USC–SIPI REPORT #304

Recursively Applied MUSIC:
A Framework for
EEG and MEG Source Localization

by

John C. Mosher and Richard M. Leahy

October 1996

Signal and Image Processing Institute
UNIVERSITY OF SOUTHERN CALIFORNIA
Department of Electrical Engineering-Systems
3740 McClintock Avenue, Room 404
Los Angeles, CA 90089-2564 U.S.A.

Abstract

The multiple signal characterization (MUSIC) algorithm locates multiple asynchronous dipolar sources from electroencephalography (EEG) and magnetoencephalography (MEG) data. A signal subspace is estimated from the data, then the algorithm scans a single dipole model through a three-dimensional head volume and computes projections onto this subspace. To locate the sources, the user must search the head volume for local peaks in the projection metric. This task is time consuming and subjective. Here we describe an extension of this approach which we refer to as RAP (Recursively APplied) MUSIC. This new procedure automatically extracts the locations of the sources through a recursive use of subspace projections. The new method is also able to deal with synchronous sources through the use of a spatially independent topographies (SPIT) model. This model defines a source as one or more non-rotating dipoles with a single time course. Within this framework, we are able to locate fixed, rotating and synchronous dipoles. The recursive subspace projection procedure that we introduce here uses the metric of subspace angles as a multi-dimensional form of correlation analysis between the model subspace and the data subspace. By using subspace angle computations, we recursively build up a model for the sources that account for a given set of data. We demonstrate here how RAP-MUSIC can easily extract multiple asynchronous dipolar sources that are difficult to find using the original MUSIC scan. We then demonstrate RAP-MUSIC applied to the more general SPIT model and show results for combinations of fixed, rotating, and synchronous dipoles.

1.0 Introduction

The problem of localizing the sources of event related scalp potentials (the electroencephalogram or EEG) and magnetic fields (the magnetoencephalogram or MEG) can be formulated in terms of finding a least squares fit of a set of current dipoles to the observed data. Early attempts at source localization were based on fitting the multiple dipole model to a single time sample of the measurements across the E/MEG (EEG and/or MEG) array [4], [19], [28]. By noting that physiological models for the current sources typically assume that they are spatially fixed for the duration of a particular response, researchers were able to justify fitting the multiple dipole model to a complete spatio-temporal data set [2],[16], [17]. The spatio-temporal model can result in substantial improvements in localization accuracy. Processing the entire data set leads to a large increase in the number of unknown parameters since the times series for each source must now be estimated in addition to the dipole location and orientation. However, in [10] we show that since these time series parameters are linear with respect to the data, they can be factored out and the source locations found without explicit computation of their associated time series.

While factoring out the linear parameters can reduce the dimensionality of the search required to localize the sources of the measured fields, the fundamental problem remains that the cost function is non-convex with respect to the locations of the dipoles. Consequently inverse methods based on direct minimization of the squared error through gradient-based optimization or simplex searches often lead to improper locations of the sources due to trapping in local minima. In an attempt to overcome this problem, we examined the use of signal subspace methods that are common in the array signal processing literature [7]. The method that we used in [10], which was originally referred to as the MUSIC (for MUltiple SIgnal Characterization) algorithm in [18], replaces the multiple dipole directed search with a procedure in which a single dipole is scanned through a grid

confined to a three dimensional head or source volume. At each point on this grid, the forward model for a dipole at this location is projected against a signal subspace that has been computed from the E/MEG data. The locations on this grid where the source model gives the best projection onto the signal subspace correspond to the dipole locations. We also show in [10] that we do not need to test all possible dipole orientations at each location, but instead can solve a generalized eigenvalue problem whose solution gives us the orientation of the dipole which gives the best fit to the signal subspace for a source at that location.

One of the major problems with the MUSIC method, and one that is addressed by the new approach described here, is how we choose the locations which give the best projection on to the signal subspace. In the absence of noise and with perfect head and sensor models, the forward model for a source at the correct location will project entirely into the signal subspace. In practice, of course, there are errors in the estimate of the signal subspace due to noise, and errors in the forward model due to approximations in our models of the head and data acquisition system. An additional problem is that we compute the metric only at a finite set of grid points. The effect of these practical limitations is that the user is faced with the problem of searching the gridded source volume for "peaks" and deciding which of these peaks correspond to true locations. It is important to note that a local peak in this metric does not necessarily indicate the location of a source. Only when the forward model projects entirely into the signal subspace – or as close as one would expect given errors due to noise and model mismatch – can we infer that a source is at that location. The effect of this limitation is that some degree of subjective interpretation of the MUSIC "scan" is required to decide on the locations of the sources. This subjective interpretation is clearly undesirable and can also lead to the temptation to incorrectly view the MUSIC scan as an image whose intensity is proportional to the probability of a source being present at each location.

Two other problems that arise with the use of MUSIC are due to the fact that the computation of the signal subspace from the data is based on the assumption that the data are produced by a set of asynchronous dipolar sources and are corrupted by additive spatially-white noise. Often both of these assumptions are incorrect in clinical or experimental data. If two dipoles have synchronous activation, then the two-dimensional signal subspace that would have been produced if they were asynchronous collapses into a one dimensional subspace. Scanning of a single dipole against this subspace using MUSIC will fail to localize either of the sources. The RAP MUSIC algorithm described here is able to localize synchronous sources through the use of a modified source representation which we refer to as the spatially independent topographies (SPIT) model. This model is described in detail in Section 3. The second problem, the issue of nonwhite noise, is not addressed in depth here. However, we note that it is straightforward to modify both the original and RAP MUSIC algorithms to cope with colored noise through a standard pre-whitening procedure [20]. In practice the pre-whitening could be achieved by using pre-stimulus data to estimate the covariance of the background noise.

The MUSIC localization technique uses the SVD (singular value decomposition) of the data mtrix or eigenanalysis of the data correlation to generate an estimate of the "signal subspace." MUSIC was preceded in the E/MEG literature by dipole fitting using principal components analysis (PCA) with techniques such as the Varimax rotation [3], [8], [9], [29]. While the initial PCA used in these methods is essentially identical to the SVD approach of MUSIC, the subsequent processing methods are very different. Fundamental to most of the earlier work on PCA was the assumption that

there was a one-to-one correspondence between each principal component and a particular source. In general this is not correct, as discussed in [3], [8], [9], [29]. In contrast, the MUSIC method is based on the far less restrictive observation that the signal subspace and that generated by the correct combination of sources should coincide.

We begin the paper with a combined formulation of the E/MEG forward problem in which we develop a standard matrix notation for the relationship between the source and data. In Section 3 we then describe the spatially independent topographies model in which, rather than treating individual current dipoles as sources, we define a source as one or more non-rotating dipoles with a single time course. In this way our model is constrained to consist of a number of sources equal to the rank of the signal subspace. In Section 4 we review the definition and properties of the signal subspace. In Section 5, we introduce the use of subspace correlations as a metric for computing the goodness of fit of putative sources to the signal subspace. We then review the MUSIC algorithm in Section 6 in the light of the preceding development. The new RAP MUSIC algorithm is then developed in Section 7. We present some examples of the application of RAP MUSIC to fixed, rotating and synchronous sources in Section 8. Finally, in Section 9, we discuss the properties, limitations and some extensions of the RAP MUSIC approach.

2.0 Background

Quasi-static approximations of Maxwell's equations govern the relationship between neural current sources and the E/MEG data that they produce. For the signal subspace methods for source localization that are described here, these relationships must be expressed in matrix form. In [14] we examined explicit forms of the "lead field" for EEG and MEG measurements, for both spherical and general BEM head models. In each case, the measurements can be expressed as an explicit function of *primary current* activity; the passive volume currents are implicitly embedded in the lead field formula. The model should also account for the sensor characteristics of the measurement modality, such as gradiometer orientation and configuration in MEG or differential pairs in EEG. We show in [14] that these effects can be incorporated into simple transformations that modify the basic lead field kernels. The result is that our EEG or MEG measurement $f_m(\boldsymbol{r})$ at sensor location $\boldsymbol{r}$ may be expressed as

$$f_m(\boldsymbol{r}) = \int_V \boldsymbol{g}(\boldsymbol{r}, \boldsymbol{r}') \cdot \boldsymbol{j}(\boldsymbol{r}') d\boldsymbol{r}' \qquad (1)$$

where V is the volume of sources, $\boldsymbol{j}(\boldsymbol{r}')$ represents the *primary current density* at any point $\boldsymbol{r}'$ in the volume, and $\boldsymbol{g}(\boldsymbol{r}, \boldsymbol{r}')$ is commonly known as the "lead field vector" (cf. [24]). The scalar function $f_m(\boldsymbol{r})$ represents either the voltage potential or the magnetic field component that may be observed at observation (sensor) point $\boldsymbol{r}$.

If we assume that the primary current exists only at a discrete point $\boldsymbol{r}_q$, i.e., the primary current is $\boldsymbol{j}(\boldsymbol{r}')\delta(\boldsymbol{r}' - \boldsymbol{r}_q)$, where $\delta(\boldsymbol{r}' - \boldsymbol{r}_q)$ is the Dirac delta function, then *(1)* simplifies in E/MEG to

$$f_m(\boldsymbol{r}) = \boldsymbol{g}(\boldsymbol{r}, \boldsymbol{r}_q) \cdot \boldsymbol{q} \qquad (2)$$

where $\boldsymbol{q}$ is the *moment* of a *current dipole* at $\boldsymbol{r}_q$. We assume here that our source consists of p current dipole sources. We assume simultaneous recordings at m sensors for n time instances. We can express the m - by n spatio-temporal data matrix as

$$\begin{bmatrix} f_m(\boldsymbol{r}_1, t_1) & \dots & f_m(\boldsymbol{r}_1, t_n) \\ \dots & \dots & \dots \\ f_m(\boldsymbol{r}_m, t_1) & \dots & f_m(\boldsymbol{r}_m, t_n) \end{bmatrix} = \begin{bmatrix} \boldsymbol{g}(\boldsymbol{r}_1, \boldsymbol{r}_{q1})^T & \dots & \boldsymbol{g}(\boldsymbol{r}_1, \boldsymbol{r}_{qp})^T \\ \dots & \dots & \dots \\ \boldsymbol{g}(\boldsymbol{r}_m, \boldsymbol{r}_{q1})^T & \dots & \boldsymbol{g}(\boldsymbol{r}_m, \boldsymbol{r}_{qp})^T \end{bmatrix} \begin{bmatrix} \boldsymbol{q}_1(t_1) & \dots & \boldsymbol{q}_1(t_n) \\ \dots & \dots & \dots \\ \boldsymbol{q}_p(t_1) & \dots & \boldsymbol{q}_p(t_n) \end{bmatrix}. \quad (3)$$

or

$$\boldsymbol{F}_m = \left[\boldsymbol{G}(\boldsymbol{r}_{q1}) \dots \boldsymbol{G}(\boldsymbol{r}_{qp})\right]\boldsymbol{Q}^T. \quad (4)$$

We refer to $\boldsymbol{G}(\boldsymbol{r}_{qi})$ as the dipole "gain matrix" [10] that maps a dipole at $\boldsymbol{r}_{qi}$ into a set of measurements. The three columns of the gain matrix represent the possible *forward fields* that may be generated by the three orthogonal orientations of the i th dipole at the m sensor locations $\{\boldsymbol{r}_1, \dots, \boldsymbol{r}_m\}$. Each row of the full gain matrix $\left[\boldsymbol{G}(\boldsymbol{r}_{q1}) \dots \boldsymbol{G}(\boldsymbol{r}_{qp})\right]$ represents the *lead field,* sampled at the discrete dipole locations $\{\boldsymbol{r}_{q1}, \dots, \boldsymbol{r}_{qp}\}$. The matrix $\boldsymbol{F}_m$ is our *model matrix* of perfect measured data, i.e., the magnetic field component or scalp potential data we would observe, in the absence of noise, given our model.

