymbol{d}_i \boldsymbol{d}_i^T) \equiv \boldsymbol{C}_{\boldsymbol{d}}$, and both are assumed zero-mean for convenience here. The major difference between techniques will be differences in these priors.

We will assume independence between noise and currents, linking the three covariances as

$$\boldsymbol{C}_{\boldsymbol{d}} = (\boldsymbol{L}\boldsymbol{C}_{\boldsymbol{j}}\boldsymbol{L}^T + \boldsymbol{C}_{\boldsymbol{v}}). \qquad (2)$$

3. LEAD-FIELD SURROGATES

The lead field matrix $\boldsymbol{L}$ is assumed here to be a dense (5,000 - 500,000) mesh of elemental cortical dipoles. The sensor array is about 100-300 sensors, and $\boldsymbol{L}$ therefore represents a system of highly underconstrained equations. Various techniques exist to combine regions of interest into much smaller models, such as equivalent cortical dipoles (ECDs) or multipoles ([5]), or simply subset selections of $\boldsymbol{L}$. We therefore can substitute surrogate models for the lead-field, and our general linear model is

$$\boldsymbol{d} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{v}, \qquad (3)$$

where $\boldsymbol{x}$ is the corresponding abstract surrogate for $\boldsymbol{j}$ (e.g. ECD moment). We note that $\boldsymbol{A}$ may represent over or underconstrained equations, while we will always assume that $\boldsymbol{L}$ is underconstrained and $\boldsymbol{j}$ represents the moments of a dense grid of cortical dipoles.

4. THE APPROACHES

4.1 Minimum Norm Approach

We concatenate model and noise,

$$\boldsymbol{d} = [\boldsymbol{A}, \boldsymbol{I}] \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{v} \end{bmatrix} \equiv \boldsymbol{B}\boldsymbol{y}. \qquad (4)$$

Since $\boldsymbol{B}$ is always "wide" by design (underconstrained), then no single unique solution for $\boldsymbol{y}$ exists. To limit the solution uniquely, we constrain $\boldsymbol{y}$ where $\boldsymbol{C}$ is specified and $\boldsymbol{c}$ is arbitrary,

$$\boldsymbol{y} = \boldsymbol{C}\boldsymbol{B}^T\boldsymbol{c}, \; \boldsymbol{d} = \boldsymbol{B}\boldsymbol{C}\boldsymbol{B}^T\boldsymbol{c}. \qquad (5)$$

The matrix $\boldsymbol{B}\boldsymbol{C}\boldsymbol{B}^T$ is now square and invertible, and we can estimate $\boldsymbol{c}$ uniquely from the data,

This work was supported in part by the National Institute for Biomedical Imaging and Bioengineering under grant R01-MH53213, and by Los Alamos National Laboratory, operated by the University of California for the United States Department of Energy, under Contract W-7405-ENG-36.

Generalized Linear Solution with Priors

$$\hat{\boldsymbol{j}} = \boldsymbol{C_j}\boldsymbol{L}^T\boldsymbol{C_d}^{-1}\boldsymbol{d} = \boldsymbol{C_j}\boldsymbol{L}^T(\boldsymbol{L}\boldsymbol{C_j}\boldsymbol{L}^T + \boldsymbol{C_v})^{-1}\boldsymbol{d} = (\boldsymbol{L}^T\boldsymbol{C_v}^{-1}\boldsymbol{L} + \boldsymbol{C_j}^{-1})^{-1}\boldsymbol{L}^T\boldsymbol{C_v}^{-1}\boldsymbol{d} \quad (A)$$

if $\boldsymbol{C_j} = \mathrm{diag}(\sigma_i^2)$, *then each dipole amplitude is* $\hat{j}_i = \sigma_i^2\boldsymbol{a}_i^T\boldsymbol{C_d}^{-1}\boldsymbol{d}$, *where* $\boldsymbol{L} \equiv [\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_p]$. *(B)*

