Local Linear Estimators for the Bioelectromagnetic Inverse Problem

Richard E. Greenblatt*, Alexei Ossadtchi, and Mark E. Pflieger

Abstract-**Linear estimators have been used widely in the bioelectromagnetic inverse problem, but their properties and relationships have not been fully characterized. Here, we show that the most widely used linear estimators may be characterized by a choice of norms on signal space and on source space. These norms depend, in part, on assumptions about the signal space and source space covariances. We demonstrate that two estimator classes (standardized and weight vector normalized) yield unbiased estimators of source location for simple source models (including only the noise-free case), but biased estimators of source magnitude. In the presence of noise, we show that the non-adaptive standardized estimator is a biased estimator of source location, while the adaptive weight vector normalized estimator remains unbiased. A third class (distortionless) is an unbiased estimator of source magnitude, but a biased estimator of source location.**

I. INTRODUCTION

The bioelectromagnetic inverse problem consists in estimating the locations and magnitudes of a set of equivalent current sources that can account for measured electroencephalographic (EEG) and/or magnetoencephalographic (MEG) data (jointly EMEG).

Conventionally, the source estimation problem (recently reviewed in [1]) has been defined as a global minimization problem, that is, given a set of measurements, find a set of sources that best accounts for the measurements in their entirety. We will refer to this class of solutions as *global* solutions. The global solutions in turn, may be subdivided into *discrete* and *distributed* solutions. The discrete solutions attempt to account for the data in terms of a small number of equivalent sources, while the distributed sources depend on a large number of fixed sources within the brain that may be combined to obtain the desired solution. One property of global solutions is that the parts of the solution are necessarily connected. Omitting one equivalent source, for example, will generally change the position and/or magnitude of other sources. The second class of solutions, the *local* estimators, do not attempt to account for the entire measured signal, but rather attempt to estimate the activity at points or regions of interest independently of one another. These local estimators do not depend on one another, but their summed activity is generally not equal to the measured signal.

Several different local estimators have been proposed in the literature (*e.g.*, [1]-[5]), more or less independently of one another. In this paper, we study the relationships between these estimators, demonstrating that they all may be viewed as members of a single family that differ in their choice of norms for two linear vector spaces that arise naturally out of the bioelectromagnetic inverse problem. We find that the linear estimators may be grouped into four classes: (i) "standardized" (ii) "weight vector normalized", (iii) "distortionless", and (iv) "hybrid". Within each of these classes, estimator may be either adaptive or non-adaptive. For a simple case, that of the scalar beamformer and one or two (unknown) sources, we find that the adaptive weight vector normalized estimators are unbiased estimators of source location, but biased estimators of relative source amplitude. Conversely, the distortionless estimators are unbiased estimators of relative source amplitude, but biased estimators of source location. Many of the results in this paper confirm those recently reported by Sekihara [6], using a somewhat different approach.

Although we often refer in this paper to the head and the brain, the results obtained are more general, and can be applied to the cardiographic inverse problem as well.

II. BIOELECTROMAGNETIC FORWARD AND INVERSE PROBLEMS[1]

A. Lead Fields and the Forward Problem

If we have a known source distribution, the forward problem consists in calculating the electromagnetic field produced by that distribution as a function of space and time. The first step is to state the forward solution for a single time sample and a continuous domain. If we consider a continuous primary current source vector field, $\mathbf{j}(\mathbf{r})$, within the interior volume of the head, Ω (the source volume), and a scalar-valued measurement function $v(\mathbf{r}')$, continuous in spatial parameter $\mathbf{r}'$, then

$$v(\mathbf{r}') = \int_{\Omega} d\mathbf{r}\mathbf{L}(\mathbf{r}',\mathbf{r}) \cdot \mathbf{j}(\mathbf{r}) \quad (1)$$

This work was supported in part by grant MH64343 from the U.S. National Institute of Mental Health to M.E. Pflieger. Asterisk indicates corresponding author.

*R.E. Greenblatt is with Source Signal Imaging, Inc. 2323 Broadway, San Diego CA 92102-1946 USA (e-mail reg@sourcesignal.com)

A. Ossadtchi and M.E. Pflieger are with Source Signal Imaging, Inc. 2323 Broadway, San Diego CA 92102-1946 USA

[1] A note on notation: Matrices are indicates as boldface capitals (*e.g.*, $\mathbf{G}$), column vectors as lowercase bold (*e.g.*, $\mathbf{g}$), and scalars as lowercase italics (*e.g.*, ξ). $\|\cdot\|$ indicates the L2 norm, and $\cdot^T$ the transpose operator. $\mathbf{I}$ is the identity matrix and $\mathbf{1}$ is vectors of 1's. $N(\mu,\sigma^2)$ is the univariate normal distribution with mean μ and variance σ^2, and $\mathbf{N}(\boldsymbol{\mu},\boldsymbol{\Sigma})$ is the multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$.

is the forward solution, where the kernel $\mathbf{L}(\mathbf{r}',\mathbf{r})$ is the current density lead field, a vector field over Ω. The lead field may be calculated using one of several models [7], all of which represent the current flow within the conducting volume of the head with differing degrees of realism.

In most practical cases when we are dealing with experimental data, the measurements are made at discrete (and non-varying) locations and orientations, so we replace $v(\mathbf{r}')$ with the ideal measurement vector $\mathbf{v}$. In addition, the continuous current density distribution $\mathbf{j}(\mathbf{r})$ may be replaced by a vector of current dipoles $\mathbf{q}$. Then the discrete form of the forward problem is given by the matrix equation $\mathbf{v}=\mathbf{Lq}$, where $\mathbf{L}$ is the dipolar lead field matrix.

B. Gain Matrix

While magnetic field measurements may be reference-free, this is not true of the electric field, since the electric potential is only defined as a difference between two locations. For many forward solution calculations, one of these points may be infinity, but typically the measurement is made differentially with respect to a specific reference electrode placed somewhere on or near the head. It is useful to define the measurement (or montage) matrix, **M**, that accounts for the reference. In the case of a single reference electrode, **M** takes the form $\mathbf{M}_{\text{PhysRef}}=[\mathbf{I},-\mathbf{1}]$. Similarly, we can define the average reference, or centering, montage as $\mathbf{M}_{\text{AvgRef}}=\mathbf{I}-\frac{1}{S}\mathbf{1}\mathbf{1}^T$, where S is the number of sensors. We adopt the nomenclature that **M** maps *sensors* to *channels*. Furthermore, **M** need not be square. There may be, for example, fewer channels than sensors. Typically there is no channel that corresponds to the reference sensor. In the case of magnetic measurements, **M** is typically the identity matrix, although many multichannel magnetometers use reference channels to cancel environmental signals, and these magnetic reference channels may be accommodated easily into our formalism.

It may be sometimes the case that the data have been preprocessed by a spatial filter, *e.g.*, a projection matrix obtained from independent components analysis. We represent the spatial filtering process by the derivation matrix **D**. Then the gain matrix **G** is defined by

$$\mathbf{G}=\mathbf{DML} \tag{2}$$

and the forward solution (from dipolar sources to measurement channels) becomes $\mathbf{v}=\mathbf{Gq}$, which we use as our representation of the forward solution for a single time sample in the work we describe here.

C. The inverse problem

We represent the forward problem abstractly as $\mathbf{Q}\xrightarrow{\mathbf{G}}\mathbf{V}$, where $\mathbf{Q}=\mathbb{R}^N$ is the N-dimensional real vector space of sources (the source space, consisting of N sources), and $\mathbf{V}=\mathbb{R}^M$ is the M-dimensional real vector space of measurements (the measurement, or signal, space, consisting of M channels). Then the inverse problem consists in finding a mapping $\mathbf{V}\xrightarrow{\mathbf{G}^{\oplus}}\mathbf{Q}$ from signals to sources. Since $N>>M$, this problem is ill-posed, and solutions must be based on models which reduce the dimension of the solution space.