The columns of $\boldsymbol{Q}$ represent the time series associated with each of the three orthogonal components of each dipole, i.e., with each column of the gain matrix. For the "fixed" dipole model, whose moment orientation is time invariant, we can separate the orientation of each source from the moments [10] as:

$$\boldsymbol{F}_m = \left[\boldsymbol{G}(\boldsymbol{r}_{q1}) \dots \boldsymbol{G}(\boldsymbol{r}_{qp})\right] \begin{bmatrix} \boldsymbol{u}_{q1} & & 0 \\ & \dots & \\ 0 & & \boldsymbol{u}_{qp} \end{bmatrix} \begin{bmatrix} s_{q1}(t_1) & \dots & s_{q1}(t_n) \\ \dots & \dots & \dots \\ s_{qp}(t_1) & \dots & s_{qp}(t_n) \end{bmatrix} \quad (5)$$

such that $\boldsymbol{q}_i(t_i) = \boldsymbol{u}_{qi} s_{qi}(t_i)$, where $\boldsymbol{u}_{qi}$ is a unit norm orientation vector. The scalar time series $s_{qi}(t_i)$ are the *linear* parameters of our model, the corresponding dipole locations $\boldsymbol{r}_{qi}$ are the *nonlinear* parameters, and the dipole orientations $\boldsymbol{u}_{qi}$ are the *quasi-linear* parameters.

The data model defined by *(5)* assumes a collection of p fixed dipoles. In [10], we also considered a "rotating" dipole as one whose time series could not be decomposed into a single fixed orientation and time series. Physically, a rotating dipole may be viewed as two nearly collocated dipoles with independent time series, such that they are indistinguishable from a model comprising a single dipole whose orientation is allowed to vary with time. The "hybrid" models in [10] comprised both fixed and rotating dipoles. In the following section we will return to the fixed dipole model as the basic element of our source model. From this we develop a representation for the data that we will refer to as the spatially independent topographies model.

3.0 Spatially Independent Topographies

Our goal in this section is build up a model for the data as the sum of contributions from a fixed number of spatially independent topographies. Associated with each of these topographies is a source consisting of one or more fixed dipoles which have a single time course. By building up the model in this way we are guaranteed that the rank of the signal subspace that we compute in Section 4 is equal to the number of these sources. This representation then provides a convenient framework for describing and implementing the RAP MUSIC algorithm that is the main contribution of this paper.

3.1 The Single Dipole Case

In the following we adopt a fixed-dipole only model, with rotating dipoles modeled as two collocated fixed dipoles. We elaborate on this point, since its construction generalizes to the "spatially independent topographies" model we will introduce. Consider a single dipole model and its corresponding time series, expanded in Cartesian coordinates to reveal the "elemental" (cf. [10]) dipoles in directions x, y, and z,

$$G(r_{q1})Q_1^T = \left[g_x(r_{q1})\ g_y(r_{q1})\ g_z(r_{q1})\right]\begin{bmatrix} q_x(t_1) & \dots & q_x(t_n) \\ q_y(t_1) & \dots & q_y(t_n) \\ q_z(t_1) & \dots & q_z(t_n) \end{bmatrix}. \tag{6}$$

Our construction hinges on the dependence of the elemental time series, or equivalently, on the rank of the inner product matrix,

$$R_1 = Q_1^T Q_1 = \Phi_1 \Lambda_1 \Phi_1^T \tag{7}$$

where Φ_1 is the matrix of eigenvectors and Λ_1 is the diagonal matrix of eigenvalues. Equivalently, we may perform a singular value decomposition of Q_1^T,

$$Q_1^T = U_1 \Sigma_1 V_1^T \tag{8}$$

where U_1 and V_1 are the left and right singular vectors, respectively, and Σ_1 is a diagonal matrix of singular values. If the dipole is fixed in orientation, then Λ_1 (Σ_1) comprises only a single nonzero eigenvalue (singular value) with a corresponding eigenvector (left singular vector). Equivalently, we say R_1 (Q_1) is of rank one. In this instance, we may designate this eigenvector or left singular vector as the dipole orientation vector u_{q1}, and we use the corresponding right singular vector and singular value to create a time series $s_{q1} = v_{q1}\sigma_{q1}$. Our one dipole model then becomes

$$G(r_{q1})Q_1^T = \left[g_x(r_{q1})\ g_y(r_{q1})\ g_z(r_{q1})\right]u_{q1}\sigma_{q1}v_{q1}^T \tag{9}$$

$$= [G(r_{q1})u_{q1}]s_{q1}^T \tag{10}$$

$$= a(r_{q1}, u_{q1}) s_{q1}^T \tag{11}$$

where, in general for the i th dipole,

$$a(r_{qi}, u_{qi}) \equiv G(r_{qi}) u_{qi}. \tag{12}$$

We refer to this column vector as a "single dipolar topography," since it represents the spatial pattern of a single fixed orientation dipole across the m sensors.

If Q_1 is rank two, with corresponding left singular vectors u_{q1} and u_{q2}, then we may specify our one dipole model as

$$G(r_{q1}) Q_1^T = \left[g_x(r_{q1}) \; g_y(r_{q1}) \; g_z(r_{q1}) \right] [u_{q1}, u_{q2}] \begin{bmatrix} \sigma_{q1} & 0 \\ 0 & \sigma_{q2} \end{bmatrix} [v_{q1}, v_{q2}]^T \tag{13}$$

In [10], we designated this form as a "rotating" dipole. Here, we find it more convenient to continue to use our single dipolar topography model by rewriting *(13)* as the sum of two single dipolar topographies,

$$G(r_{q1}) Q_1^T = a(r_{q1}, u_{q1}) s_{q1}^T + a(r_{q1}, u_{q2}) s_{q2}^T \tag{14}$$

where $s_{q2} = v_{q2} \sigma_{q2}$. This combines the same gain matrix $G(r_{q1})$ with two different orientations u_{q1} and u_{q2} to form two different single dipolar topographies $a(r_{q1}, u_{q1})$ and $a(r_{q1}, u_{q2})$. The case for Q_1 of rank three follows similarly. We note that the specific orientations in the rotating dipole model are arbitrary and an infinite number of expansions of the form in *(14)* are possible.

This combination of single dipolar topographies can be rewritten by expressing *(14)* in matrix form,

$$G(r_{q1}) Q_1^T = [a(r_{q1}, u_{q1}), a(r_{q2}, u_{q2})] [s_{q1}, s_{q2}]^T \tag{15}$$

$$= A(\rho, \theta) S^T \tag{16}$$

where $\rho \equiv \{r_{q1}, r_{q2}\}$ is the set of dipole locations, $\theta \equiv \{u_{q1}, u_{q2}\}$ is the set of dipole orientations, and $A(\rho, \theta) \equiv [a(r_{q1}, u_{q1}), a(r_{q2}, u_{q2})]$ is the *spatially independent topographies (SPIT)* matrix formed from the gain matrix $G(\rho)$ in the appropriate linear combinations determined by θ. In this rotating dipole example, $r_{q1} = r_{q2}$, i.e., the two dipoles in $A(\rho, \theta)$ are collocated. Their corresponding time series are the columns of S. *Our requirement in the SPIT model is that* $A(\rho, \theta)$ *and* S *are each of full column rank, and that their product* $A(\rho, \theta) S^T$ *is of the same rank.*

3.2 The Multiple Dipole Case

For the case of multiple dipoles, we enforce the full column rank requirement by examining the "outer product" of our full spatio-temporal model matrix,

$$\boldsymbol{R}_m \equiv \boldsymbol{F}_m \boldsymbol{F}_m^T = \boldsymbol{G}(\rho)\boldsymbol{Q}^T \boldsymbol{Q}\boldsymbol{G}(\rho)^T \tag{17}$$

We perform an eigenanalysis of $\boldsymbol{R}_m$ to determine the *rank* of the model, or equivalently, we perform an SVD of $\boldsymbol{G}(\rho)\boldsymbol{Q}$ to determine the number of nonzero singular values. Assuming that the dipoles are well-separated, the rank of $\boldsymbol{R}_m$ will be less than or equal to $3p$, and the determining factor will be the rank of $\boldsymbol{Q}$. If each of the dipoles is fixed in orientation, and these fixed dipole time series are linearly independent, then the rank of $\boldsymbol{Q}$ and equivalently $\boldsymbol{R}_m$ will be at most p. In this instance, a model of p single fixed dipole topographies is appropriate,

$$\boldsymbol{R}_m = \boldsymbol{G}(\rho)\boldsymbol{Q}\boldsymbol{Q}^T \boldsymbol{G}(\rho)^T = \boldsymbol{A}(\rho, \theta)\boldsymbol{S}^T \boldsymbol{S}\boldsymbol{A}(\rho, \theta)^T \tag{18}$$

where $\rho \equiv \{\boldsymbol{r}_{q1}, \ldots, \boldsymbol{r}_{qp}\}$ and $\theta = \{\boldsymbol{u}_{q1}, \ldots, \boldsymbol{u}_{qp}\}$, the set of dipole locations and orientations, respectively, have been generalized to the case of p dipoles, and each column of $\boldsymbol{A}(\rho, \theta)$ is formed as in *(12)*. If any of the dipoles are rotating, we simply add that dipole's location and its other orientation vector or vectors into the sets ρ and θ, i.e. we treat each of the additional independent orientations as an additional fixed dipole at the same location contributing its own independent topography. For simplicity, we will assume below that all dipoles are fixed in orientation.