Noise Significance for Diagonal Source Covariance

$$E\{\hat{j}_i^2\} = \sigma_i^4\boldsymbol{a}_i^T\boldsymbol{C_d}^{-1}\boldsymbol{a}_i, \text{ if } \boldsymbol{d} \equiv \boldsymbol{v} \text{ (no signal), then } E\{\hat{j}_i^2\big|_{\boldsymbol{d}=\boldsymbol{v}}\} = \sigma_i^4\boldsymbol{a}_i^T\boldsymbol{C_d}^{-1}\boldsymbol{C_v}\boldsymbol{C_d}^{-1}\boldsymbol{a}_i, \quad (C)$$

and "z-score" is $\hat{j}_i/(\hat{j}_i^2\big|_{\boldsymbol{d}=\boldsymbol{v}})^{1/2} = \boldsymbol{a}_i^T\boldsymbol{C_d}^{-1}\boldsymbol{d}/(\boldsymbol{a}_i^T\boldsymbol{C_d}^{-1}\boldsymbol{C_v}\boldsymbol{C_d}^{-1}\boldsymbol{a}_i)^{1/2}$ *(note no explicit* σ_i^2 *dependency).* *(D)*

$$\hat{\boldsymbol{c}} = (\boldsymbol{B}\boldsymbol{C}\boldsymbol{B}^T)^{-1}\boldsymbol{d}, \; \hat{\boldsymbol{y}} = \boldsymbol{C}\boldsymbol{B}^T(\boldsymbol{B}\boldsymbol{C}\boldsymbol{B}^T)^{-1}\boldsymbol{d}, \quad (6)$$

which for $\boldsymbol{C} = \boldsymbol{I}$ ($\boldsymbol{C} \neq \boldsymbol{I}$) is the (weighted) "minimum norm" solution because of all solutions satisfying *(4)*, this one has the smallest L_2 norm (in the space weighted by $\boldsymbol{C}$).

Assuming independence (cf. *(2)*), the expected covariance of $\boldsymbol{y}$ is

$$\boldsymbol{C_y} = \begin{bmatrix} \boldsymbol{C_x} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{C_v} \end{bmatrix}. \quad (7)$$

For statistical reasons discussed below, we let $\boldsymbol{C} = \boldsymbol{C_y}$, and expanding $\boldsymbol{B}$ and $\boldsymbol{y}$ yields

$$\begin{bmatrix} \hat{\boldsymbol{x}} \\ \hat{\boldsymbol{v}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{C_x}\boldsymbol{A}^T \\ \boldsymbol{C_v} \end{bmatrix}(\boldsymbol{A}\boldsymbol{C_x}\boldsymbol{A}^T + \boldsymbol{C_v})^{-1}\boldsymbol{d} \quad (8)$$

and finally *(A)* for the specific case of cortical dipoles.

4.2 Least-Squares Approach

If $\boldsymbol{A}$ is "tall," then in general the equations are overconstrained and no solution exists. The classical approach is to form the *normal equations* that minimize the norm of the error. The generalized (or weighted) least squares solution accounts explicitly for differences in noise variances by forming $\boldsymbol{C_v} = \boldsymbol{W_v}\boldsymbol{W_v}^T$, then "prewhitening" the equations to yield

$$(\boldsymbol{W_v}^{-1}\boldsymbol{d}) = (\boldsymbol{W_v}^{-1}\boldsymbol{A})\boldsymbol{x} + (\boldsymbol{W_v}^{-1}\boldsymbol{v}) \quad (9)$$

Multiplying both sides by $(\boldsymbol{W_v}^{-1}\boldsymbol{A})^T$ yields the normal equations

$$(\boldsymbol{A}^T\boldsymbol{C_v}^{-1})\boldsymbol{d} = (\boldsymbol{A}^T\boldsymbol{C_v}^{-1}\boldsymbol{A})\boldsymbol{x} + (\boldsymbol{A}^T\boldsymbol{C_v}^{-1})\boldsymbol{v} \quad (10)$$

and now $(\boldsymbol{A}^T\boldsymbol{C_v}^{-1}\boldsymbol{A})$ is square and invertible (but only if $\boldsymbol{A}$ is tall). The well-known "generalized least-squares" solution is

$$\hat{\boldsymbol{x}} = (\boldsymbol{A}^T\boldsymbol{C_v}^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{C_v}^{-1}\boldsymbol{d}. \quad (11)$$