III. LINEAR ESTIMATORS

A. Linear estimator algebra

We have introduced the two principal vector spaces, **Q** and **V**, source space and signal (or measurement) space, respectively. To solve the linear estimation problem, we need to add additional structure, *i.e.*, a metric or norm. A conventional way to represent this is through the use of dual spaces $\mathbf{Q}^*$ and $\mathbf{V}^*$ [8]. Then $\mathbf{Q}\xrightarrow{\Theta_\mathbf{Q}}\mathbf{Q}^*$ and $\mathbf{V}\xrightarrow{\Theta_\mathbf{V}}\mathbf{V}^*$, where $\Theta_\mathbf{Q}$ and $\Theta_\mathbf{V}$ are symmetric, positive definite matrices. Since $\mathbf{V}^*\xrightarrow{\mathbf{G}^T}\mathbf{Q}^*$, we can write linear inverse solutions as

$$\mathbf{V}\xrightarrow{\Theta_\mathbf{V}}\mathbf{V}^*\xrightarrow{\mathbf{G}^T}\mathbf{Q}^*\xrightarrow{\Theta_\mathbf{Q}^{-1}}\mathbf{Q} \tag{3}$$

or $\mathbf{G}^{\oplus}=\Theta_Q^{-1}\mathbf{G}^T\Theta_V$. Given $\mathbf{G}$, the problem is solved once we know (or choose) suitable metrics on **Q** and **V**. If we want the inverse estimator to be a pseudoinverse (*i.e.*, $\mathbf{G}\mathbf{G}^{\oplus}=\mathbf{I}$), then once we choose $\Theta_\mathbf{Q}$, $\Theta_\mathbf{V}$ is constrained by the requirement that $\mathbf{G}\mathbf{G}^{\oplus}=\mathbf{I}$. If we choose $\Theta_\mathbf{Q}=\mathbf{I}$, we find that $\Theta_\mathbf{V}=\left(\mathbf{G}\mathbf{G}^T\right)^{-1}$, then $\mathbf{G}^{\oplus}=\mathbf{G}^T\left(\mathbf{G}\mathbf{G}^T\right)^{-1}$ is the well-known minimum norm solution. Of course, we can choose some other form for $\Theta_\mathbf{Q}$ (such as a Sobolev norm [9], of which Loreta [10] is an example) or a lead field normalized metric [11].

If we relax the constraint that $\mathbf{G}\mathbf{G}^{\oplus}=\mathbf{I}$, then we are free to choose not only $\Theta_\mathbf{Q}$, but also $\Theta_\mathbf{V}$ as well. This path leads us to local estimators. While we are relatively free in making these choices, not all admissible metrics will lead to useful estimators. In this paper, we will derive several useful metrics as the solution to a set of constrained minimization problems, where the constraints have explicit physical and statistical significance.

Note that, from the pseudoinverse, we can get an estimate for a single scalar source at the target location $\mathbf{r}_0$ as

$$\mathbf{w}_{r_0}^T=\mathbf{g}_{r_0}^T\left(\mathbf{G}\mathbf{G}^T\right)^{-1} \tag{4}$$

where $\mathbf{g}_{r_0}$ is the column of $\mathbf{G}$ corresponding to the source located at $\mathbf{r}_0$. The weight vector $\mathbf{w}_{r_0}^T$ in equation (4) is an example of a local estimator. For an arbitrary set of weight vectors, we can define the scan matrix, $\mathbf{G}^{\oplus}$, as

$$\mathbf{G}^{\oplus}\triangleq\left[\mathbf{w}_0,\ldots,\mathbf{w}_{N-1}\right]^T \tag{5}$$

If $\mathbf{G}\mathbf{G}^{\oplus} = \mathbf{I}$, then $\mathbf{G}^{\oplus}$ is the pseudoinverse of $\mathbf{G}, \mathbf{G}^{+}$. In this paper, we will also refer to the scan matrix as a weight matrix, $\mathbf{W}^T$, where each row is a weight vector, $\mathbf{w}_{\mathbf{r}_0}^T$, that maps a measurement vector into a source space magnitude. This notation is consistent with that frequently used in the beamformer literature.

B. Resolution kernel

Formally, the resolution kernel, $R(\mathbf{r}_0, \mathbf{r})$ may be defined implicitly ([12], [13]) as

$$\hat{q}(\mathbf{r}_0) = \int_{\Omega} d\mathbf{r} R(\mathbf{r}_0, \mathbf{r}) q(\mathbf{r}) \tag{6}$$

where $\hat{q}(\mathbf{r}_0)$ is the estimated source activity at target location $\mathbf{r}_0$ for source distribution $q(\mathbf{r})$. In the discrete case, we can represent the resolution kernel for all target locations and a given scan matrix $\mathbf{W}^T$ as the $N \times N$ matrix $\mathbf{K}$ given by

$$\mathbf{K} = \mathbf{W}^T \mathbf{G} \tag{7}$$

Then for a given true source distribution, $\mathbf{q}$, the estimated source distribution obtained by application of the scan matrix $\mathbf{W}^T$ will be given by

$$\hat{\mathbf{q}} = \mathbf{W}^T \mathbf{G} \mathbf{q} \tag{8}$$

where equation (8) is the discrete version of equation (6). In words, the resolution kernel has the effect of projecting the source space up to the measurement space, then back down into the source space again.

Consider two different views on the physical interpretation of the resolution kernel, each obtained by selecting a single location for analysis. First, assume that the true source distribution is given by a unit source at location $\mathbf{r}_0$. Then equation (8) leads to the interpretation of the resolution kernel as a point spread function, measuring how much the inverse estimate for a point source is "blurred" by the scan matrix. For such a unit point source, we may rewrite (8) as

$$\hat{\mathbf{q}} = \mathbf{W}^T \mathbf{g}_{\mathbf{r}_0} \tag{9}$$

where $\mathbf{g}_{\mathbf{r}_0}$ is the lead field (column) vector for location $\mathbf{r}_0$.

Now consider an alternative view. Choose a target location $\mathbf{r}_0$, and successively place sources of unit strength at each source space location; this can be represented by the $N \times N$ identity matrix. Symbolically we can write $\hat{\mathbf{q}}^T = \mathbf{w}_{\mathbf{r}_0}^T \mathbf{G}\mathbf{I}$, which measures the extent to which sources at all locations project through the estimator for the target location. We may also consider a random distribution of sources across the entire source space, each source independent and identically distributed as $q \sim N(0,1)$. Then the source space covariance is the identity matrix, $\mathbf{I}$. We can ask how much of this uniform activity will be projected onto a target location $\mathbf{r}_0$. Since there is only a single target location, we may consider only the row from $\mathbf{W}^T$ that projects onto the target location, $\mathbf{w}_{\mathbf{r}_0}^T$, and obtain from equation (8) the estimated power, $\hat{q}_{\mathbf{r}_0}^2 = \mathbf{w}_{\mathbf{r}_0}^T \mathbf{G}\mathbf{I}\mathbf{G}^T \mathbf{w}_{\mathbf{r}_0}$. We view $\mathbf{w}_{\mathbf{r}_0}^T \mathbf{G}$ as an interference function, measuring the interference that the scan matrix imposes on our estimate, given an identity source space covariance. This expression plays a central role in our derivation of local estimators, and we will refer to it as the resolution kernel at $\mathbf{r}_0$, or

$$\mathbf{k}_{\mathbf{r}_0}^T = \mathbf{w}_{\mathbf{r}_0}^T \mathbf{G} \tag{10}$$

For the minimum norm inverse, $\mathbf{K}$ is symmetric, so the rows and columns of $\mathbf{K}$ are identical, and thus (9) and (10) lead to the same distribution (*i.e.*, the point spread function and the interference function are identical in the minimum norm case). However, the resolution kernel in the more general case is not necessarily symmetric, so the point spread and interference functions generally will not be the same.

C. Non-adaptive scalar estimators

Our approach is to derive a family of local estimators by finding a set of estimators that minimize the norm of the resolution kernel at $\mathbf{r}_0$, subject to one or more constraints, or, from equation (10)

$$\min \left\| \mathbf{k}_{\mathbf{r}_0} \right\|^2 \tag{11}$$

The constraints, described below, avoid the trivial solution $\left\| \mathbf{k}_{\mathbf{r}_0} \right\|^2 = 0$, and also assure that $\mathbf{w}_{\mathbf{r}_0}$ is actually a useful estimator. So the problem posed is to find the local estimator $\mathbf{w}_{\mathbf{r}_0}$ that is the solution to equation (11), subject to one or more constraints.

From equation (10), we can rewrite the squared norm in equation (11) as

$$\left\| \mathbf{k} \right\|^2 = \mathbf{w}^T \mathbf{G} \mathbf{G}^T \mathbf{w} \tag{12}$$

Then for target location $\mathbf{r}_0$, the problem described in equation. (11) becomes

$$\mathbf{w}_{\mathbf{r}_0} = \underset{\mathbf{w}}{\operatorname{argmin}} \left(\mathbf{w}^T \mathbf{G} \mathbf{G}^T \mathbf{w} \right) \tag{13}$$

A natural (but not unique) constraint is obtained by requiring that

$$\mathbf{w}^T \mathbf{g}_{r_0} = 1 \tag{14}$$

where $\mathbf{g}_{r_0}$ is the lead field (column vector) for location $\mathbf{r}_\mathbf{o}$. In words, equation (14) means that the solution must have unit gain for a source located at $\mathbf{r}_\mathbf{o}$, or, given a measurement distribution corresponding to a point source, the desired solution $\mathbf{w}$ will return an unbiased magnitude estimate for that source. It has therefore been referred to as a "distortionless" beamformer [5].