Next we consider the case when the first two out of p dipoles are synchronous in their time series. This synchronicity may occur in bilateral response studies, for example. We will continue to assume that the other $p-2$ dipoles have time series independent of the first two. In this instance, $\boldsymbol{R}_m$ is of rank $p-1$, and we alter our SPIT model by considering the rank-one decomposition of the combined dipolar time series,

$$\begin{bmatrix} \boldsymbol{q}_1(t_1) & \ldots & \boldsymbol{q}_1(t_n) \\ \boldsymbol{q}_2(t_1) & \ldots & \boldsymbol{q}_2(t_n) \end{bmatrix} = \boldsymbol{u}_1 \sigma_1 \boldsymbol{v}_1^T \tag{19}$$

Since we have assumed these two dipoles are synchronous, we have only one nonzero singular value in the decomposition. We then alter our matrix of single dipolar topographies to be

$$\boldsymbol{F}_m = \begin{bmatrix} \boldsymbol{a}(\rho_1, \boldsymbol{u}_1) & \boldsymbol{a}(\boldsymbol{r}_{q3}, \boldsymbol{u}_{q3}) & \ldots \end{bmatrix} \begin{bmatrix} \boldsymbol{s}_1^T \\ \boldsymbol{s}_{q3}^T \\ \ldots \end{bmatrix} \tag{20}$$

where the set $\rho_1 \equiv \{\boldsymbol{r}_{q1}, \boldsymbol{r}_{q2}\}$ comprises the two synchronous dipoles, and the gain matrices for these two dipoles have been combined into a single column vector $\boldsymbol{a}(\rho_1, \boldsymbol{u}_1) \equiv [\boldsymbol{G}(\boldsymbol{r}_{q1}), \boldsymbol{G}(\boldsymbol{r}_{q2})]\boldsymbol{u}_1$.

This process combines the first two fixed dipoles into a single "2-dipolar topography," with its corresponding time series $\boldsymbol{s}_1 \equiv \boldsymbol{v}_1 \sigma_1$ from the decomposition in *(19)*. The vector $\boldsymbol{u}_1$ is now dimension six with unit norm. The other columns of the topographies matrix remain single dipolar topographies, and the number of columns in our topographies matrix $\boldsymbol{A}(\rho, \theta)$ and our time series matrix

$\boldsymbol{S}$ is now $p-1$. By design, both matrices are now of full column rank $p-1$, as is their product, thereby maintaining the SPIT model requirement.

The generalization to arbitrary topographies follows. We cluster the p dipoles into r subsets and redefine *(4)* as

$$\boldsymbol{F}_m = \boldsymbol{G}(\rho)\boldsymbol{Q} \tag{21}$$

$$= \left[\boldsymbol{a}(\rho_1, \boldsymbol{u}_1) \ \ldots \ \boldsymbol{a}(\rho_r, \boldsymbol{u}_r)\right] \begin{bmatrix} \boldsymbol{s}_1^T \\ \ldots \\ \boldsymbol{s}_r^T \end{bmatrix} \tag{22}$$

$$= \boldsymbol{A}(\rho, \theta)\boldsymbol{S}^T \tag{23}$$

where $\rho \equiv \{\rho_1, \ldots, \rho_r\}$ represents r clusters of dipoles, with the ith cluster comprising p_i dipoles with the location parameter set $\rho_i \equiv \{\boldsymbol{r}_{q1}, \ldots, \boldsymbol{r}_{qp_i}\}$. The set $\theta \equiv \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r\}$ contains the corresponding unit norm vectors that correspond to the left singular vectors found as the extension of *(19)* to p_i dipoles. The ith column of $\boldsymbol{A}(\rho, \theta)$ is found by extending *(12)*:

$$\boldsymbol{a}(\rho_i, \boldsymbol{u}_i) \equiv \left[\boldsymbol{G}(\boldsymbol{r}_{q1}) \ \ldots \ \boldsymbol{G}(\boldsymbol{r}_{qp_i})\right]\boldsymbol{u}_i \tag{24}$$

We refer to this column vector as a "p_i-dipolar topography." Thus each column of $\boldsymbol{A}(\rho, \theta)$ is now a p_i-dipolar topography, with a corresponding time series found as the ith column of $\boldsymbol{S}$. By construction, the time series in $\boldsymbol{S}$ are also independent. Note that in the SPIT model, the concept of dipole orientation is generalized to represent the orientation of all dipoles in a given p_i-dipolar topography. The vector $\boldsymbol{u}_i$ is a single unit norm orientation parameter for the ith topography of dimension $3p_i$. This orientation parameter defines the particular weightings necessary to combine all of the dipoles together into a single gain vector.

In this general multiple dipole model we assume that each independent topography comprises one or more well-separated dipoles. The case where dipoles are so spatially close that their corresponding topography is ambiguous with other simpler spatial models will be addressed in a future paper. Implicit in the requirement that $\boldsymbol{A}(\rho, \theta)$ is of full column rank is that the number of sensors (number of rows) sufficiently exceeds the number of dipoles, such that the SPIT matrix remains well-conditioned. In fact, the number of sensors should well exceed this number for the MUSIC methods described below.

We will conclude this section with some examples. Three asynchronous dipoles yield a rank three model, and hence our SPIT model comprises three single dipolar topographies. If two of these dipoles are synchronous, but the third remains asynchronous from the others, then the rank of the model is two, and our SPIT model comprises one single dipolar topography and one 2-dipolar topography. If all three are synchronous, then our SPIT model comprises a single 3-dipolar topography. A single time slice of a p dipole model is always rank one, and our SPIT model is therefore a single p-dipolar topography. For the MEG spherical head model, a single "rotating" dipole

becomes two fixed orientation dipoles corresponding to two single dipolar topographies, with $\rho_1 = \rho_2 = r_{q1}$; for other head models, the dipole may possibly rotate in three dimensions, $\rho_1 = \rho_2 = \rho_3 = r_{q1}$. Note that the $m \times 2$ (or $m \times 3$) SPIT matrix $\left[a(\rho_1, u_1)\ a(\rho_2, u_2)\right]$ (or $\left[a(\rho_1, u_1)\ a(\rho_2, u_2)\ a(\rho_3, u_3)\right]$) properly remains of full column rank.

To summarize, all "sources" in a SPIT model must be mutually spatially and temporally independent. Under this requirement, the *rank* of the noiseless spatio-temporal data matrix determines the number of independent topographies. Each topography in turn may comprise multiple dipoles with synchronous time courses. The SPIT model can handle arbitrary combinations of synchronous and asynchronous, and fixed and rotating dipoles.

4.0 Signal Subspaces

We will now investigate the relationship between the SPIT model and the signal subspace that we estimate from the data. Assume that a random noise matrix $N \equiv \left[n_1 \ \ldots \ n_n\right]$ is added to the data $F_m = A(\rho, \theta)S^T$, to produce an $m \times n$ "noisy" spatio-temporal data set,

$$F = A(\rho, \theta)S^T + N \tag{25}$$

For convenience we will occasionally drop the explicit dependence of A on its parameters. The goal of the inverse problem is to estimate the parameters, $\{\rho, \theta, S\}$, given the data set F. We will use the common assumption that the noise is zero-mean and white, i.e.:

$$E\{n_i\} = 0,\ E\{n_i n_i^T\} = \sigma_n^2 I,\ and\ E\{n_i n_j^T\} = 0,\ i \neq j. \tag{26}$$

where $E\{\bullet\}$ denotes the expected value of the argument, and I is the identity matrix. The case for colored noise is readily treated with standard pre-whitening methods [20], provided a reasonable estimate of the noise covariance is available. For event related studies the noise covariance can probably be estimated using sufficiently long periods of pre-stimulus data.

In *signal subspace* processing, we focus on the matrix outer product FF^T. Under the white noise assumption

$$R_F \equiv E\{FF^T\} = AS^TSA^T + \sum_{i=1}^{n} E(n(t_i)n^T(t_i)) \tag{27}$$

$$= AS^TSA^T + n\sigma_N^2 I \tag{28}$$

Here we have assumed that our model parameters and the dipolar time series are deterministic. From our independent topographies model, we know that AS^TSA^T is rank r. We designate the singular value decomposition of A as $A = U_A \Sigma_A V_A^T$, where U_A retains only the r left singular vectors associated with the r nonzero singular values. The outer product AS^TSA^T may be factored in terms of the SVD of A:

$$R_s \equiv AS^TSA^T = U_A(\Sigma_A V_A^T S^T S V_A \Sigma_A)U_A^T \tag{29}$$

and also using the eigendecomposition:

$$\boldsymbol{R}_s = \Phi_s \Lambda_s \Phi_s^T \tag{30}$$

where Φ_s contain the $m \times r$ eigenvectors of $\boldsymbol{R}_s$ corresponding to its non zero eigenvalues and Λ_s is the corresponding $r \times r$ diagonal matrix of nonzero eigenvalues. The remaining $m - r$ eigenvectors are contained in $\Phi_s^{\perp}$. The notation $\Phi_s^{\perp}$ reflects the property that its columns are orthogonal to those of Φ_s. We can therefore decompose the correlation matrix $\boldsymbol{R}_F$ of the noisy data by combining *(28)* and *(30)* as:

$$\boldsymbol{R}_F = [\Phi_s, \Phi_s^{\perp}] \begin{bmatrix} \Lambda_s + n\sigma_N^2 & 0 \\ 0 & n\sigma_N^2 \end{bmatrix} [\Phi_s, \Phi_s^{\perp}]^T. \tag{31}$$

$$= \Phi_s \Lambda_F \Phi_s^T + \Phi_s^{\perp} \Lambda_n \Phi_s^{\perp T} \tag{32}$$

where $\Lambda_F \equiv \Lambda_s + n\sigma_N^2 \boldsymbol{I}$ is the $r \times r$ diagonal matrix combining both the model and noise eigenvalues, and $\Lambda_n \equiv n\sigma_N^2 \boldsymbol{I}$ is the $(m-r) \times (m-r)$ diagonal matrix of noise only eigenvalues. Therefore an eigendecomposition of $\boldsymbol{R}_F$ yields r eigenvalues greater than $n\sigma_N^2$.

By eigendecomposing the term in parentheses in *(29)* into its $r \times r$ eigenvector and eigenvalue matrices, it is straightforward to show that $\boldsymbol{U}_A \boldsymbol{T}_{\Phi A} = \Phi_s$, where $\boldsymbol{T}_{\Phi A}$ is a full rank $r \times r$ transformation matrix relating the two. The r dimensional subspace spanned by Φ_s is therefore the same as the r dimensional subspace spanned by $\boldsymbol{U}_A$. We refer to this space as the *signal subspace*. The remaining $m - r$ vectors in $\Phi_s^{\perp}$ span the *orthogonal signal subspace*, or *noise-only subspace*, which is also often, but somewhat erroneously, referred to as the *noise subspace*.