We can easily include our prior on $\boldsymbol{x}$. We augment the equations and redefine as

$$\begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{d} \end{bmatrix} = \begin{bmatrix} \boldsymbol{I} \\ \boldsymbol{A} \end{bmatrix}\boldsymbol{x} + \begin{bmatrix} -\boldsymbol{x} \\ \boldsymbol{v} \end{bmatrix} \quad (12)$$

$$\delta = \boldsymbol{B}\boldsymbol{x} + \upsilon \quad (13)$$

By design, $\boldsymbol{B}$ is always tall, regardless of the size of $\boldsymbol{A}$, and we can proceed as before. We denote the covariance of our augmented noise vector υ as $\boldsymbol{C}_\upsilon$, yielding

$$\hat{\boldsymbol{x}} = (\boldsymbol{B}^T\boldsymbol{C}_\upsilon^{-1}\boldsymbol{B})^{-1}\boldsymbol{B}^T\boldsymbol{C}_\upsilon^{-1}\delta. \quad (14)$$

Assuming independence between noise and signal (cf. *(2)*) yields the prior on this augmented noise as $\boldsymbol{C}_\upsilon = \boldsymbol{C_y}$ from *(7)*, yielding after simple substitutions

$$\hat{\boldsymbol{x}} = (\boldsymbol{A}^T\boldsymbol{C_v}^{-1}\boldsymbol{A} + \boldsymbol{C_x}^{-1})^{-1}\boldsymbol{A}^T\boldsymbol{C_v}^{-1}\boldsymbol{d} \quad (15)$$

and identically *(A)*.

Note that *(11)* and *(15)* differ only in the inclusion of the source prior $\boldsymbol{C_x}$, such that *(15)* is also known as "generalized least-squares with prior information" ([11]). Note that if $\boldsymbol{A}$ is tall, we can let the source variance "go to infinity" (i.e. essentially unknown), such that $\boldsymbol{C_x}^{-1} \to \boldsymbol{0}$, and *(15)* becomes *(11)*. If $\boldsymbol{A}$ is wide, then the matrix in *(11)* cannot be inverted, i.e. prior information is necessary in order to form *(15)*.

Finally, note that $\hat{\boldsymbol{x}}$ in *(8)* and *(15)* are *identical*, proven by applying the *matrix inversion lemma* (e.g. [11]) to equate them, and hence the equality in *(A)*.

4.3 Gaussian Priors: Max Likelihood and MAP

We now validate the above approaches in a more formal approach, as we first presented in [6]. Consider the probability density function $p(\boldsymbol{d}|\boldsymbol{x})$, also known as the *likelihood function*, since we state the probability of the measured data as a function of the parameters. Since many pdfs are exponential, we also find it convenient to take the log of this pdf to obtain the *log likelihood function* $\ln p(\boldsymbol{d}|\boldsymbol{x})$. For statistical reasons (cf. Sorenson) we want those values of $\boldsymbol{x}$ that maximize this function, found by solving the *log likelihood equation*,

$$\frac{\partial}{\partial \boldsymbol{x}}\ln p(\boldsymbol{d}|\boldsymbol{x}) = \boldsymbol{0}. \quad (16)$$

If we assume that $\boldsymbol{v}$ is independent of $\boldsymbol{x}$, then ([11] p. 209)

$$p(\boldsymbol{d}|\boldsymbol{x}) \equiv p_{\boldsymbol{v}}((\boldsymbol{d} - \boldsymbol{A}\boldsymbol{x})|\boldsymbol{x}) \quad (17)$$

$$\equiv p(\boldsymbol{v}|\boldsymbol{x}) = p(\boldsymbol{v}) \quad (18)$$

If we further assume that $\boldsymbol{v}$ is zero-mean Gaussian with covariance $\boldsymbol{C_v}$, then the *maximum likelihood* (ML) solution is identically *(11)*.