Now the minimization problem may be stated as

$$\mathbf{w}_{\mathbf{r}_0} = \underset{\mathbf{w}}{\operatorname{argmin}} \left(\mathbf{w}^T \mathbf{G} \mathbf{G}^T \mathbf{w} \right) \text{ s.t. } \mathbf{w}^T \mathbf{g}_{r_0} = 1 \tag{15}$$

The solution to (15) may be obtained using variational methods combined with the use of a Lagrange multiplier ([4], [14]). The Lagrange functional associated with (15) is given by

$$L(\mathbf{w},\lambda)=\mathbf{w}^T\mathbf{G}\mathbf{G}^T\mathbf{w}+2\lambda\left(\mathbf{w}^T\mathbf{g}_{r_0}-1\right) \tag{16}$$

and the solution is found at

$$\left.\frac{\partial L(\mathbf{w},\lambda)}{\partial \mathbf{w}}\right|_{\mathbf{w}=\mathbf{w}_{\mathbf{r}_0}}=0 \tag{17}$$

to be

$$\mathbf{w}_{\mathbf{r}_0}=\frac{\left(\mathbf{G}\mathbf{G}^T\right)^{-1}\mathbf{g}_{r_0}}{\mathbf{g}_{\mathbf{r}_0}^T\left(\mathbf{G}\mathbf{G}^T\right)^{-1}\mathbf{g}_{r_0}} \tag{18}$$

Finally, the estimated source magnitude at $\mathbf{r}_0$ is given by $\hat{q}_{\mathbf{r}_0}=\mathbf{w}_{\mathbf{r}_0}^T\mathbf{v}$.

In the derivation of equation (18), we have ignored any spatial properties of the interference. If we wish to include spatial information, a natural way to do this is to modify equation (12) by the incorporation of a diagonal matrix $\mathbf{\Delta}$, such that $d_{ii}=\left\|\mathbf{r}_0-\mathbf{r}_i\right\|^2$. Then equation (13) becomes

$$\mathbf{w}_{\mathbf{r}_0}=\underset{\mathbf{w}}{\operatorname{argmin}}\left(\mathbf{w}^T\mathbf{G}\mathbf{\Delta}\mathbf{G}^T\mathbf{w}\right) \tag{19}$$

and the derivation proceeds accordingly. This route leads to a solution essentially equivalent to that described in [15]. The principal results we report in this paper obtain whether or not we include $\mathbf{\Delta}$.

What we have obtained with equation (18) is, by construction, an unbiased estimator of the dipole magnitude at any location (due to constraint (14)). However, simulations demonstrate that when equation (18) is used to scan for the maximum current, it will provide a biased estimate of the dipole location. This is discussed in greater detail below.

The physical explanation for the failure of (18) to identify the correct location lies is the nonlinear dependence of the lead field on depth. In order to overcome this bias, we need to add an additional constraint that will explicitly overcome the depth dependence of the lead field. As proposed in [5] (see also[16]), we will constrain the solution for uniform gain as

$$\mathbf{w}^T\mathbf{w}=1 \tag{20}$$

The constraint $\mathbf{w}^T\mathbf{w}=1$ assures that the gain will be the same at all locations. Then to find the solution to (21), we follow [5] by adding a new (pseudo)constraint $\mathbf{w}^T\mathbf{g}=\xi$ where the unknown scalar ξ may be found from (20) after solving the minimization problem(22).

$$\mathbf{w}_{\mathbf{r}_0}=\underset{\mathbf{w}}{\operatorname{argmin}}\left(\mathbf{w}^T\mathbf{G}\mathbf{G}^T\mathbf{w}\right)\text{ s.t. }\mathbf{w}^T\mathbf{g}=\xi\text{ and }\mathbf{w}_\xi^T\mathbf{w}_\xi=1 \tag{21}$$

What equation (21) means in words is that we first find a one-parameter family of solutions to the constrained minimization problem, $\mathbf{w}_\xi$. All of these solutions "point" in the same direction in measurement space, but their norm (equivalently, gain) is undefined. Then we use the second constraint, $\mathbf{w}_\xi^T\mathbf{w}_\xi=1$, to single out a specific length that satisfies the condition. We can solve this in two steps. First solve the variational problem

$$\mathbf{w}_{\mathbf{r}_0,\xi}=\underset{\mathbf{w}}{\operatorname{argmin}}\left(\mathbf{w}^T\mathbf{G}\mathbf{G}^T\mathbf{w}\right)\text{ s.t. }\mathbf{w}^T\mathbf{g}=\xi \tag{22}$$

The solution to (22) is obtained by steps essentially similar to equations (16) and (17) and we obtain

$$\mathbf{w}_{\mathbf{r}_0,\xi}=\xi\frac{\left(\mathbf{G}\mathbf{G}^T\right)^{-1}\mathbf{g}_{\mathbf{r}_0}}{\mathbf{g}_{\mathbf{r}_0}^T\left(\mathbf{G}\mathbf{G}^T\right)^{-1}\mathbf{g}_{\mathbf{r}_0}} \tag{23}$$

To determine the value of ξ, combine (20) and (23) to obtain

$$1=\mathbf{w}_{\mathbf{r}_0,\xi}^T\mathbf{w}_{\mathbf{r}_0,\xi}=\xi^2\left[\frac{\left(\mathbf{G}\mathbf{G}^T\right)^{-1}\mathbf{g}_{r_0}}{\mathbf{g}_{\mathbf{r}_0}^T\left(\mathbf{G}\mathbf{G}^T\right)^{-1}\mathbf{g}_{\mathbf{r}_0}}\right]^T\left[\frac{\left(\mathbf{G}\mathbf{G}^T\right)^{-1}\mathbf{g}_{\mathbf{r}_0}}{\mathbf{g}_{\mathbf{r}_0}^T\left(\mathbf{G}\mathbf{G}^T\right)^{-1}\mathbf{g}_{\mathbf{r}_0}}\right]$$

After some rearrangement, we obtain

$$\xi=\frac{\mathbf{g}_{r_0}^T\left(\mathbf{G}\mathbf{G}^T\right)^{-1}\mathbf{g}_{r_0}}{\left[\mathbf{g}_{r_0}^T\left(\mathbf{G}\mathbf{G}^T\right)^{-2}\mathbf{g}_{r_0}\right]^{1/2}} \tag{25}$$

Then taking (23) and (25) and simplifying, we get

$$\mathbf{w}_{\mathbf{r}_0}=\frac{\left(\mathbf{G}\mathbf{G}^T\right)^{-1}\mathbf{g}_{\mathbf{r}_0}}{\left[\mathbf{g}_{\mathbf{r}_0}^T\left(\mathbf{G}\mathbf{G}^T\right)^{-2}\mathbf{g}_{\mathbf{r}_0}\right]^{1/2}} \tag{26}$$

and the estimated dipole magnitude given by $\hat{q}_{\mathbf{r}_0}=\mathbf{w}_{\mathbf{r}_0}^T\mathbf{v}$. As we show in the Appendix, the weight vector obtained from equation (26) yields a biased estimator of source location.

Notice that constraints (15) or (21) introduced a new norm into the estimator problem when demanding that $\mathbf{w}_\xi^T\mathbf{w}_\xi=1$, or equivalently $\mathbf{w}_\xi^T\mathbf{I}\mathbf{w}_\xi=1$. It is, of course, possible to choose some other metric on the space of weight vectors. One natural choice is the matrix $\mathbf{G}\mathbf{G}^T$, which constraints the expected value of the interference from brain sources at all locations to be uniform. Then the relevant minimization problem become

$$\underset{\mathbf{w}}{\operatorname{argmin}}\left(\mathbf{w}^T\mathbf{G}\mathbf{G}^T\mathbf{w}\right)\text{ s.t. }\mathbf{w}^T\mathbf{g}=\xi\text{ and }\mathbf{w}_\xi^T\mathbf{G}\mathbf{G}^T\mathbf{w}_\xi=1 \tag{27}$$

Following steps equivalent to equations (22) and (23), and combining with the constraint $\mathbf{w}_\xi^T\mathbf{G}\mathbf{G}^T\mathbf{w}_\xi=1$, we obtain

$$\xi=\left[\mathbf{g}_{\mathbf{r}_0}^T\left(\mathbf{G}\mathbf{G}^T\right)^{-1}\mathbf{g}_{r_0}\right]^{1/2} \tag{28}$$

Then substituting (28) into (23), we obtain the desired weight vector as

$$\mathbf{w}_{\mathbf{r}_0} = \frac{\left(\mathbf{G}\mathbf{G}^T\right)^{-1}\mathbf{g}_{r_0}}{\left[\mathbf{g}_{r_0}^T\left(\mathbf{G}\mathbf{G}^T\right)^{-1}\mathbf{g}_{r_0}\right]^{1/2}} \tag{29}$$

In [17], Pascual-Marqui obtained a local estimator, which he calls sLoreta. The power output of the estimator obtained from equation (29) is exactly equivalent to that obtained using sLoreta. However, scanning estimators derived from equation (29), including sLoreta, are biased estimator of source location in the presence of noise, as we show in the Appendix.

The estimators which we describe in this section have been called non-adaptive [6] since their functional form depends only on the forward solution, and not on the data.