In summary, the spatio-temporal matrix of noisy data has an expected correlation with precisely r eigenvalues that exceed $n\sigma_N^2$. The corresponding eigenvectors span the signal subspace which is identical to the space spanned by the columns of the SPIT version of the gain matrix $\boldsymbol{A}(\rho, \theta)$. Therefore we can determine the parameters (ρ, θ) of the sources as those that cause the column space of $\boldsymbol{A}(\rho, \theta)$ to coincide with that of Φ_s. To estimate the signal subspace Φ_s from the data, we can use an eigendecomposition of the sample covariance $\hat{\boldsymbol{R}}_F = \boldsymbol{F}\boldsymbol{F}^T$ or, equivalently, compute the SVD of the data $\boldsymbol{F}$. If the number of time slices is large, we may find the eigendecomposition of $\hat{\boldsymbol{R}}_F$ more practical. In either case, we select the r dominant eigenvectors or left singular vectors to represent our estimate $\hat{\Phi}_s$. In the subspace discussions below, we will find it notationally convenient to refer to this estimate $\hat{\Phi}_s$ as the *signal subspace estimate*, although strictly $\hat{\Phi}_s$ is a matrix whose columns *span* an estimate of the true signal subspace.

5.0 Subspace Correlations

As noted above, we can find the source parameters, and hence the dipole locations, by comparing the column space of the SPIT matrix $\boldsymbol{A}(\rho, \theta)$ to an estimated signal subspace. In this section we will describe how we use the metric of *subspace correlations* [6] to measure the fit between these

two subspaces. Since the signal subspace is spanned both by the columns of $\boldsymbol{A}(\rho, \theta)$ and the eigenvectors in Φ_s, there must exist a full rank $(r \times r)$ transformation matrix $\boldsymbol{T}$ such that

$$\boldsymbol{A}(\rho, \theta)\boldsymbol{T} = \Phi_s. \tag{33}$$

Since the signal subspace is computed from the data, however, we only have an estimate of the signal subspace matrix Φ_s

$$\boldsymbol{A}(\rho, \theta)\boldsymbol{T} \approx \hat{\Phi}_s. \tag{34}$$

One approach to source localization using *(34)* is the *weighted subspace fitting (WSF)* method (cf. [7], [21], [25], [26]) in which the parameters $\{\rho, \theta, \boldsymbol{T}\}$ are found by minimizing the squared error

$$c_{ls} = \left\| \hat{\Phi}_s \boldsymbol{W} - \boldsymbol{A}(\rho, \theta)\boldsymbol{T} \right\|_F^2 \tag{35}$$

where $\boldsymbol{W}$ is a weighting matrix designed to improve the estimator performance [26], and $\|\bullet\|_F^2$ is the square of the Frobenius norm of the matrix (i.e., the sum of the squares of all elements in the enclosed matrix). Here we propose an alternative procedure in which, rather than solving directly for the parameter set $\{\rho, \theta, \boldsymbol{T}\}$, we instead examine the angles between the subspaces spanned by $\boldsymbol{A}(\rho, \theta)$ and Φ_m using *subspace correlations.*

The subspace correlation function $subcorr\{\boldsymbol{A}, \hat{\Phi}_s\} = \{s_1, s_2, \ldots, s_r\}$ defined in Appendix A yields a set of r ordered scalars $1 \geq s_1 \geq \ldots \geq s_r \geq 0$. These scalars are equal to the cosines of the *principal angles* between pairs of *principal vectors* chosen from the two subspaces $\boldsymbol{A}$ and $\hat{\Phi}_s$, where r is the minimum of the ranks of the two subspaces. The set of r principal vectors for each of the two subspaces are orthonormal. The first pair of principal vectors are chosen using one vector from each of the two subspaces so as to minimize the angle between the two vectors. The second pair are again chosen to minimize the angle between the vectors from the two spaces, but under the constraint that the second principal vector from $\boldsymbol{A}$ must be orthogonal to the first principal vector from $\boldsymbol{A}$, and similarly for the first two principal vectors from $\hat{\Phi}_s$. The process is repeated until a set of r pairs of principal vectors have been found, along with the associated ordered correlations corresponding to the cosines of the angles between each pair.

The computation of the subspace correlation between the signal subspace Φ_s and the SPIT matrix $\boldsymbol{A}(\rho, \theta)$ provides the fundamental basis for the RAP MUSIC algorithm. We describe how this can be computed using the SVD in Appendix A; additional details can be found in [6]. The significance of the subspace correlation function is that if one subspace is entirely contained within another, then the cosines of all the principal angles will equal unity. Conversely, if the two spaces are orthogonal, the cosines of all the principal angles will equal zero. For cases between these extremes, the set of cosine values provide a measure of the similarity between the two subspaces. We will see below that the MUSIC metric corresponds to computing a subspace correlation between a single topography in a SPIT matrix and the estimated signal subspace. The subspace correlations leads to a natural extension of MUSIC. We can recursively build up our source estimate by appending putative sources to the SPIT matrix and using the minimum of the subspace correlations as a metric for adding a new source. We expand on this below, but first consider an illustrative example of this concept.

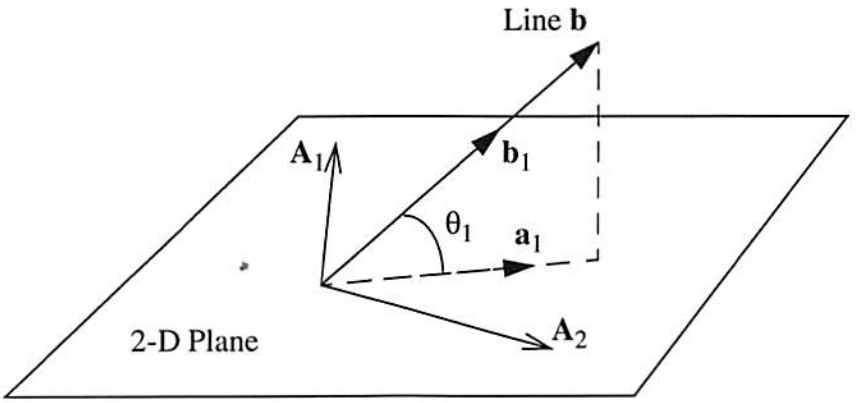

Figure 1. Geometric interpretation of subspace angles. The two-dimensional plane is spanned by the two columns, A_1 and A_2, of the matrix A. These vectors and the vector b pass through the coordinate origin. The function *subcorr*$\{A, b\}$ returns the *principal correlation* $s_1 = \cos\theta_1$, where θ_1 is the *principal angle* between the line and the plane. The *principal vectors* are a_1 and b_1, which are unit length vectors in the plane and line, respectively.

To give an intuitive geometric insight into these subspace correlations, consider *subcorr*$\{A, b\}$, where we briefly suspend our E/MEG notation and define the columns of a 3×2 matrix A to represent two vectors that form a basis for a two-dimensional plane in a three-dimensional space. Similarly, let the 3×1 vector b represent a one-dimensional vector (line). The subspaces A and b both pass through the origin. In this case, *subcorr*$\{A, b\}$ yields a single correlation coefficient, representing the cosine of the angle between the line and the plane. From the discussion in Appendix A, we can directly form $a_1 = Ax_1$, which is the unit length vector in the plane of A closest to b. We illustrate this case in Figure 1. If the correlation is unity, then b lies in the plane of A and $a_1 = b/\|b\|$; if the correlation is zero, then b is perpendicular to the plane, and x_1 is arbitrary.

Next, consider a second 3×2 matrix B, and again the planes formed by the columns of both A and B pass through the origin. We find that the first (maximum) correlation of *subcorr*$\{A, B\}$ is always unity, since two such planes always intersect along a line, namely the line found by Ax_1 or By_1. The second correlation is the cosine of the angle between the planes, the angle we intuitively picture when visualizing two intersecting planes.

Resuming our E/MEG notation, we define the function *distance*$\{A, \hat{\Phi}_s\}$ [6] as

$$distance\{A, \hat{\Phi}_s\} = \sqrt{1 - min\{s_1, s_2, \ldots, s_r\}^2} = \sqrt{1 - s_r^2}. \qquad (36)$$

Assuming that the rank of $A(\rho, \theta)$ is less than or equal to that of the signal subspace estimate $\hat{\Phi}_s$, the distance as defined in *(36)* will approach zero as the column space of $A(\rho, \theta)$ matches that of $\hat{\Phi}_s$. Consequently we can determine the parameters $\{\rho, \theta\}$ of the sources that produced the esti-

mated signal subspace as the set that jointly minimize the distance between our topographies matrix $\boldsymbol{A}(\rho, \theta)$ and our estimated signal subspace $\hat{\Phi}_s$:

$$\{\hat{\rho}, \hat{\theta}\}_{\text{mindist}} = \arg\min\{\mathit{distance}\{\boldsymbol{A}(\rho, \theta), \hat{\Phi}_s\}\}. \tag{37}$$

For each candidate set of parameters $\{\rho, \theta\}$, we generate a spatial topographies matrix $\boldsymbol{A}(\rho, \theta)$. We then compute the minimum subspace correlation s_r of this SPIT matrix with our estimated signal subspace. We can then find the set $\{\hat{\rho}, \hat{\theta}\}_{\text{mindist}}$ that maximizes this minimum correlation or equivalently minimizes $\mathit{distance}\{\boldsymbol{A}, \hat{\Phi}_s\}$, such that the subspace spanned by our topographies matrix is as close as possible to "parallel" with our signal subspace estimate.

For multiple dipoles, the key concept that makes subspace distance easier to use than least-squares fitting is that if $\boldsymbol{A}(\rho, \theta)$ is parallel to $\hat{\Phi}_s$, then so is each column (each topography) of $\boldsymbol{A}(\rho, \theta)$. Since the principal vector of $\boldsymbol{A}$ associated with $s_r = \min\{\mathit{subcorr}\{\boldsymbol{A}, \hat{\Phi}_s\}\}$ corresponds to the linear combination of the columns of $\boldsymbol{A}$ that minimizes the correlation between the two spaces, it follows that the ith column of $\boldsymbol{A}$, i.e., the ith independent topography, must have a correlation greater than or equal to this minimum subspace correlation,

$$\mathit{subcorr}\{\boldsymbol{a}(\rho_i, \boldsymbol{u}_i), \hat{\Phi}_s\} \geq \min\{\mathit{subcorr}\{\boldsymbol{A}(\rho, \theta), \hat{\Phi}_s\}\}. \tag{38}$$

In our SPIT model, each column of $\boldsymbol{A}$ represents an independent topography, where each topography may comprise multiple synchronous dipoles. For exemplary purposes, let us assume that each topography represents a single current dipole. Let us further assume that we have a perfect signal subspace estimate Φ_s, in which case the minimum subspace correlation will be unity for the true parameters $\{\rho, \theta\}$. From *(38)*, each of the independent topographies formed by each dipole must also have a correlation of unity with the subspace. We can therefore find the dipole parameters by searching for the p dipole locations that each have unity correlation. Thus a search strategy for minimizing the distance between the topographies matrix and the rank r signal subspace estimate is to search for a single dipole model whose subspace correlation is maximized with respect to $\hat{\Phi}_s$. We should find r such dipole locations in our dipolar space, each yielding a correlation value of unity. This search strategy is the basis of the MUSIC algorithm that we described in [10] and will briefly review in the next section.