Next, we may also consider $\boldsymbol{x}$ to also be random. We compare the *a priori* density function $p(\boldsymbol{x})$ to its *a posteri-*

ori probability density function $p(\boldsymbol{x}|\boldsymbol{d})$, i.e. the change in the density function given the data. *Bayes rule* allows us to easily relate the joint and conditional probability density functions as

$$p(\boldsymbol{x}|\boldsymbol{d})p(\boldsymbol{d}) = p(\boldsymbol{d}|\boldsymbol{x})p(\boldsymbol{x}) \quad (19)$$

from which we readily see the relationship between prior and posterior pdfs as

$$p(\boldsymbol{x}|\boldsymbol{d}) = (p(\boldsymbol{d}|\boldsymbol{x})p(\boldsymbol{x}))/p(\boldsymbol{d}). \quad (20)$$

A second common estimator is to therefore find the parameters that maximize this alternative likelihood function $p(\boldsymbol{x}|\boldsymbol{d})$. The denominator is not dependent on $\boldsymbol{x}$ and only serves to normalize the pdf, so the common alternative is to drop the denominator and effectively maximize instead the joint distribution $p(\boldsymbol{d}, \boldsymbol{x}) = p(\boldsymbol{d}|\boldsymbol{x})p(\boldsymbol{x})$. Following the same approach as above, we set the partial w.r.t. the parameters of the log of the pdf equal to zero. The likelihood equation becomes equivalent to ([11] p. 201)

$$\frac{\partial}{\partial \boldsymbol{x}} \ln p(\boldsymbol{d}|\boldsymbol{x}) + \frac{\partial}{\partial \boldsymbol{x}} \ln p(\boldsymbol{x}) = \boldsymbol{0}. \quad (21)$$

Estimators which maximize the posterior pdf are referred to as the *maximum a posteriori* (MAP).

If we assume the joint pdf for the p parameters in $\boldsymbol{x}$ and the m_T parameters in $\boldsymbol{v}$ to be Gaussian $[\boldsymbol{x}^T, \boldsymbol{v}^T]^T \in N(\boldsymbol{0}, \boldsymbol{C})$, and we make the common assumption that $\boldsymbol{x}$ and $\boldsymbol{v}$ are independent (cf. *(2)*), then $\boldsymbol{C}$ is shown in *(7)*. The MAP solution is then identically *(15)* and *(A)*.

4.4 Non-Gaussian Prior: Linear MMS

We first propose as a *loss function* or *performance function* the widely-known mean-square error,

$$J^2_{\mathrm{MS}} = E\{(\boldsymbol{x} - \hat{\boldsymbol{x}})^T(\boldsymbol{x} - \hat{\boldsymbol{x}})\}. \quad (22)$$

The estimator that minimizes this loss function is the *minimum mean-square estimator* (MMS)

$$\hat{\boldsymbol{x}}_{\mathrm{MS}} = E\{\boldsymbol{x}|\boldsymbol{d}\}, \quad (23)$$

i.e., the estimator that equals the expected value of $\boldsymbol{x}$, given the data. In general, this estimate is more difficult to formulate than the MAP.

For the case of our Gaussian assumptions above, however, the solution is well-known, identically *(15)*, and therefore we see that the MAP for Gaussian assumptions also achieves the optimal mean-square solution (cf. [11] p. 149).

In this framework, we can further relax our assumptions about the pdf. We now only assume the following statistics,

$$p(\boldsymbol{x}, \boldsymbol{v}) = p(\boldsymbol{x})p(\boldsymbol{v}), \quad (24)$$

$$E\{\boldsymbol{x}\} = E\{\boldsymbol{v}\} = \boldsymbol{0}, \quad (25)$$

$$E\{\boldsymbol{x}\boldsymbol{x}^T\} = \boldsymbol{C}_{\boldsymbol{x}},\ E\{\boldsymbol{v}\boldsymbol{v}^T\} = \boldsymbol{C}_{\boldsymbol{v}}, \quad (26)$$

in other words, we know only the first two moments of our pdf, and we again assume independence between the noise and the parameters of interest.