D. *Adaptive scalar estimators*

If $\mathbf{q}(t)$ is a multivariate random process with covariance $\mathbf{R}_\mathbf{Q}$, and $\mathbf{G}$ is a linear mapping, $\mathbf{Q}\xrightarrow{\mathbf{G}}\mathbf{V}$, then (see [24]), the signal space covariance of $\mathbf{v}(t)$ in the absence of instrumental and environmental noise is $\mathbf{R}_\mathbf{V} = \mathbf{G}\mathbf{R}_\mathbf{Q}\mathbf{G}^T$. This is the bridge from the estimators that we have just derived to minimum variance beamformers typically used in signal processing applications, including bioelectromagnetic applications (*e.g.*, [4], [5], [15], [17], [18]). If we assume that $\mathbf{R}_\mathbf{Q}$ is unit variance and uncorrelated (*i.e.*, the identity matrix) then $\mathbf{R}_\mathbf{V} = \mathbf{G}\mathbf{G}^T$. Of course, we can estimate $\mathbf{R}_\mathbf{V}$ directly from the data, assuming that the underlying random process is stationary. Using the estimated $\mathbf{R}_\mathbf{V}$ in place of the theoretical $\mathbf{R}_\mathbf{V}$, the form of the weight vector equations we have derived remains unchanged, but all instances of $\mathbf{G}\mathbf{G}^T$ are replaced with $\mathbf{R}_\mathbf{V}$. This is shown in Table 3, where we represent either $\mathbf{G}\mathbf{G}^T$ or $\mathbf{R}_\mathbf{V}$ as the matrix $\mathbf{\Gamma}$. We call the $\mathbf{\Gamma} = \mathbf{G}\mathbf{G}^T$ case the non-adaptive case, and $\mathbf{\Gamma} = \mathbf{R}_\mathbf{V}$, the adaptive case [6].

In general, the adaptive solution will have better resolution that the non-adaptive solution [6], but the accuracy of the solution will depend critically on a suitable, data-based, estimate of the signal space covariance matrix. Since the background (or noise) covariance is generally different from the signal covariance, using the noise covariance generally will give an incorrect estimate of the source location and magnitude.

Table 1 shows that within each class (non-adaptive or adaptive) the local estimators that we have derived differ by a scale factor, *i.e.*, a scalar field over the source space. The entries in the table may by calculated by observing that within the adaptive or non-adaptive classes, the estimators differ only in the source space metric. What this means is that once the source space covariance is chosen (either $\mathbf{I}$ or implicitly from $\mathbf{R}_\mathrm{V}$), the subclasses (distortionless, weight vector normalized, or standardized) are all represented by vector fields that are directionally equivalent (that is, at any location, the three weight vectors all point in the same direction), and these vectors differ only in magnitude.

	Distortion-less	*weight vector normalized*	*standardized*
Distortion-less	1	$\dfrac{\left(\mathbf{g}_r^T\mathbf{\Gamma}^{-2}\mathbf{g}_r\right)^{1/2}}{\mathbf{g}_r^T\mathbf{\Gamma}^{-1}\mathbf{g}_r}$	$\left(\mathbf{g}_r^T\mathbf{\Gamma}^{-1}\mathbf{g}_r\right)^{-1/2}$
weight vector normalized	$\dfrac{\mathbf{g}_r^T\mathbf{\Gamma}^{-1}\mathbf{g}_r}{\left(\mathbf{g}_r^T\mathbf{\Gamma}^{-2}\mathbf{g}_r\right)^{1/2}}$	1	$\dfrac{\left(\mathbf{g}_r^T\mathbf{\Gamma}^{-1}\mathbf{g}_r\right)^{1/2}}{\left(\mathbf{g}_r^T\mathbf{\Gamma}^{-2}\mathbf{g}_r\right)^{1/2}}$
Standard-ized	$\left(\mathbf{g}_r^T\mathbf{\Gamma}^{-1}\mathbf{g}_r\right)^{1/2}$	$\dfrac{\left(\mathbf{g}_r^T\mathbf{\Gamma}^{-2}\mathbf{g}_r\right)^{1/2}}{\left(\mathbf{g}_r^T\mathbf{\Gamma}^{-1}\mathbf{g}_r\right)^{1/2}}$	1

Table 1. Each cell in the table indicates the scalar field equation over the source volume, Ω, that may be used to map the estimator indicated by the j^th column to the estimator indicated by the i^th row, when estimators are of the same type (either non-adaptive or adaptive). To change between non-adaptive and adaptive, Γ (as defined in Table 3) should be substituted as indicated, equivalent to a tensor field over Ω.

E. *Hybrid estimators*

All of the linear estimators that we have so far described are obtained as the solution to a minimization problem subject to specified constraints. However, several estimators described in the literature (*e.g.*, [4], [18], [19]) do not fall neatly into this constrained optimization framework, although they are encompassed by the linear estimator model described in III.A above. These hybrid estimators begin as one of the location-biased estimators that we have described in III.C and III.D. For example, [4] and [18] begin with a minimum variance distortionless beamformer, while [19] begins with a lead field normalized minimum norm estimator. Then the hybrid methods add a second normalization (equivalent to imposing a new norm on source space) to more-or-less successfully compensate for the location bias in the initial estimator. This is done generally by using the instrumental noise covariance projected back to the source space by the initial estimator matrix to obtain an estimate of the source space covariance (*i.e.*, a new source space metric). In this paper, we do not study the properties of these hybrid estimators in further detail.

F. *Vector estimators*

The scalar estimators described in Tables 1 and 2 assume that we seek to estimate the dipole magnitude at a target location, given a fixed orientation. This would be the case, for example, when scanning with a cortically constrained source model, where the dipoles are constructed to be oriented normal to the cortical surface. Often, the dipole orientations are not known *a priori*, and there is a three-space of current dipoles at each location. The solution to this problem involves the application of vector estimators. Such estimators, clearly of considerable practical importance, are described by a number of authors (including [4], [5], [15], [17], [18]), and we refer the interested reader to these sources. Vector estimators may be understood according to the same methods used to develop Table 3, and their detailed discussion here would not add any new conceptual information to the questions that we address here.

In Table 2, we classify several previously reported vector estimators in terms of the descriptions used in this paper.

Class	*Relatives*
Non-adaptive distortionless	WROP [15]
Non-adaptive weight vector normalized	
Non-adaptive standardized	sLoreta [17]
Adaptive distortionless	Distortionless LCMV [4], [5], [6]
Adaptive weight vector normalized	Borgiotti-Kaplan [5]
Adaptive standardized	
Hybrid	Neural activity index [4]; SAM [18]; DSPM [19]

Table 2. This table compares the linear estimator classes that we derive with those reported in the literature for use with the bioelectromagnetic inverse problem. We analyze scalar estimators in this paper, but practical applications typically use vector estimators, so not all properties of these vector estimators may be accounted for by the classification system we propose. For example, SAM [18] is an orientation-adaptive vector beamformer, and the orientation adaptation algorithm has no clear scalar counterpart.

G. Estimation Bias

A statistical estimator is unbiased if the expected value of the estimated parameter is the value of the physical quantity to be estimated, symbolically $E(\hat{x}) = x$ for some parameter *x*. In this work we consider two different estimators, both obtained from the same set of weight vectors. The dipole magnitude estimator (or beamformer) is given by

$$\hat{q} = \mathbf{w}^T \mathbf{v} \tag{30}$$

and the location (or scanning) estimator is given by

$$\hat{\mathbf{r}} = \arg\max_{\mathbf{r}} \left(\mathbf{w}_{\mathbf{r}}^T \mathbf{v}\mathbf{v}^T \mathbf{w}_{\mathbf{r}} \right) \tag{31}$$

For the estimators given by equations (30) and (31), we now consider three different conditions for determining estimator bias.

1) Source space localization estimation for a single dipolar source. Often estimators are used in the scanning mode to estimate the location of one or more sources in the brain. This may be done iteratively or through use of the scan matrix (equation *(5)*). If the true source location is $\mathbf{r}_0$, then the scan is unbiased iff $E(\hat{\mathbf{r}}) = \mathbf{r}_0$. In the Appendix, we show that, in the presence of additive white noise, this condition is met only by the adaptive weight vector normalized estimators. However, as we also show in the Appendix, the non-adaptive standardized estimator (sLoreta) is unbiased only in the absence of noise. This is consistent with the simulations reported in [17] where correct location is reported for the maximum entropy standardized beamformer (sLoreta), but only in the noise-free case. These conclusions are also consistent with the results reported in [6].