Before proceeding to a description of MUSIC, we first address the problem of finding the orientation vector $\boldsymbol{u}_i$. The dipole parameters are chosen to maximize

$$\mathit{subcorr}\{\boldsymbol{a}(\rho_i, \boldsymbol{u}_i), \hat{\Phi}_s\} \tag{39}$$

However, $\boldsymbol{u}_i$ simply represents a linear combination of the columns of the gain matrix $\boldsymbol{G}(\rho_i)$ (see equation *(24)* and preceding discussion). We can avoid searching for the optimal orientation vector by noting that the maximum of the subspace correlation vector $\mathit{subcorr}\{\boldsymbol{G}(\rho_i), \hat{\Phi}_s\}$ gives us the best way of combining the columns of $\boldsymbol{G}(\rho_i)$ so that they are as close as possible to the signal subspace. We can therefore find the optimal orientation vector $\boldsymbol{u}_i$ for each candidate location ρ_i as that which maximizes the subspace correlation at that location, i.e.:

$$\max\{\mathit{subcorr}\{\boldsymbol{G}(\rho_i), \hat{\Phi}_s\}\} \tag{40}$$

Therefore we can find the dipole locations by solving *(40)* at each candidate dipole location, and then searching for the true locations at which this maximum correlation equals, or is sufficiently close to, unity. Once we find these locations, we can then explicitly form the corresponding best orientation (from Appendix A, set $\boldsymbol{u}_i$ to "$\boldsymbol{x}_1$" and scale to unity norm), to determine the independent topography vector $\boldsymbol{a}(\rho_i, \boldsymbol{u}_i)$.

6.0 Classical MUSIC

In [10] we adapted a "diversely polarized" form of Schmidt's original MUSIC algorithm [5], [18] to the problem of multiple point dipoles. We briefly review and update that presentation here to include our discussion of subspace correlations. The steps are:

1. Obtain a spatio-temporal data matrix $\boldsymbol{F}$, comprising information from m sensors and n time slices. Decompose $\boldsymbol{F}$ using the SVD to yield $\boldsymbol{F} = \boldsymbol{U}_F \Sigma_F \boldsymbol{V}_F^T$.

2. Examine the singular values in Σ_F and select the rank of the signal subspace to obtain $\hat{\Phi}_s$. Overspecifying the true rank by a couple of dimensions usually has little effect on performance. Underspecifying the rank can dramatically reduce the performance.

3. Create a relatively dense grid of dipolar source locations. At each grid point, form the gain matrix $\boldsymbol{G}$ for the dipole. At each grid point, calculate the subspace correlations $subcorr\{\boldsymbol{G}, \hat{\Phi}_s\}$.

4. As a graphical aid, plot the inverse of $\sqrt{1 - s_1^2}$, where s_1 is the maximum subspace correlation. Correlations close to unity will exhibit sharp peaks (indeed, perfect correlation yields an infinite spike). Locate r or fewer peaks in the grid. At each peak, refine the search grid to improve the location accuracy, and check the second subspace correlation. A large second subspace correlation is an indication of a "rotating dipole."

The measure $\sqrt{1 - s_1^2}$ is equivalent to the correlation with the noise-only subspace, the original proposal by Schmidt. As we discussed in [10], plotting the inverse of this measure makes graphical location of the peaks easier; however, since that publication we have found it more informative to plot the principal correlation, since correlation is a direct measure of how well the model fits the data.

7.0 RAP-MUSIC

Problems with the use of MUSIC arise when there are errors in the estimate of the signal subspace and the subspace correlation is computed at only a finite set of grid points. The largest peak is usually easily located by searching over the grid for the largest correlation; however, the second and subsequent peaks must be located by means of a three-dimensional "peak-picking" routine. Suppose that an incorrect set of locations are picked. While individually each of the dipoles may have good correlations with the signal subspace, there is no guarantee that their combined SPIT model has a small distance from the signal subspace, since we test only one dimension at a time. The RAP-MUSIC methods overcomes this problem by recursively building up the SPIT model and comparing this full model to the signal subspace. A second problem with MUSIC as discussed in

[10] is that it will fail to locate synchronous sources - this problem is also addressed by the RAP-MUSIC algorithm.

In the following we assume that our independent topographies each comprise one or more dipoles. We search first for the single dipolar topographies, then the two-dipolar topographies, and so forth. As we discover each topography model, we add it to our existing SPIT model and continue the search. We build the source model by recursively applying the subspace correlation measure, the key metric of MUSIC, to successive subspace correlations.

For exemplary purposes, we assume that the r independent topographies each comprise a single dipole. Conceptually, RAP-MUSIC begins by finding the first dipole location as that which maximizes *(40)*. Single dipole locations are readily found by scanning the head volume. At each point in the volume, we calculate

$$\{s_1, s_2, \ldots\} = subcorr\{\boldsymbol{G}(\boldsymbol{r}_q), \hat{\Phi}_s\} \tag{41}$$

where $\{s_1, s_2, \ldots\}$ is the set of subspace correlations. We find the dipole location $\hat{\boldsymbol{r}}_{q1}$ which maximizes the primary correlation s_1. As discussed in Appendix A, the corresponding dipole orientation $\hat{\boldsymbol{u}}_1$ is easily obtained from $subcorr\{\boldsymbol{G}(\hat{\boldsymbol{r}}_{q1}), \hat{\Phi}_s\}$, and we designate our topography model comprising this first dipole as

$$\hat{\boldsymbol{A}}^{(1)} = \boldsymbol{a}(\hat{\boldsymbol{r}}_{q1}, \hat{\boldsymbol{u}}_1). \tag{42}$$

To search for the second dipole, we again search the head volume; however, at each point in the head, we first form the model matrix $\boldsymbol{M} = [\hat{\boldsymbol{A}}^{(1)}, \boldsymbol{G}(\boldsymbol{r}_q)]$. We then calculate

$$\{s_1, s_2, \ldots\} = subcorr\{\boldsymbol{M}, \hat{\Phi}_s\} \tag{43}$$

but now we find the dipole point that maximizes the *second* subspace correlation, s_2; the first subspace correlation should already account for $\boldsymbol{a}(\hat{\boldsymbol{r}}_{q1}, \hat{\boldsymbol{u}}_1)$ in the model. The corresponding dipole orientation $\hat{\boldsymbol{u}}_2$ may be readily obtained by projecting this second topography against the subspace, $subcorr\{\boldsymbol{G}(\hat{\boldsymbol{r}}_{q2}), \hat{\Phi}_s\}$, and we append this to our model to form

$$\hat{\boldsymbol{A}}^{(2)} = [\boldsymbol{a}(\hat{\boldsymbol{r}}_{q1}, \hat{\boldsymbol{u}}_1), \boldsymbol{a}(\hat{\boldsymbol{r}}_{q2}, \hat{\boldsymbol{u}}_2)]. \tag{44}$$

We repeat the process r times, maximizing the kth subspace correlation at the kth pass, $k = 1, \ldots, r$. The final iteration is effectively attempting to minimize the subspace distance between the full r topographies matrix and the signal subspace estimate.

If the r topographies comprise r_1 single-dipolar topographies and r_2 2-dipolar topographies, then RAP-MUSIC will first extract the r_1 single dipolar models. At the $(r_1 + 1)$th iteration, we will find no single dipole location that correlates well with the subspace. We then increase the number of dipole elements per topography to two. We must now search simultaneously for two dipole locations, such that

$$\{s_1, s_2, \ldots\} = subcorr\left\{[\hat{\boldsymbol{A}}^{(r_1)}, \boldsymbol{G}(\rho)], \hat{\Phi}_s\right\} \tag{45}$$

is maximized for the subspace correlation s_{r_1+1}, where $\rho = \{\boldsymbol{r}_{q1}, \boldsymbol{r}_{q2}\}$ comprises two dipoles. If the combinatorics are not impractical, we can exhaustively form all pairs on our grid and compute maximum subspace correlations for each pair. The alternative is to begin a two-dipole nonlinear search with random initialization points to maximize this correlation. This low-order dipole search can be easily performed using standard minimization methods.

We proceed in this manner to build the remaining r_2 2-dipolar topographies. As each pair of two dipoles is found to maximize the appropriate subspace correlation, the corresponding pair of dipole orientations may be readily obtained from $subcorr\{\boldsymbol{G}(\rho), \hat{\Phi}_s\}$, as detailed in Appendix A. Extensions to more dipoles per independent topography are straightforward, although the complexity of the search obviously increases. In any event, the complexity of the search will always remain less than the least-squares search required for finding all dipoles simultaneously.

To summarize, we assume that our forward model has been corrupted by additive noise, and that this noise is zero mean, i.i.d., with a known spatial covariance matrix $\sigma_n^2\boldsymbol{I}$. The case of non i.i.d. noise is readily treated with standard "pre-whitening" procedures [20]. We then SVD $\boldsymbol{F}$ to yield $\boldsymbol{F} = \boldsymbol{U}_F\Sigma_F\boldsymbol{V}_F^T$ or alternatively eigendecompose $\boldsymbol{F}\boldsymbol{F}^T$ to yield $\boldsymbol{F}\boldsymbol{F}^T = \boldsymbol{U}_F\Sigma_F^2\boldsymbol{U}_F^T$, where we retain only those components corresponding to strictly nonzero singular values.

We examine the singular values to determine the rank r of the signal subspace, and we retain only the corresponding r singular vectors in $\boldsymbol{U}_F$ to form $\hat{\Phi}_s$, which is our estimate of a set of vectors that span the signal subspace. If the rank is uncertain, we should err towards overspecifying the signal subspace rank. If we overselect the rank, the additional subspace vectors should span an arbitrary subspace of the noise-only subspace, and the probability that these vectors correlate with our model is small; hence, we may in general overspecify the rank of the signal subspace. However, as the overspecification of the noise-only subspace increases, so does the probability that we may inadvertently obtain a noise subspace component that correlates with our models, so some prudence is called for in rank selection.

We design a sufficiently dense grid in our volume of interest, and at each grid point $\boldsymbol{l}_i$ we form the head model for the single dipole gain matrix $\boldsymbol{G}(\boldsymbol{l}_i)$.We initialize the topography complexity as "1-dipolar topography," i.e., each topography comprises a single dipole. We then proceed as follows:

1. For *index* from 1 to rank r:

2. Let $\hat{\boldsymbol{A}} = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{(index-1)}]$ be the model extracted as of the previous loop ($\hat{\boldsymbol{A}}$ is a null matrix for the first loop).

3. Form sets of grid points λ_i, where for a 1-dipolar topography each set consists of the location of a single grid point. For a 2-dipolar topography, λ_i contains the locations of pairs of grid points, and so on for higher order dipolar topographies. If the combinatorics make it impractical to consider all possible combinations of grid points, choose a random subset of the possible

combinations.