We make the important constraint that the estimator *must be linear* in the data, $\hat{\boldsymbol{x}} = \boldsymbol{W}^T\boldsymbol{d}$, a reasonable constraint in our case, since our data are linear in the parameters. Under these assumptions, then the linear mean square estimator (LMMS) is again the same as the MMS for the Gaussian priors above, yielding identically *(15)* and *(A)* ([11] p. 152).

4.5 Weighted Subspace Fitting

The above approaches specify/estimate $\boldsymbol{C}_{\boldsymbol{x}}$, and by implication therefore set $\boldsymbol{C}_{\boldsymbol{d}}$ through *(2)*. Alternative methods instead directly specify or estimate $\boldsymbol{C}_{\boldsymbol{d}}$. If $\boldsymbol{A}$ is tall, then we can exploit *(2)* to yield

$$\boldsymbol{C}_{\boldsymbol{x}} = \boldsymbol{A}^{\dagger}(\boldsymbol{C}_{\boldsymbol{d}} - \boldsymbol{C}_{\boldsymbol{v}})\boldsymbol{A}^{\dagger T} \quad (27)$$

and $\hat{\boldsymbol{x}}$ in *(8)* specializes to

$$\hat{\boldsymbol{x}} = \boldsymbol{A}^{\dagger}(\boldsymbol{C}_{\boldsymbol{d}} - \boldsymbol{C}_{\boldsymbol{v}})\Pi_{\boldsymbol{A}}\boldsymbol{C}_{\boldsymbol{d}}^{-1}\boldsymbol{d} \quad (28)$$

where $\Pi_{\boldsymbol{A}} \equiv \boldsymbol{A}\boldsymbol{A}^{\dagger}$ is the idempotent subspace projection operator, and this form of the estimate is referred to as *weighted subspace fitting*. This approach has seen little use in MEG [8].

4.6 Imaging

Looking at the form of *(15)* and expanding the data model $\boldsymbol{d} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{v}$ yields

$$\hat{\boldsymbol{x}} = \boldsymbol{R}\boldsymbol{x} + (\boldsymbol{A}^T\boldsymbol{C}_{\boldsymbol{v}}^{-1}\boldsymbol{A} + \boldsymbol{C}_{\boldsymbol{x}}^{-1})^{-1}(\boldsymbol{A}^T\boldsymbol{C}_{\boldsymbol{v}}^{-1})\boldsymbol{v} \quad (29)$$

where

$$\boldsymbol{R} \equiv (\boldsymbol{A}^T\boldsymbol{C}_{\boldsymbol{v}}^{-1}\boldsymbol{A} + \boldsymbol{C}_{\boldsymbol{x}}^{-1})^{-1}(\boldsymbol{A}^T\boldsymbol{C}_{\boldsymbol{v}}^{-1}\boldsymbol{A}). \quad (30)$$

Each column of the *resolution kernel* $\boldsymbol{R}$ is the *point spread function* for an element in $\boldsymbol{x}$. If $\boldsymbol{A}$ is tall, we can let $\boldsymbol{C}_{\boldsymbol{x}}^{-1} \to \boldsymbol{0}$, then $\boldsymbol{R}$ becomes the identity matrix and we achieve "perfect resolution," cf. *(11)*. If $\boldsymbol{A}$ is wide, we must specify a source prior, limiting the resolution (cf. [3]).

4.7 Linearly Constrained Minimum Variance

A common assumption is that each source is independent of all others. The result is that $\boldsymbol{C}_{\boldsymbol{x}}$ is diagonal, and the solution for each element is shown in *(B)*. Most published applications of *(A)* to dense cortical sources use a diagonal (e.g. [2]) or a highly sparse covariance matrix.