2) *Source magnitude estimation for a pair of uncorrelated sources.* The other common use of weight vectors is to estimate a source time series, given a location. In the case of a pair of uncorrelated sources, the distortionless estimators will correctly estimate the time series for each of a pair of sources, or $E(\hat{q}(t)) = q(t)$. This is imposed by the constraint that $\mathbf{w}_{\mathbf{r}_0}^T \mathbf{g}_{r_0} = 1$ (equation *(14)*), for both the non-adaptive and the adaptive distortionless estimators. In the adaptive case, however, the condition will only be met if the sources are uncorrelated (see III.G.3) below). However, both the standardized and the weight vector normalized estimators will be biased estimators of relative source magnitude. This may be seen by considering two sources at different locations of unit magnitude. Then, for the standardized beamformer (from equation *(29)*)

$$\frac{E(\hat{q}_0)}{E(\hat{q}_1)} = \frac{\left(\mathbf{g}_{r_0}^T \mathbf{\Gamma}^{-1} \mathbf{g}_{r_0} E(q_0)\right)^{1/2}}{\left(\mathbf{g}_{r_1}^T \mathbf{\Gamma}^{-1} \mathbf{g}_{r_1} E(q_1)\right)^{1/2}} \neq 1 \text{ in general} \tag{32}$$

A similar analysis applies to all but the distortionless estimators. This bias has an effect on the scanning beamformer, since deeper sources will appear smaller than shallower sources of the same magnitude when the scanner output is viewed as a map.

3) Source magnitude estimation for a pair of correlated sources. It is known (*e.g.*, [4]) that sources whose time series are highly correlated in time will be mislocalized by minimum variance estimators. This may be understood by considering the signal space covariance structure. When two sources are completely correlated, the rank of the covariance will be reduced, since the signals are always represented by a single vector direction in measurement space. Because the non-adaptive estimators do not use the estimated signal space covariance, they will not exhibit this problem. This correlation bias has different implications for electric and magnetic measurements. In the electrical case, a single tangential dipole may appear quite similar in its signal space projection to a pair of completely correlated radial dipoles, and linear estimators often have practical problems distinguishing between these two possibilities. If we assume that real sources are generally only partially correlated, the correlation bias imposed by the minimum variance estimators may actually be of benefit in identifying these tangential sources. The problem of distinguishing between radial and tangential sources arises to a much more limited extent with magnetic measurements, since magnetic measurements have little or no signal derived from radial sources.

IV. DISCUSSION

We have shown that linear estimators used for bioelectromagnetic source estimation may be classified by the metrics that they impose on source space and signal space. As a general rule, the signal space metric is the principal determinant of the spatial resolution, and the source space metric is the principal determinant of estimator bias. These observations lead to some practical guidelines for the selection of linear estimators. When the signal space covariance can be estimated accurately, which in turn depends critically on the assumption of stationarity, minimum variance estimators will yield better spatial resolution than will maximum entropy estimators. In addition, only the adaptive weight vector normalized estimator can be used as an unbiased estimator of source location in the presence of noise. However, if an unbiased comparison is required between source magnitudes at different locations, then distortionless estimators should be used.

The use of weight vector normalized scanning estimators is limited, however, by its magnitude ratio estimation bias, which will lead to deeper sources to appear weaker when using these estimators. We consider briefly two methods to overcome this problem. The first uses a recursively applied local estimator algorithm, and is based on methods described originally in [20], and also shares some common features with the RAP Music algorithm [21]. In this approach, we first apply a scanning estimator to identify a source location peak. We can then construct a new constrained scanning beamformer that has zero gain at the previously identified target location. Successive applications of these increasingly constrained scanning estimators until some stopping criterion is satisfied will yield a set of sources that can account for the observed data.

The second approach to the multiple source problem that we consider is randomization testing [22]. This approach is applicable to event related datasets, where multiple samples are available. Randomization tests have the added advantage that the estimated parameters may be associated with data-derived distribution, and thus with probability values.

Although we have motivated our derivation of linear estimators beginning with the need to minimize interference, this report does not describe methods for estimating the interference between the target location and the rest of the brain. This problem, essentially that of describing receiver operator characteristics for source estimators, is addressed elsewhere [25].

Finally, it is worth noting that nothing we have proposed permits us to circumvent the non-uniqueness problem, *i.e.*, suitably arranged linear source combinations may always be found that will mimic any other source combination. The methods we have investigated here assume, however, that satisfactory models for many problems may be obtained from simple, point-like equivalent sources, since these are what the linear estimators are searching for.

APPENDIX

In the appendix, we provide detailed calculations on the extrema of several scanning estimators, both in the noise-free case, and in the presence of noise.

We consider as the statistical model

$$\mathbf{v}(t) = \mathbf{g}_{\mathbf{r}_0} q(t) + \boldsymbol{\eta}(t) \tag{A.1}$$

$$q \sim N(0,1),\ \boldsymbol{\eta} \sim \mathbf{N}(\mathbf{0},\mathbf{I}) \tag{A.2}$$

$$\mathbf{R}_v = \mathbf{g}_{\mathbf{r}_0}\mathbf{g}_{\mathbf{r}_0}^T + \mathbf{I} \tag{A.3}$$

The general problem is: given a scanning beamformer with weight vectors $\mathbf{w}_\mathbf{r}$, $\mathbf{r} \in \Omega$, and a single scalar source $q_{\mathbf{r}_0}$, find the estimator

$$\hat{\mathbf{r}} = \arg\max_{\mathbf{r}} \left(\mathbf{w}_\mathbf{r}^T E\left(\mathbf{v}\mathbf{v}^T\right) \mathbf{w}_\mathbf{r} \right) \tag{A.4}$$

where $E\left(\hat{j}_D^2\right) = \dfrac{\mathbf{g}^T \boldsymbol{\Gamma}^{-1} \mathbf{R}_v \boldsymbol{\Gamma}^{-1} \mathbf{g}}{\left(\mathbf{g}^T \boldsymbol{\Gamma}^{-1} \mathbf{g}\right)^2}$. However, we can solve a simpler problem by assuming that $E(\hat{\mathbf{r}}) = \mathbf{r}_0$, then determine if it is true that $\left.\dfrac{\partial\left(\mathbf{w}_\mathbf{r}^T \mathbf{R}_v \mathbf{w}_\mathbf{r}\right)}{\partial \mathbf{w}}\right|_{\mathbf{w}=\mathbf{w}_{\mathbf{r}_0}} = 0$. In this Appendix, we consider only the first derivative, since it is sufficient for our purposes to determine if the scanning estimators show an extremum at the target location.

A. Adaptive weight vector normalized estimator

We will consider first the case of the adaptive weight vector normalized estimator in the presence of noise. Other scalar cases may be solved using similar methods to those detailed here.

Begin with $E\left(\hat{j}_{\text{A-WVN}}^2\right) = \dfrac{\mathbf{g}^T \mathbf{R}_v^{-1} \mathbf{R}_v \mathbf{R}_v^{-1} \mathbf{g}}{\mathbf{g}^T \mathbf{R}_v^{-2} \mathbf{g}} = \dfrac{\mathbf{g}^T \mathbf{R}_v^{-1} \mathbf{g}}{\mathbf{g}^T \mathbf{R}_v^{-2} \mathbf{g}}$. Then, by an abuse of notation,

$$\frac{1}{2}\frac{\partial \hat{j}^2}{\partial \mathbf{g}} = \frac{\mathbf{R}_v^{-1}\mathbf{g}\left(\mathbf{g}^T \mathbf{R}_v^{-2}\mathbf{g}\right) - \left(\mathbf{g}^T \mathbf{R}_v^{-1}\mathbf{g}\right)\mathbf{R}_v^{-2}\mathbf{g}}{\left(\mathbf{g}^T \mathbf{R}_v^{-2}\mathbf{g}\right)^2} \tag{A.5}$$

From the matrix inversion lemma [20],

$$\mathbf{R}_v^{-1} = \left(\mathbf{I} - \left(\frac{1}{1 + \mathbf{g}_{\mathbf{r}_0}^T \mathbf{g}_{\mathbf{r}_0}} \right) \mathbf{g}_{\mathbf{r}_0} \mathbf{g}_{\mathbf{r}_0}^T \right) \tag{A.6}$$

Let. $a = \mathbf{g}_{\mathbf{r}_0}^T \mathbf{g}_{\mathbf{r}_0}$ Then, after substitution into (A.6), multiplication, and rearrangement of terms, we get

$$\mathbf{R}_v^{-2} = \mathbf{I} - \frac{2+a}{(1+a)^2}\mathbf{g}_{\mathrm{r}_0}\mathbf{g}_{\mathrm{r}_0}^T \qquad \text{(A.7)}$$

It will also be useful to see that, from equation (A.6)

$$\mathbf{g}_{\mathrm{r}_0}^T\mathbf{R}_v^{-1}\mathbf{g}_{\mathrm{r}_0} = \frac{a}{1+a} \qquad \text{(A.8)}$$

and from equation (A.7)

$$\mathbf{g}_{\mathrm{r}_0}^T\mathbf{R}_v^{-2}\mathbf{g}_{\mathrm{r}_0} = \frac{a}{(1+a)^2} \qquad \text{(A.9)}$$