4. For each set of grid points λ_i, form the grid model $\boldsymbol{M}_i = [\hat{\boldsymbol{A}}, \boldsymbol{G}(\lambda_i)]$, i.e., concatenate the set of grid point models to the present extracted model.

5. Calculate the set of subspace correlations, $\{s_1, s_2, \ldots\} = subcorr\{\boldsymbol{M}_i, \hat{\Phi}_s\}$, using the algorithm described in Appendix A.

6. Find the maximum over all sets of grid points λ_i for s_{index}, e.g., for $index = 2$, find the maximum *second* subspace correlation.

7. Optionally, if the set of grid points $\{\lambda_i\}$ is not particularly dense or complete, then use a nonlinear optimization method (e.g., Nelder-Meade simplex) to maximize s_{index}, beginning the optimization at the best set of grid point λ_i. If the grid is dense and our sets in Step 3 complete, this step may not be necessary.

8. Is the correlation at the location of the maximum "sufficient," i.e., does s_{index} indicate a good correlation? If the correlation is adequate, proceed to Step 11. If it is not, proceed to Step 9.

9. (Insufficient correlation in Step 8). We have two situations to consider. We may have over-specified the true rank of the signal subspace, in which case we are now attempting to fit a topography into a noise-only subspace component. We can test for this condition by forming the projection operator $\boldsymbol{P}_{\hat{A}} = \hat{\boldsymbol{A}}\hat{\boldsymbol{A}}^{\dagger}$ (where $\hat{\boldsymbol{A}}^{\dagger}$ is the *pseudoinverse*, see [10] and references therein) from the existing estimated model, then forming the residual $\boldsymbol{F}_{res} = \boldsymbol{F} - \boldsymbol{P}_{\hat{A}}\boldsymbol{F}$. Inspection and testing of the residual should reveal whether or not we believe a signal is still present. If we believe the residual is simply "noise," break this loop. Otherwise, proceed to Step 10.

10. (Signal still apparent in the residual) Increase the complexity of the topography (e.g. from one to two dipolar) and return to Step 3 without increasing the loop *index*.

11. (Good correlation in Step 8) We have found the best set of locations $\hat{\rho}_{index}$ of the next independent topography, with corresponding gain matrix $\boldsymbol{G}_{index}(\hat{\rho}_{index})$. We need the best fitting orientation. From the discussion in Appendix A, calculate the principal orientation vector $\boldsymbol{x}_1$ from $subcorr\{\boldsymbol{G}_{index}(\hat{\rho}_{index}), \hat{\Phi}_s\}$, normalize $\hat{\boldsymbol{u}}_{index} = \boldsymbol{x}_1 / \|\boldsymbol{x}_1\|$, and form the topography vector $\hat{\boldsymbol{a}}_{index} = \boldsymbol{G}_{index}(\hat{\rho}_{index})\hat{\boldsymbol{u}}_{index}$.

12. Increment the *index* and loop to Step 1 for the next independent topography.

In Step 8, we recommend here that the correlation exceed at least 95%. In [12], we discuss some of the means for determining if a MUSIC peak represents "adequate" or "sufficient" correlation, and we will return to this topic in future papers. Our recommendation here of 95% reflects our empiricism that a "good" solution should generate a topography which explains at least 90% (the square of the correlation, i.e., the "R-squared" statistic) of the variance of the topography identified in the data. If we overselect the rank of the signal subspace, then we will in general break out of

the loops at Step 9, once we have found the true number of sources and have only noise left in the residual. We will not address the determination of statistical "sufficiency" of the model in this paper. The interested reader may refer to ([1], [23]) among others for discussions on the testing of the residual for remnant signals.

If the grid is dense or we performed Step 7 for each topography, we may find the RAP-MUSIC set of parameters is already a good solution. The RAP-MUSIC algorithm has maximized a set of subspace correlations, a metric different from that of the more traditional least-squares approach. If so desired, we may proceed with a local least-squares search:

13. Our RAP-MUSIC search has yielded an estimate of the full spatial topographies gain matrix $\hat{A} = [a_1, \ldots, a_r]$, which is a function of the estimated full set of dipole locations $\hat{\rho}$ and orientations $\hat{\theta}$. Beginning with these parameters, initialize a nonlinear search for the cost function,

$$\{\hat{\rho}, \hat{\theta}\}_{\text{ls}} = arg\ max\left\{ \left\| \boldsymbol{U}^T_{A(\rho,\theta)} \boldsymbol{U}_F \Sigma_F \right\|^2_F \right\} \tag{46}$$

where $\boldsymbol{U}_{A(\rho,\theta)}$ is an orthogonal matrix spanning the same subspace as $\boldsymbol{A}(\rho, \theta)$, found for example by an SVD or QR decomposition of $\boldsymbol{A}(\rho, \theta)$. In *(46)*, we have stripped away the norm preserving orthogonal matrices in order to increase the computational speed of what is otherwise a straightforward nonlinear least-squares search; see [10] for more discussions on the details of implementing efficient least-squares solutions.

Step 13 represents an increase in the complexity in the nonlinear search over that of RAP-MUSIC, at possibly diminishing returns in terms of improvement in the solution. Each iteration of the nonlinear search must now adjust the parameters of all of the dipoles, not just a single topography as in RAP-MUSIC. Analysis of the benefits of one estimator over another usually require statistical assumptions that may be difficult to justify for short sequences of data (cf. [27]). We may also find in practice that other modeling assumption errors, such as head models, sensor models, and source models, as well as environmental contaminations and unaccounted "brain noise", may be sufficiently confounding to preclude the additional refinement of any specific source model.

Once we find the optimal $\{\hat{\rho}, \hat{\theta}\}$, we can find the remaining linear temporal parameters as

$$\boldsymbol{S}^T = \boldsymbol{V}_{A(\hat{\rho},\hat{\theta})} \Sigma^{-1}_{A(\hat{\rho},\hat{\theta})} \boldsymbol{U}^T_{A(\hat{\rho},\hat{\theta})} \boldsymbol{F} \tag{47}$$

i.e., we use the pseudoinverse of $\boldsymbol{A}(\hat{\rho}, \hat{\theta})$, as discussed in [10] and its references.

8.0 Computer Simulations

We present two simulations to illustrate some of the features of our proposed SPIT model and the RAP-MUSIC algorithm. In the first simulation, we arranged 255 EEG sensors about the upper region of an 8.8 cm single shell sphere, with a nominal spacing between sensors of one cm. For illustrative purposes, we arranged three dipolar sources in the same plane, $z = 7$ cm, and the three

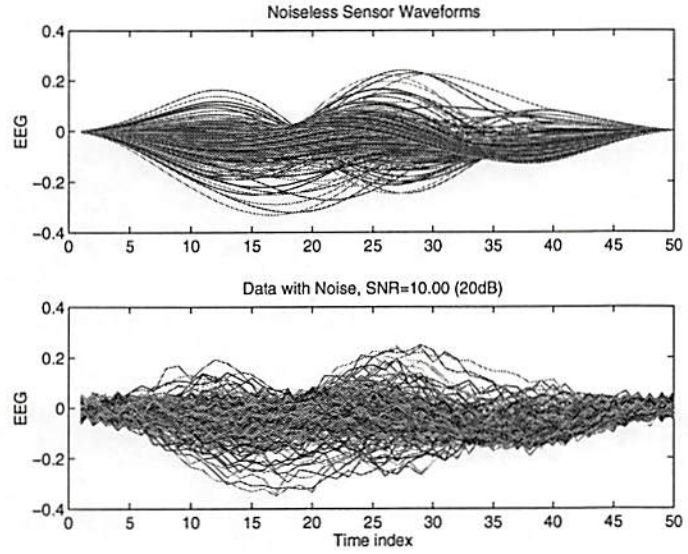

Figure 2. The upper plot is the overlay of the response of all 255 EEG sensors to the simulated three dipolar sources. The sources were given independent, overlapping time series. White Gaussian noise was added such that the SNR was 20 dB, and the percent variance explained by the true solution was 91% of the total variance.

sources were given independent, overlapping time courses. The overlay of the responses of all sensors is given in the upper plot of Figure 2. We then added white Gaussian noise to all data points, scaled such that the squared Frobenius norm of the noise matrix was one-tenth that of the squared Frobenius norm of the noiseless signal matrix, for an SNR of 20 dB. The lower plot of Figure 2 shows the overlay of all sensors for the signal plus noise data.

The singular value spectrum was clearly rank three, but we selected rank five to illustrate the robustness of rank over-selection. We created a one mm grid in the $z = 7$ cm plane and calculated the correlation between a single dipole model and the signal subspace. Figure 3 displays these correlation as an image whose intensities are proportional to the primary correlation s_1. In Figure 4, we have replotted the same data, but in this case plot $1/(\sqrt{1-s_1^2})$ in order to graphically intensify the appearance of the peaks. This image is essentially the original MUSIC scan proposed in [10]. The largest correlation of 99.8% is easily found at $[-1, -1, 7]$. The peak at $[1, 1, 7]$ is apparent in Figure 4, but the peak at $[0, 0, 7]$ is obscured in either figure. Graphically or computationally declaring the location of these other two peaks is non-obvious without subjective interpretation by the viewer.

We generated the forward field for this first dipole, then re-scanned the subspace correlation on the same grid with the combined model. Figure 5 displays the *second* subspace correlation; in this and subsequent figures, we will resume plotting the correlation value directly, rather than the inverted metric. We can now more clearly see the peaks corresponding to the two remaining sources, and the first source has been suppressed. The maximum peak of this image at 99.7% is easily located at $[1, 1, 7]$.

We generated the forward field for this second dipole and appended it to the first dipole's forward field. We then re-scanned the subspace correlations on the same grid with the combined model. Figure 6 displays the *third* subspace correlation, where we now readily observe the single remaining peak of the third source, 99.6% at $[0, 0, 7]$. Visual examination of the residual at this point

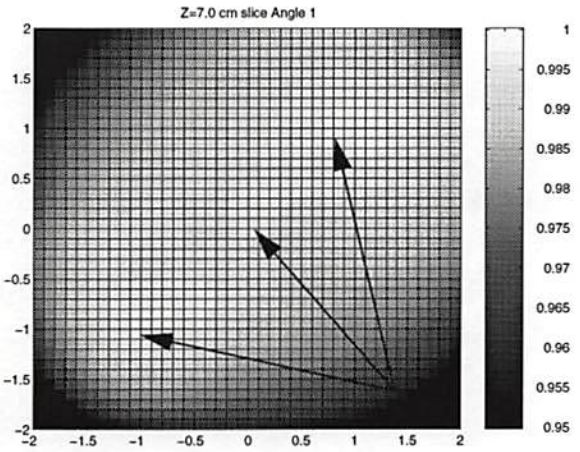

Figure 3. Principal correlation between an EEG dipolar model and the rank five signal subspace extracted from the data. The correlations were calculated in the $z = 7$ cm plane on a one mm grid. Each grid point was then scaled in intensity per the color bar on the right side of the figure. The largest correlation of 99.8% is at $[-1, -1, 7]$; however, this peak and the other two peaks, as indicated by the arrows, are not readily discernible, either graphically or computationally.