We combine the above ideas by first restricting all sources to be independent, yielding the dipole solution in *(B)*. Again inserting *(1)* yields the expression for a particular element

$$\hat{x}_i = \sigma_i^2 \boldsymbol{a}_i^T\boldsymbol{C}_{\boldsymbol{d}}^{-1}(\boldsymbol{a}_i x_i + \overline{\boldsymbol{A}_i\boldsymbol{x}_i} + \boldsymbol{v}) \quad (31)$$

where $\overline{\boldsymbol{A}_i\boldsymbol{x}_i}$ is the rest of the model not including the element of interest. If we set $\sigma_i^2 = (\boldsymbol{a}_i^T\boldsymbol{C}_{\boldsymbol{d}}^{-1}\boldsymbol{a}_i)^{-1}$, then we achieve "perfect resolution" of this element, and the estimator for each element is

$$\hat{x}_i = (\boldsymbol{a}_i^T\boldsymbol{C}_{\boldsymbol{d}}^{-1}\boldsymbol{a}_i)^{-1}\boldsymbol{a}_i^T\boldsymbol{C}_{\boldsymbol{d}}^{-1}\boldsymbol{d} \quad (32)$$

Although derived from a completely different viewpoint ([10], [12], [13]), this is identically the "linearly constrained minimum variance" solution.

5. NOISE SIGNIFICANCE

We focus again on the specific case of assumed diagonal source covariance $\hat{x}_i = \sigma_i^2\boldsymbol{a}_i^T\boldsymbol{C}_{\boldsymbol{d}}^{-1}\boldsymbol{d} \equiv \boldsymbol{w}_i^T\boldsymbol{d}$, although the approach can be generalized. If $\boldsymbol{d}$ only contains noise, i.e. $\boldsymbol{d} = \boldsymbol{v}$, then we expect the variance of our estimate to be

$$E(\hat{x}_i^2)\Big|_{\boldsymbol{d}=\boldsymbol{v}} = \boldsymbol{w}^T \boldsymbol{C}_{\boldsymbol{v}} \boldsymbol{w}. \qquad (33)$$

If we therefore scale our estimate by this noise expectation, we achieve a "z-statistic" (or "t-statistic" if $\boldsymbol{C}_{\boldsymbol{v}}$ is estimated) shown in *(D)*, and we declare a particular element to be significant if it is substantially greater than unity (e.g. two or three sigma). This approach is the theoretical basis for the "neural activity index" of ([13]), the "SPM-z" of SAM ([10]), and the "dynamic SPM" of ([2]).

Note that as shown in *(D)*, the actual value of σ_i^2 factors out, such that the important specification is to make $\boldsymbol{C}_{\boldsymbol{x}}$ diagonal, but the explicit values are not explicitly needed. We do need, however, $\boldsymbol{C}_{\boldsymbol{d}}$, which is implicitly dependent on $\boldsymbol{C}_{\boldsymbol{x}}$. We note that for the case of wide $\boldsymbol{A}$, most of the fine detail of the diagonal source covariance matrix $\boldsymbol{C}_{\boldsymbol{x}}$ will be compressed into a smaller data covariance matrix $\boldsymbol{C}_{\boldsymbol{d}}$.

6. DIFFERENCES

We have thoroughly manipulated the same basic equation and shown many forms that should be theoretically equivalent. Why are there such disparate differences in the literature? We continue the assumptions that $\boldsymbol{L}$ and $\boldsymbol{C}_{\boldsymbol{v}}$ are the same in all cases.

- Obviously, different specifications of the source covariance $\boldsymbol{C}_{\boldsymbol{j}}$.

But consider that specifying $\boldsymbol{C}_{\boldsymbol{j}}$ implicitly specifies $\boldsymbol{C}_{\boldsymbol{d}}$, the data covariance. So for the exact same data set, different "min norm" techniques begin implicitly with different specifications of a data covariance for these data. Simple statistical tests could be applied to confirm the consistency of the data with the prior, yet surprisingly, many applications appear never to check their data.

- Estimation of $\boldsymbol{C}_{\boldsymbol{d}}$ from the data, rather than specification, from which the source covariance $\boldsymbol{C}_{\boldsymbol{j}}$ is estimated, such as in weighted subspace fitting or LCMV.

Given the same raw data stream, however, different groups segment the data differently, possibly averaging, both in time and frequency domains, and "regularizers" may be added to the estimation. Thus the real distinction between these groups is the specification of what is signal vs. what is noise in the estimation process.