Now we have established the preliminaries to evaluate $\left.\frac{\partial \hat{j}^2}{\partial \mathbf{g}}\right|_{\mathbf{g}=\mathbf{g}_0}$. Substituting (A.8) and (A.9) into (A.5), we obtain

$$\left.\frac{1}{2}\frac{\partial \hat{j}^2}{\partial \mathbf{g}}\right|_{\mathbf{g}=\mathbf{g}_0} = \frac{\left(I - \frac{1}{1+a}\mathbf{g}_{\mathrm{r}_0}\mathbf{g}_{\mathrm{r}_0}^T\right)\mathbf{g}_{\mathrm{r}_0}\frac{a}{(1+a)^2} - \frac{a}{1+a}\left(I - \frac{2+a}{(1+a)^2}\mathbf{g}_{\mathrm{r}_0}\mathbf{g}_{\mathrm{r}_0}^T\right)\mathbf{g}_{\mathrm{r}_0}}{\left(\mathbf{g}^T\mathbf{R}_v^{-2}\mathbf{g}\right)^2}$$

The solution is found at

$$\left(I - \frac{1}{1+a}\mathbf{g}_{\mathrm{r}_0}\mathbf{g}_{\mathrm{r}_0}^T\right)\mathbf{g}_{\mathrm{r}_0}\frac{a}{(1+a)^2} = \frac{a}{1+a}\left(I - \frac{2+a}{(1+a)^2}\mathbf{g}_{\mathrm{r}_0}\mathbf{g}_{\mathrm{r}_0}^T\right)\mathbf{g}_{\mathrm{r}_0}$$

which after some rearrangement becomes

$$\mathbf{n} = \left(I - \frac{\mathbf{g}_{\mathrm{r}_0}\mathbf{g}_{\mathrm{r}_0}^T}{1+a}\right)\frac{a}{(1+a)^2}\mathbf{g}_{\mathrm{r}_0} - \frac{a}{1+a}\left(I - \frac{2+a}{(1+a)^2}\mathbf{g}_{\mathrm{r}_0}\mathbf{g}_{\mathrm{r}_0}^T\right)$$

Then it is straightforward (if tedious) to show that $\mathbf{n} = \mathbf{0}$. So we have demonstrated that, given the statistical model in (A.1)-(A.3), the minimum variance weight vector normalized estimator is unbiased with respect to source location. We omit the demonstration that it is unbiased in the absence of noise as well.

B. Non-adaptive weight vector normalized – uncorrelated noise

Begin with $E\left(\hat{j}^2_{\text{NA-WVN}}\right) = \frac{\mathbf{g}^T\mathbf{\Gamma}^{-1}\mathbf{R}_v\mathbf{\Gamma}^{-1}\mathbf{g}}{\mathbf{g}^T\mathbf{\Gamma}^{-2}\mathbf{g}}$ where $\mathbf{\Gamma} = \mathbf{G}\mathbf{G}^T$.

Then

$$\frac{\partial \hat{j}^2}{\partial \mathbf{g}} = \frac{2\left(\left(\mathbf{g}^T\mathbf{\Gamma}^{-2}\mathbf{g}\right)\mathbf{S}\mathbf{g} - \left(\mathbf{g}^T\mathbf{S}\mathbf{g}\right)\mathbf{\Gamma}^{-2}\mathbf{g}\right)}{\left(\mathbf{g}^T\mathbf{\Gamma}^{-2}\mathbf{g}\right)^2} \qquad \text{(A.11)}$$

and $\mathbf{S} = \mathbf{\Gamma}^{-1}\mathbf{R}_v\mathbf{\Gamma}^{-1}$. The solution is found at

$$\left.\frac{\partial \hat{j}^2}{\partial \mathbf{g}}\right|_{\mathbf{g}=\mathbf{g}_{\mathrm{r}_0}} = 0 \Leftrightarrow \left(\mathbf{g}_{\mathrm{r}_0}^T\mathbf{\Gamma}^{-2}\mathbf{g}_{\mathrm{r}_0}\right)\mathbf{S}_{\mathrm{r}_0}\mathbf{g}_{\mathrm{r}_0} = \left(\mathbf{g}_{\mathrm{r}_0}^T\mathbf{S}_{\mathrm{r}_0}\mathbf{g}_{\mathrm{r}_0}\right)\mathbf{\Gamma}^{-2}\mathbf{g}_{\mathrm{r}_0} \qquad \text{(A.12)}$$

For $\mathbf{R}_v = \mathbf{I} + \mathbf{g}_{\mathrm{r}_0}\mathbf{g}_{\mathrm{r}_0}^T$, the left term in (A.12) becomes

$$\begin{aligned}\left(\mathbf{g}_{\mathrm{r}_0}^T\mathbf{\Gamma}^{-2}\mathbf{g}_{\mathrm{r}_0}\right)\mathbf{S}_{\mathrm{r}_0}\mathbf{g}_{\mathrm{r}_0} &= \left(\mathbf{g}_{\mathrm{r}_0}^T\mathbf{\Gamma}^{-2}\mathbf{g}_{\mathrm{r}_0}\right)\mathbf{\Gamma}^{-1}\left(I + \mathbf{g}_{\mathrm{r}_0}\mathbf{g}_{\mathrm{r}_0}^T\right)\mathbf{\Gamma}^{-1}\mathbf{g}_{\mathrm{r}_0} \\ &= \left(\mathbf{g}_{\mathrm{r}_0}^T\mathbf{\Gamma}^{-2}\mathbf{g}_{\mathrm{r}_0}\right)\mathbf{\Gamma}^{-2}\mathbf{g}_{\mathrm{r}_0} + \left(\mathbf{g}_{\mathrm{r}_0}^T\mathbf{\Gamma}^{-1}\mathbf{g}_{\mathrm{r}_0}\right)\left(\mathbf{g}_{\mathrm{r}_0}^T\mathbf{\Gamma}^{-2}\mathbf{g}_{\mathrm{r}_0}\right)\mathbf{\Gamma}^{-1}\mathbf{g}_{\mathrm{r}_0}\end{aligned}$$

and the right term in (A.12) becomes

$$\begin{aligned}\left(\mathbf{g}_{\mathrm{r}_0}^T\mathbf{S}_{\mathrm{r}_0}\mathbf{g}_{\mathrm{r}_0}\right)\mathbf{\Gamma}^{-2}\mathbf{g}_{\mathrm{r}_0} &= \left(\mathbf{g}_{\mathrm{r}_0}^T\mathbf{\Gamma}^{-1}\left(I + \mathbf{g}_{\mathrm{r}_0}\mathbf{g}_{\mathrm{r}_0}^T\right)\mathbf{\Gamma}^{-1}\mathbf{g}_{\mathrm{r}_0}\right)\mathbf{\Gamma}^{-2}\mathbf{g}_{\mathrm{r}_0} \\ &= \left(\mathbf{g}_{\mathrm{r}_0}^T\mathbf{\Gamma}^{-2}\mathbf{g}_{\mathrm{r}_0}\right)\mathbf{\Gamma}^{-2}\mathbf{g}_{\mathrm{r}_0} + \left(\mathbf{g}_{\mathrm{r}_0}^T\mathbf{\Gamma}^{-1}\mathbf{g}_{\mathrm{r}_0}\right)^2\mathbf{\Gamma}^{-2}\mathbf{g}_{\mathrm{r}_0}\end{aligned}$$

Since, in general, $\left(\mathbf{g}_{\mathrm{r}_0}^T\mathbf{\Gamma}^{-1}\mathbf{g}_{\mathrm{r}_0}\right)\left(\mathbf{g}_{\mathrm{r}_0}^T\mathbf{\Gamma}^{-2}\mathbf{g}_{\mathrm{r}_0}\right)\mathbf{\Gamma}^{-1}\mathbf{g}_{\mathrm{r}_0} \neq \left(\mathbf{g}_{\mathrm{r}_0}^T\mathbf{\Gamma}^{-1}\mathbf{g}_{\mathrm{r}_0}\right)^2\mathbf{\Gamma}^{-2}\mathbf{g}_{\mathrm{r}_0}$, the non-adaptive weight vector normalized estimator yields a biased location estimator.

C. Distortionless estimator

Let $E\left(\hat{j}^2_D\right) = \frac{\mathbf{g}^T\mathbf{\Gamma}^{-1}\mathbf{R}_v\mathbf{\Gamma}^{-1}\mathbf{g}}{\left(\mathbf{g}^T\mathbf{\Gamma}^{-1}\mathbf{g}\right)^2}$. Then

(A.10)

$$\frac{\partial \hat{j}^2}{\partial \mathbf{g}} = \frac{\left(2\left(\mathbf{g}^T\mathbf{\Gamma}^{-1}\mathbf{g}\right)^2\mathbf{S}\mathbf{g} - \left(\mathbf{g}^T\mathbf{S}\mathbf{g}\right)\frac{\partial}{\partial \mathbf{g}}\left(\mathbf{g}^T\mathbf{\Gamma}^{-1}\mathbf{g}\right)^2\right)}{\left(\mathbf{g}^T\mathbf{\Gamma}^{-1}\mathbf{g}\right)^2} \qquad \text{(A.15)}$$

where $\mathbf{S} = \mathbf{\Gamma}^{-1}\mathbf{R}_v\mathbf{\Gamma}^{-1}$ and $\frac{\partial}{\partial \mathbf{g}}\left(\mathbf{g}^T\mathbf{\Gamma}^{-1}\mathbf{g}\right)^2 = 4\left(\mathbf{g}^T\mathbf{\Gamma}^{-1}\mathbf{g}\right)\mathbf{\Gamma}^{-1}\mathbf{g}$.