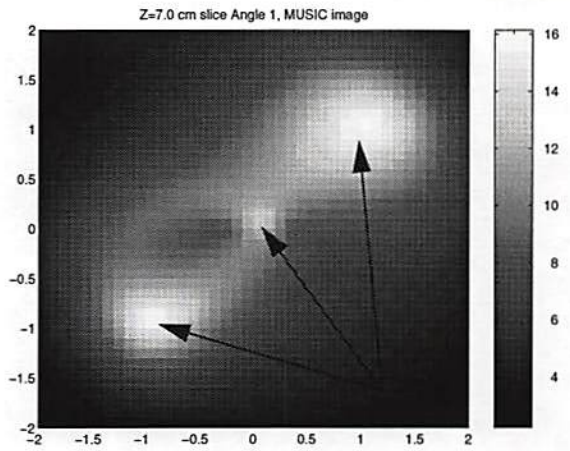

Figure 4. Rather than image each pixel as a scaled version of a correlation value s_1 as in Figure 3, here we instead image $1/(\sqrt{1-s_1^2})$. This image metric is equivalent to our MUSIC metric discussed in [10]. We see two of the three peaks rather clearly, but the central peak is somewhat obscured, since its graphical "intensity" is about 75% of the intensity of the other two peaks. Although the peaks are better defined than in Figure 3, interpretation of the intensity scale is now ambiguous, and the user must still "peak pick" graphically or algorithmically in three dimensions. Subsequent figures resume the correlation scaling of Figure 3.

indicated no remaining signal, and subspace correlations of multiple dipole models yielded no substantial correlations. We thus correctly halted the algorithm after this third topography.

The second simulation was designed to demonstrate the localization of a "rotating" dipole and a pair of synchronous dipoles, as well as to illustrate the use of a directed search algorithm to refine these locations. In this simulation, we arranged 240 MEG planar gradiometer sensors about the upper hemisphere, with a nominal spacing of about two cm and a baseline separation of one cm. A "rotating" dipole was located at $[0, 0, 7]$ cm, and a pair of dipoles with synchronous activation was located at $[-2, -2, 7]$ and $[2, 2, 7]$ cm. We then created a 1.5 mm grid in the $z = 6.5$ cm plane, i.e., in a plane displaced from the true source plane, and the gridding was slightly coarser than the first simulation. The noise level was again set to 20 dB. The true rank of the signal subspace was

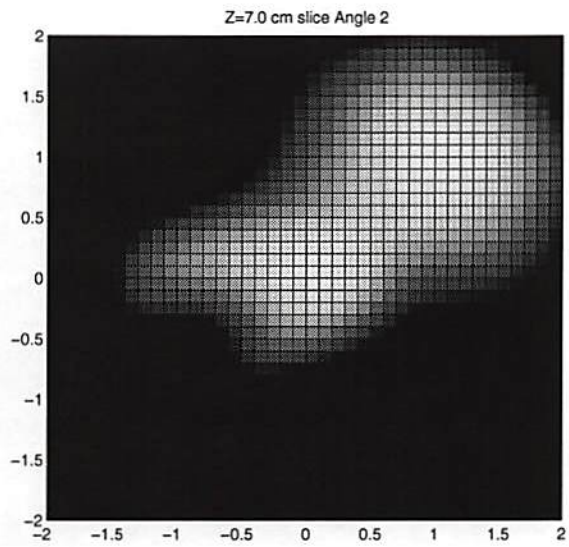

Figure 5. The forward field from the first dipolar solution was formed and concatenated with the forward field generated by each point on this grid. The subspace correlations were again computed between this combined model and the signal subspace. The *second* subspace correlation is displayed here as an image. The original peak has been suppressed, and we more clearly see two peaks in this image. The maximum correlation in this images is found at $[1, 1, 7]]$ of 99.7%.

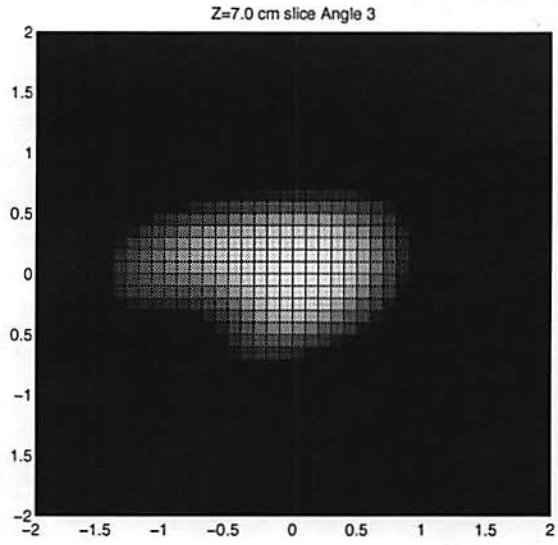

Figure 6. We generated the forward fields for the first two dipolar locations and concatenated that set with the field generated by each point in the grid. We then reran the subspace correlations, then imaged the *third* subspace correlation. The remaining source solution is now easily observed at $[0, 0, 7]$, with a correlation of 99.6%. The percent variance explained by the combined three dipoles was 91%.

three, with the rotating dipole comprising two single-dipolar topographies, and the third topography comprising a two-dipolar topography. Figure 7 displays the overlay of the noiseless and noisy sensor responses.

We again overselected the rank of the signal subspace rank to be five, then scanned the one dipole model against the signal subspace. We found a single good peak at 99.3%, as displayed in Figure 8. Note the absence of any other peaks; the remaining "rotating" dipolar topography is obscured by this peak, and the other topography is not a single dipole. The peak observed in the grid was at $[0.1, -0.2, 6.5]$ cm. We initiated a directed search from this point to maximize the correlation to 99.8% at $[0.0, 0.0, 7.0]$ cm, the correct solution for the single dipole topography.

As in the previous example, we then scanned for a second dipole, observing the second subspace correlation. The maximum correlation in the grid was again high, 99.2%, at $[-0.1, -0.2, 6.5]$, as shown in Figure 9. A directed search initiated at this point maximized the second subspace correlation at 99.7% at $[0, 0, 7]$, the same dipole location as the first solution. The dipole orientations

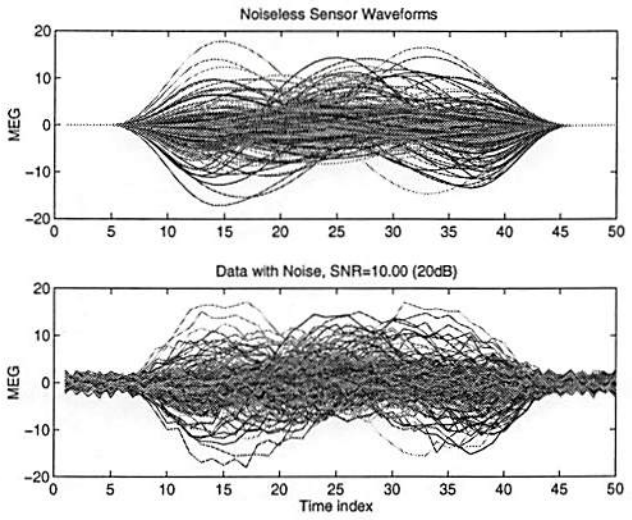

Figure 7. MEG Simulation comprising 240 planar gradiometers with a one cm baseline spaced about two cm apart, arranged about the upper region of a 12 cm sphere. A "rotating" dipole was located at [0, 0, 7] cm, and a pair of dipoles with synchronous activation was located at [–2, –2, 7] and [2, 2, 7] cm. The white Gaussian noise was scaled such that squared Frobenius norm of the signal was ten times that of the noise (20 dB). The true rank of the signal subspace was three, but we again overselected it to be five.

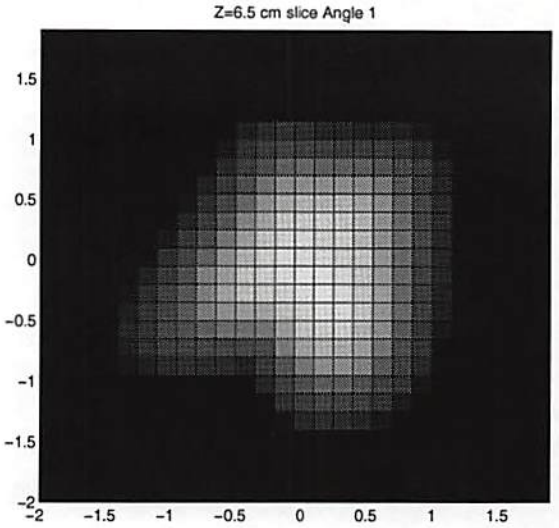

Figure 8. Principal correlation between the one dipole model and the rank five MEG subspace. A single peak is observed at the location of the "rotating" dipole; the second dipolar source is effectively buried by this peak. This plane was scanned at $z = 6.5$ cm, but the true solution lay above this scanning plane. The peak in this plane was used to initiate a directed search for the maximum correlation, which was located correctly at [0, 0, 7] cm.

of the two solutions were nearly orthogonal, [0.94, 0.34, 0] and [0.3, 0.95, 0], indicating we had correctly identified the simulated "rotating" dipole at this point.

We then scanned for a third single dipole solution, but only a peak of 88.8% was found, and a directed search maximization only improved this correlation to 88.9%. Thus this third dipole could only account for $(88.9\%)^2 = 79.0\%$ of the variance of the third topography, and we rejected this third single dipolar topography solution.

Since one dipole was inadequate to describe the third topography, we shifted to our next putative solution, that of two dipoles. Our grid comprised 729 dipole locations, and all combinations of two dipoles yielded 265,356 sets. Rather than exhaustively search all set combinations, we randomly selected a small subset of these sets for a total of about 3,000 sets. We then concatenated each of these 3,000 pairs with the first two dipole solutions, calculated the subspace correlation of the combined model and observed the third subspace correlation. The maximum third correlation of 98.4% corresponded to the pair at [–2, –1.9, 6.5], [1.8, 1.8, 6.5]. We initiated a two dipole directed search from this set and achieved a maximum correlation of 99.7% at

[–2.0, –2.0, 7.0], [2.0, 2.0, 7.0] cm, the correct solution. In Figure 10, we plot two-dimensional cross-slices of this six-dimensional function, holding constant the correct z plane and the true location of one of the two dipoles. We clearly observe the correlation metric peaking at the correct solution. As in the first example, visual examination of the residual from this model revealed that no signal was present, and further correlations with multiple dipole models yielded no substantial correlations. The algorithm was thus correctly halted after this third topography.