An important assumption (among many) is that $E(\boldsymbol{j}_i \boldsymbol{j}_i^T) \equiv \boldsymbol{C}_{\boldsymbol{j}}$ for all i, but we generally anticipate a dynamic neural process, and thus the data covariance is also a function of time. Thus it is difficult to estimate $\boldsymbol{C}_{\boldsymbol{d}}$ from the data without careful extraction of similar time periods elsewhere in the data stream. Estimating $\boldsymbol{C}_{\boldsymbol{d}}$ from very limited data sets also leads to estimation instabilities.

Therefore, given that a majority of researchers assume independence in the source covariance, and if we consider the useful noise significance conventions, then the primary difference among techniques is the accurate specification or estimation of the data covariance, and not the details of the diagonal elements in the source covariance.

7. SUMMARY

The algorithms commonly in use in EEG and MEG linear inverse analyses are highly related, if not identical, at the theoretical level. The primary differences lie in more practical matters, particularly the estimation or specification of the data covariance. Since modern arrays and acquisition systems provide us with an abundance of data, this community should be able to establish reasonable guidelines for building good estimates or specifications of the data covariance.

Nearly an identical and important issue is the specification or estimation of the noise covariance matrix, which is often ignored, assumed, or not well-explained.

Finally, not all researchers use the cortical surface for mapping current estimations, but rather use volumetric grids, planes, discs, or remap the data as surface potentials or external fields. These alternative surfaces are not wrong, just different, but they make comparisons across studies difficult.

8. REFERENCES

[1] Baillet, Mosher, Leahy, "Electromagnetic Brain Mapping," *IEEE Signal Processing Magazine,* vol 18, issue 6, Nov. 2001, pp. 14–30.

[2] Dale, Liu, Fischl, Buckner, Belliveau, Lewine, Halgren E "Dynamic statistical parametric mapping: combining fMRI and MEG for high-resolution imaging of cortical activity" *Neuron*. 2000 Apr;26(1):55-67.

[3] Grave de Peralta-Menendez, Gonzalez-Andino "A critical analysis of linear inverse solutions to the neuroelectromagnetic inverse problem," *IEEE BME*, vol. 45, pp 440-448, 1998.

[4] Hamalainen, Hari, Illmoniemi, Knuutila, Lounasmaa "Magnetoencephalography - theory, instrumentation, and application to noninvasive studies of the working human brain, *Rev Mod Phys*, APR 1993; v.65, no.2, p.413-497.

[5] Jerbi, Mosher, Baillet, Leahy, "On MEG forward modelling using multipolar expansions," *Physics in Medicine and Biology*, 47 (21 February 2002) 523-555.

[6] Lewis, Mosher, Leahy, "Neuromagnetic source reconstruction," (invited paper) *IEEE ASSP Conference 1995*, Detroit, MI, May 7-12, 1995, vol. 5, pp. 2911–2914.

[7] Mosher, Lewis, and Leahy, "Multiple dipole modeling and localization from spatio-temporal MEG data," *IEEE TBME*, June 1992, vol. 39, pp. 541 – 557.

[8] Mosher, Leahy, "Recursive MUSIC: A framework for EEG and MEG source localization," *IEEE TBME*, vol. 45, pp. 1342-1354, November. 1998.

[9] Mosher, Leahy, Lewis, "EEG and MEG: Forward solutions for inverse methods," *IEEE TBME*, 46:245-259, March 1999.

[10] Robinson, Vrba, "Functional Neuroimaging by Synthetic Aperture Magnetometry (SAM)" Biomag 1998 Sendai, Japan.

[11] Sorenson, *Parameter Estimation*, Marcel Dekker, 1980.

[12] Spencer, Leahy, Mosher, Lewis, "Adaptive filters for monitoring localized brain activity," *IEEE Proc. 26th Asilomar Conf. on Signals, Systems, and Computers*, Pacific Grove, CA, October 1992, pp 156–160.

[13] van Veen, van Drongelen, Yuchtman, Suzuki, "Localization of brain electrical activity via linearly constrained minimum variance spatial filtering" *IEEE TBME* 1997 Sep;44(9):867-80.