For equation (A.15). to be $\mathbf{0}$, it is sufficient to show that the numerator be equal to zero,

$$\left.\frac{\partial \hat{j}^2}{\partial \mathbf{g}}\right|_{\mathbf{g}=\mathbf{g}_{\mathrm{r}_0}} = 0 \Leftrightarrow \left(\mathbf{g}_{\mathrm{r}_0}^T\mathbf{\Gamma}^{-1}\mathbf{g}_{\mathrm{r}_0}\right)\mathbf{S}\mathbf{g}_{\mathrm{r}_0} = 2\left(\mathbf{g}_{\mathrm{r}_0}^T\mathbf{S}\mathbf{g}_{\mathrm{r}_0}\right)\mathbf{\Gamma}^{-1}\mathbf{g}_{\mathrm{r}_0} \qquad \text{(A.16)}$$

Equation (A.16) cannot be true. To see this, rearrange (A.16) and left-multiply both sides by $\mathbf{g}_{\mathrm{r}_0}^T$ to obtain $\frac{\mathbf{g}_{\mathrm{r}_0}^T\mathbf{S}\mathbf{g}_{\mathrm{r}_0}}{2\left(\mathbf{g}_{\mathrm{r}_0}^T\mathbf{S}\mathbf{g}_{\mathrm{r}_0}\right)} = \frac{\mathbf{g}_{\mathrm{r}_0}^T\mathbf{\Gamma}^{-1}\mathbf{g}_{\mathrm{r}_0}}{\left(\mathbf{g}_{\mathrm{r}_0}^T\mathbf{\Gamma}^{-1}\mathbf{g}_{\mathrm{r}_0}\right)} \Leftrightarrow \frac{1}{2} = 1$, which is a contradiction. Thus distortionless beamformers cannot be unbiased location estimators. Following the same argument, this conclusion follows even in the presence of noise.

D. Lead field normalized minimum norm

Consider the lead field normalized minimum norm in the noise-free case. Rewriting the weight vector from Table 3, we have $E\left(\hat{j}^2_{\text{MN}}\right) = \tilde{\mathbf{g}}^T\mathbf{\Gamma}^{-1}\mathbf{R}_v\mathbf{\Gamma}^{-1}\tilde{\mathbf{g}},\ \tilde{\mathbf{g}} = \left|\mathbf{g}\right|^{-1}\mathbf{g}$.

Then $\left.\frac{\partial \hat{j}^2}{\partial \tilde{\mathbf{g}}}\right|_{\tilde{\mathbf{g}}=\tilde{\mathbf{g}}_{\mathrm{r}_0}} = 2\mathbf{\Gamma}^{-1}\mathbf{R}_v\mathbf{\Gamma}^{-1}\tilde{\mathbf{g}}_{r_0} \neq \mathbf{0}$ unless $\tilde{\mathbf{g}}_{r_0} = \mathbf{0}$ which means that this estimator is unbiased.

E. Non-adaptive standardized

Now consider the non-adaptive standardized estimator (sLoreta)

$$E\left(\hat{j}^2_{\text{NA-St}}\right) = \frac{\mathbf{g}^T\Gamma^{-1}\mathbf{R}_v\Gamma^{-1}\mathbf{g}}{\mathbf{g}^T\Gamma^{-1}\mathbf{g}} \quad (A.17)$$

using the same statistical model as that described by equations (A.1)-(A.3). Then

$$\frac{\partial \hat{j}^2}{\partial \mathbf{g}} = \frac{2\left(\left(\mathbf{g}^T\boldsymbol{\Gamma}^{-1}\mathbf{g}\right)\mathbf{S}\mathbf{g} - \left(\mathbf{g}^T\mathbf{S}\mathbf{g}\right)\boldsymbol{\Gamma}^{-1}\mathbf{g}\right)}{\left(\mathbf{g}^T\boldsymbol{\Gamma}^{-1}\mathbf{g}\right)^2} \quad (A.18)$$

where $\mathbf{S} = \boldsymbol{\Gamma}^{-1}\mathbf{R}_v\boldsymbol{\Gamma}^{-1}$. It is sufficient that the numerator of (A.18) be **0** at the extremum, Since $\mathbf{R}_v = \mathbf{g}_{\mathbf{r}_0}\mathbf{g}_{\mathbf{r}_0}^T + \mathbf{I}$, the numerator term (up to a multiple) becomes $\mathbf{n} = \left(\mathbf{g}^T\boldsymbol{\Gamma}^{-1}\mathbf{g}\right)\boldsymbol{\Gamma}^{-1}\left(\mathbf{I} + \mathbf{g}_{\mathbf{r}_0}\mathbf{g}_{\mathbf{r}_0}^T\right)\boldsymbol{\Gamma}^{-1}\mathbf{g} - \left(\mathbf{g}^T\boldsymbol{\Gamma}^{-1}\left(\mathbf{I} + \mathbf{g}_{\mathbf{r}_0}\mathbf{g}_{\mathbf{r}_0}^T\right)\boldsymbol{\Gamma}^{-1}\mathbf{g}\right)\boldsymbol{\Gamma}^{-1}\mathbf{g}$

Rearranging terms, we get

$$\begin{aligned}&\left(\mathbf{g}^T\boldsymbol{\Gamma}^{-1}\mathbf{g}\right)\boldsymbol{\Gamma}^{-2}\mathbf{g} + \left(\mathbf{g}^T\boldsymbol{\Gamma}^{-1}\mathbf{g}\right)^2\boldsymbol{\Gamma}^{-1}\mathbf{g}_{\mathbf{r}_0}\\ &\quad \neq \left(\mathbf{g}^T\boldsymbol{\Gamma}^{-2}\mathbf{g}\right)\boldsymbol{\Gamma}^{-1}\mathbf{g} - \left(\mathbf{g}^T\boldsymbol{\Gamma}^{-1}\mathbf{g}_{\mathbf{r}_0}\right)^2\boldsymbol{\Gamma}^{-1}\mathbf{g}\end{aligned} \quad (A.19)$$

This confirms the result reported by Sekihara [6] that sLoreta is a biased source location estimator in the presence of noise.

If we consider the noise-free case, the situation changes, since now equation (A.19) becomes

$$\mathbf{n} = \left(\mathbf{g}_{\mathbf{r}_0}^T\boldsymbol{\Gamma}^{-1}\mathbf{g}_{\mathbf{r}_0}\right)^2\boldsymbol{\Gamma}^{-1}\mathbf{g}_{\mathbf{r}_0} - \left(\mathbf{g}_{\mathbf{r}_0}^T\boldsymbol{\Gamma}^{-1}\mathbf{g}_{\mathbf{r}_0}\right)^2\boldsymbol{\Gamma}^{-1}\mathbf{g}_{\mathbf{r}_0} = \mathbf{0} \quad (A.20)$$

Thus in the noise-free case, this estimator becomes unbiased.

F. *Adaptive standardized*

For the adaptive standardized estimator

$$E\left(\hat{j}^2_{\text{A-St}}\right) = \frac{\mathbf{g}^T\mathbf{R}_v^{-1}\mathbf{R}_v\mathbf{R}_v^{-1}\mathbf{g}}{\mathbf{g}^T\mathbf{R}_v^{-1}\mathbf{g}} = 1 \quad (A.21)$$

so this estimator has no location sensitivity.

REFERENCES

[1] C. M. Michel, M. M. Murray, G. Lantz, S. Gonzalez, L. Spinelli and R. Grave de Peralta, "EEG Source Imaging,", *Clinical Neurophysiology* **115**, pp 2195–2222, 2004.

[2] M.E. Spencer, R.M. Leahy, J.C. Mosher and P.S. Lewis, "Adaptive Filters for Monitoring Localized Brain Activity from Surface Potential Time Series,", in *Conf. Rec. 26th Annual Asilomar Conference on Signals, Systems, and Computers*, Nov. 1992, pp 156-161.

[3] S.E. Robinson and D.F. Rose, "Current source image estimation by spatially filtered MEG,", in *Biomagnetism:Clinical Aspects*, Amsterdam: Excerpta Medica, pp 761-765, 1992.

[4] B.D. Van Veen, W. van Drongelen, M. Yuchtman, and A. Suzuki. "Localization of Brain Electrical Activity via Linearly Constrained Minimum Variance Spatial Filtering," *IEEE Trans. BME* **44**(9) pp 867-880, 1997.

[5] K . Sekihara, S.S. Nagarajan, D. Poeppel, A. Marantz, and Y. Miyashita. "Reconstructing Spatio-Temporal Activities of Neural Sources Using an MEG Vector Beamformer Technique," *IEEE Trans. BME* **48**(7), pp 760-771, 2001.

[6] K. Sekihara, "An overview of beamformer approaches: formulations, underlying assumptions, and major causes of errors," in *Proc. 14th Intl. Cong. Biomag.*, 2004, p 113.

[7] J.C. Mosher, R.M. Leahy and P.S. Lewis, "EEG and MEG: Forward Solutions for Inverse Methods," *IEEE Trans. BME* **46**(3), pp 245-259, 1999.

[8] L.H. Loomis and S. Sternberg, Advanced Calculus, Reading, MA: Addison-Wesley, 1968, p 85.

[9] A. Tarantola, *Inverse Problem Theory*, Amsterdam, Elsevier, 1987, p526.

[10] R. Pascual-Marqui, C.M. Michel, and D. Lehmann, "Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain," *Intl. J. Psychophysiol*. **18**, pp 49-65, 1994.

[11] T.A. Crowley, C.D. Haupt, and D.B. Kynor, "A Weighting Matrix to Remove Depth Bias in the Linear Biomagnetic Inverse Problem with Applications to Cardiology," in *Proc. Tenth Intl. Conf. Biomag*, 1996, pp 197-200.

[12] G.E. Backus and J.F. Gilbert, "Numerical applications of a formalism for geophysical inverse problems." *Geophys. J. Roy. Astron. Soc*. **13,** pp 247-276, 1967.

[13] R. Grave de Peralta and S.L. Gonzalez Andino. "Backus and Gilbert Method for Vector Fields," *Human Brain Mapping* **7**, pp 161-165, 1999.

[14] T.J, Loredo and R.I. Epstein, "Analysing Gamma-Ray Burst Spectral Data," *Astrophysical J.* **336** pp 896-919, 1989.

[15] R. Grave de Peralta, O. Hauk, S.L. Gonzalez Andino, H. Vogt, and C. Michel, "Linear inverse solutions with optimal resolution kernels applied to the electromagnetic tomography," *Human Brain Mapping* **5**(6), pp 454-467, 1997.

[16] G. Borgiotti and L.J. Kaplan, "Superresolution of uncorrelated interference sources by using adaptive array technique," *IEEE Trans. Antennas Propagat.*, **AP-27**, pp 842-845, 1979.

[17] R.D. Pascual-Marqui,. "Standardized low resolution brain electromagnetic tomography (sLORETA): technical details," *Methods and Findings in Experimental and Clinical Pharmacology* **24D** pp 5-12, 2002.

[18] S.E. Robinson and J. Vrba, "Functional Neuroimaging by Synthetic Aperture Magnetometry (SAM)," in *Recent Advances in Biomagnetism*, Sendai JP, Tohoku Univ. Press, 1999, pp 302-305.

[19] A.M. Dale, A.K. Liu, B.R. Fischl, 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* **26,** pp 55-67, 2000.

[20] M. Scherg, P. Berg, D. Weckesser, E. Hoechstetter, H. Bornfleth, N. Ille, and T. Bast, "Multiple source analysis of MEG and EEG data: Tools for validation," in *Proc. 14th Intl. Cong. Biomag.*, 2004, p 112.

[21] J.C. Mosher and R.M. Leahy, Recursive MUSIC: A Framework for EEG and MEG Source Localization," *IEEE Trans. BME* **45**(11), pp 1342-1354, 1998.

[22] R.E. Greenblatt and M.E. Pflieger, "Randomization-based Hypothesis Testing form Event-related data", *Brain Topography* **16**(4), pp 225-232, 2004.

[23] S. Haykin, *Adaptive Filter Theory*, Englewood Cliffs, NJ: Prentice Hall, 1991, p 480.

[24] K.V. Mardia, J.T. Kent, and J.M. Bibby, *Multivariate Analysis*. London: Academic Press, 1979, p62.

[25] M.E. Pflieger, R.E. Greenblatt and J Kirkish, "Regional Resolving Power of Combined MEG/EEG", in *Proc. 14th Intl. Cong. Biomag.*, 2004, pp 673-674.

Method Type	*Class*		*Weight vector* ($\mathbf{w}^T$)	Γ	$\mathbf{V} \to \mathbf{V}^*$	$\mathbf{Q} \to \mathbf{Q}^*$	*Location unbiased*	*Relative magnitude unbiased*
Global (pseudo-inverse)	*Minimum L2-norm*		$\mathbf{g}_r^T \boldsymbol{\Gamma}^{-1}$	$\mathbf{G}\mathbf{G}^T$	$\left(\mathbf{G}\mathbf{G}^T\right)^{-1}$	$\mathbf{I}$		
	Minimum B-norm		$\mathbf{b}_r^{-1} \mathbf{g}_r^T \boldsymbol{\Gamma}^{-1}$	$\mathbf{G}\mathbf{B}^{-1}\mathbf{G}^T$	$\left(\mathbf{G}\mathbf{B}^{-1}\mathbf{G}^T\right)^{-1}$	$\mathbf{B}$		
Local	*Non-adaptive*	*Distortion-less*	$\frac{\mathbf{g}_r^T \boldsymbol{\Gamma}^{-1}}{\mathbf{g}_r^T \boldsymbol{\Gamma}^{-1} \mathbf{g}_r}$	$\mathbf{G}\mathbf{G}^T$	$\left(\mathbf{G}\mathbf{G}^T\right)^{-1}$	$\operatorname{diag}\left[\mathbf{G}^T \boldsymbol{\Gamma}^{-1} \mathbf{G}\right]$		✓✓
		weight vector normalized	$\frac{\mathbf{g}_r^T \boldsymbol{\Gamma}^{-1}}{\left(\mathbf{g}_r^T \boldsymbol{\Gamma}^{-2} \mathbf{g}_r\right)^{1/2}}$	$\mathbf{G}\mathbf{G}^T$	$\left(\mathbf{G}\mathbf{G}^T\right)^{-1}$	$\operatorname{diag}\left[\left(\mathbf{G}^T \boldsymbol{\Gamma}^{-2} \mathbf{G}\right)^{1/2}\right]$		
		Standard-ized	$\frac{\mathbf{g}_r^T \boldsymbol{\Gamma}^{-1}}{\left(\mathbf{g}_r^T \boldsymbol{\Gamma}^{-1} \mathbf{g}_r\right)^{1/2}}$	$\mathbf{G}\mathbf{G}^T$	$\left(\mathbf{G}\mathbf{G}^T\right)^{-1}$	$\operatorname{diag}\left[\left(\mathbf{G}^T \boldsymbol{\Gamma}^{-1} \mathbf{G}\right)^{1/2}\right]$	✓	
	Adaptive	*Distortion-less*	$\frac{\mathbf{g}_r^T \boldsymbol{\Gamma}^{-1}}{\mathbf{g}_r^T \boldsymbol{\Gamma}^{-1} \mathbf{g}_r}$	$\mathbf{R}$	$\mathbf{R}^{-1}$	$\operatorname{diag}\left[\mathbf{G}^T \boldsymbol{\Gamma}^{-1} \mathbf{G}\right]$		✓✓
		weight vector normalized	$\frac{\mathbf{g}_r^T \boldsymbol{\Gamma}^{-1}}{\left(\mathbf{g}_r^T \boldsymbol{\Gamma}^{-2} \mathbf{g}_r\right)^{1/2}}$	$\mathbf{R}$	$\mathbf{R}^{-1}$	$\operatorname{diag}\left[\left(\mathbf{G}^T \boldsymbol{\Gamma}^{-2} \mathbf{G}\right)^{1/2}\right]$	✓✓	
		Standard-ized	$\frac{\mathbf{g}_r^T \boldsymbol{\Gamma}^{-1}}{\left(\mathbf{g}_r^T \boldsymbol{\Gamma}^{-1} \mathbf{g}_r\right)^{1/2}}$	$\mathbf{R}$	$\mathbf{R}^{-1}$	$\operatorname{diag}\left[\left(\mathbf{G}^T \boldsymbol{\Gamma}^{-1} \mathbf{G}\right)^{1/2}\right]$		

Table 3. The linear estimators considered in this paper may be global or local, and the local estimators may be non-adaptive ($\Gamma = \mathbf{G}\mathbf{G}^T$) or adaptive($\Gamma = \mathbf{R}$). Each weight vector class may be seen to constitute a vector field over the source volume, Ω. G is the gain matrix, and R is the estimated signal space covariance. In addition, local estimators may be distortionless, weight vector normalized, or standardized. Each valid combination leads to a different weight vector defining equation, or equivalently, the choice of signal space and source space metrics (shown in the last two columns of the table, respectively). Two checkmarks in the cells of the last two column show the estimators derived from these weight vectors are unbiased with respect to location and relative magnitude respectively in the presence of noise, while one checkmark indicates that the estimator is only unbiased in the noise-free case (described in III.G).