This relatively simple pair of simulations has illustrated some of the key concepts of the RAP-MUSIC algorithm and the SPIT model. Both simulations used relatively dense grids of EEG or MEG sensors, such that sensor spacing was not an issue; see [11] for analysis of the effects of EEG and MEG sensor spacing on dipole localization performance. In both simulations, we overselected the true rank of the signal subspace to illustrate the robustness to such an error; we repeated the localization results with the true rank and achieved nearly identical results to those presented here. In these simulations, as in [10], the subspace scans were presented as images to highlight the MUSIC peaks; however, the RAP-MUSIC algorithm readily extracts these peaks without the need for the user to manually observe and select these solutions. Indeed, in these simulations, the set of MUSIC peaks would have been difficult to discriminate either graphically or computationally, due to their proximity and the noise.

In practice, after we have scanned on a relatively good grid for any of the single or multiple dipolar solutions, we always then initiate a directed search from these points to maximize the correlation. By optimizing the correlation in this manner, we bypass some of the concerns of coarse or inadequate gridding. In the first simulation, each of the three dipoles was located with a single dipole search of three location parameters; by contrast, a full nonlinear least-squares would have required nine parameters. In the second simulation, we performed two single-dipole searches, followed by a two-dipole search of six nonlinear parameters. A full nonlinear least-squares search would have required a twelve parameter search.

9.0 Discussion

We have introduced the concept of *spatially independent topographies (SPIT)* to describe the spatio-temporal data matrix. The number of independent topographies is equal to the rank of the signal subspace. In the forward problem, we can easily determine this rank by an SVD of the noiseless spatio-temporal data matrix. In the inverse problem, the conditioning of the data matrix may be too poor to allow estimation of the full signal subspace in the presence of the noise. In these instances, the number of independent topographies has effectively dropped, and we adjust our model of asynchronous and synchronous dipoles accordingly. The RAP-MUSIC procedure handles initial overspecification of the signal subspace rank and allows the user at each iteration to examine the true rank.

RAP-MUSIC was designed to address two common problems in E/MEG processing. Due to either bisynchronous activation or strong noise, we often encounter dipolar sources that are effectively fully correlated in their time sequence activation. The SPIT model allows a straightforward interpretation of these correlated sources that in turn allows us to proceed theoretically with a simple extension of our original MUSIC approach [10]. The second problem, however, arises from the fact that while other nonlinear optimization methods require a search for a single global optimum,

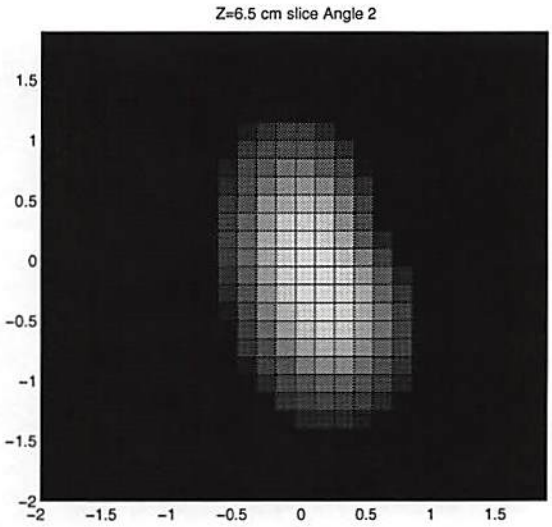

Figure 9. Second subspace correlation between the one dipole model and the rank five MEG subspace, using the solution from Figure 8. The peak in this plane was used to initiate a directed search for the maximum second correlation, which was located correctly, again at $[0, 0, 7]$ cm as in Figure 8. We have thus located a "rotating" dipole. The remaining SPIT source was not a single dipole and thus not observed in these single dipole scans.

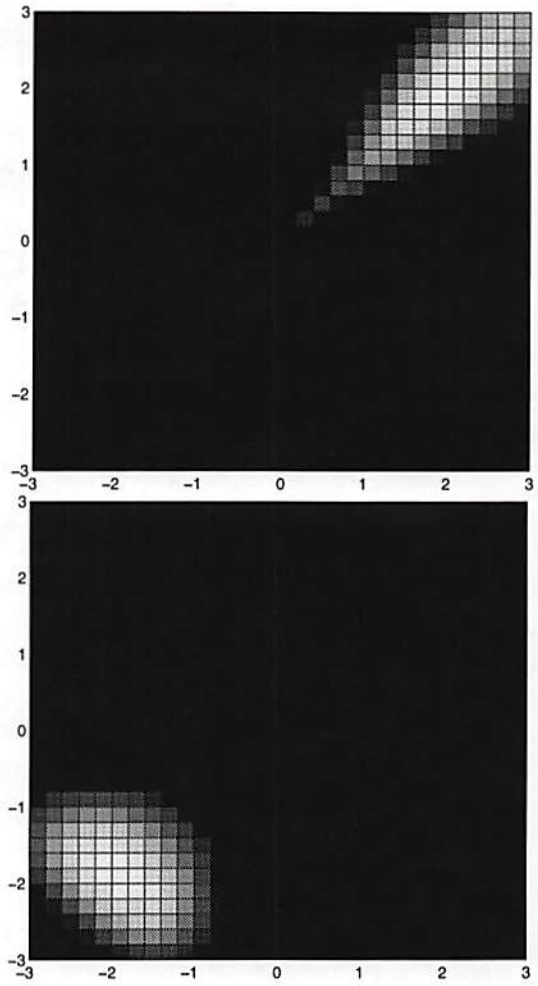

Figure 10. Third subspace correlation of the MEG simulation, computed in a six-dimensional space comprising the synchronous two-dipole topography. Pairs of dipoles were concatenated to the model identified in the first two topographies, then the third subspace correlation computed. The top figure is a two-dimensional slice through this six-dimensional space, holding fixed the true location of one of the dipoles and the true $z = 7$ cm value of the second dipole. We observe the correlation peaking correctly at $[2, 2, 7]$. The bottom figure holds fixed this dipole and the same z value, and peaks correctly at the other dipole, $[-2, -2, 7]$.

MUSIC requires a search for multiple peaks whose amplitudes have been perturbed by noise. While "peak-picking" multiple peaks in a single parameter case (the common presentation in much of the array signal processing literature on MUSIC) is possible, we found the problem confounding in even our simplest case of single dipolar topographies, where we must search for peaks in three-dimensions. Graphically searching for multiple peaks in two-dipolar topographies (a six-dimensional space) is generally not practical.

By maximizing each successive subspace correlation, the RAP-MUSIC approach solved the multiple peak search problem, since each peak corresponds to a separate correlation value. We also solved a second, more subtle issue regarding this search. In a subsequent review of the signal processing literature for similar approaches, we found two comparable MUSIC algorithms, S-MUSIC [15] and IES-MUSIC [22], with the latter introduced as an extension of the former. These "successive" MUSIC algorithms were described for two single-parameter independent sources. The possibility of extending the successive approach to more sources in a manner similar to RAP-MUSIC is mentioned, but specifically not pursued ([22], Remark 2). Both methods, however, implement the successive search in a projection matrix approach different from the subspace correlations approach of RAP-MUSIC. As pointed out in [22], when searching for the second solution, S-MUSIC is undefined at the location of the first solution, and the first solution location must therefore be algorithmically avoided. IES-MUSIC still retains a peak at the first solution, and therefore the search for the second source must also algorithmically avoid this prior location. By recursively shifting to the next subspace correlation, the RAP-MUSIC algorithm bypasses this problem of previous solution points and simply maximizes each subsequent correlation.

In this paper, we have used multiple dipoles as our source model, increasing the independent topography complexity by simply increasing the number of synchronous dipoles. In future work, we will introduce other source models that can account for more distributed sources and show that these models follow the same source framework established here. We will find these "focal source" models useful in situations where many closely spaced dipoles have been synchronously activated, such that errors preclude our ability to resolve the individual clustered dipoles.

Appendix A Subspace Correlation

Here we summarize the definition and method for computation of the subspace correlation. Given two matrices, $\boldsymbol{A}$ and $\boldsymbol{B}$, where $\boldsymbol{A}$ is $m \times p$, and $\boldsymbol{B}$ is $m \times q$, let r be the minimum of the ranks of the two matrices. We wish to calculate a function $subcorr\{\boldsymbol{A}, \boldsymbol{B}\} = \{s_1, s_2, \ldots, s_r\}$, where the scalars s_k are defined as follows.

$$s_k = \max_{a \in A} \max_{b \in B} \; a^T b = a_k^T b_k \tag{48}$$

subject to:

$$\begin{aligned} \|a\| &= \|b\| = 1 \\ a^T a_i & \qquad i = 1, \ldots, k-1 \\ b^T b_i & \qquad i = 1, \ldots, k-1 \end{aligned} \tag{49}$$

The vectors $\{a_1, \ldots, a_r\}$ and $\{b_1, \ldots, b_r\}$ are the *principal vectors* between the subspaces spanned by $\boldsymbol{A}$ and $\boldsymbol{B}$, and by construction, each set of vectors represents an orthonormal basis. Note that $1 \geq s_1 \geq s_2 \geq \ldots \geq s_r \geq 0$. The angles $\cos\theta_k = s_k$ are the *principal angles*, representing the geometric angle between a_k and b_k, or analogously, s_k is the *correlation* between these two vectors. The steps to compute the subspace correlations are as follows [6] (p. 585),

1. If $\boldsymbol{A}$ and $\boldsymbol{B}$ are already orthogonal matrices, we redesignate them as $\boldsymbol{U}_A$ and $\boldsymbol{U}_B$ and skip to Step 2. Otherwise, perform a singular value decomposition (SVD) of $\boldsymbol{A}$, such that $\boldsymbol{A} = \boldsymbol{U}_A \Sigma_A \boldsymbol{V}_A^T$. Similarly decompose $\boldsymbol{B}$. Retain only those components of $\boldsymbol{U}_A$ and $\boldsymbol{U}_B$ that correspond to nonzero singular values, i.e., the number of columns in $\boldsymbol{U}_A$ and $\boldsymbol{U}_B$ correspond to their ranks, and the other matrices are square, with dimension equal to the ranks. If $\boldsymbol{A}$ and $\boldsymbol{B}$ are known to be of full column rank, other orthogonal decompositions, such as QR, are also suitable.

2. Form $\boldsymbol{C} = \boldsymbol{U}_A^T \boldsymbol{U}_B$.

3. If only the correlations are desired, then compute only the singular values of $\boldsymbol{C}$ (the extra computation for the singular vectors is not required). The r ordered singular values $1 \geq s_1 \geq \ldots \geq s_r \geq 0$ are the subspace or principal correlations between $\boldsymbol{A}$ and $\boldsymbol{B}$.

4. If the principal vectors are also desired, then compute the full singular value decomposition, $\boldsymbol{C} = \boldsymbol{U}_C \Sigma_C \boldsymbol{V}_C^T$. The r ordered singular values are extracted from the diagonal of Σ_C. Form the sets principal vectors $\boldsymbol{U}_a = \boldsymbol{U}_A \boldsymbol{U}_C$ and $\boldsymbol{U}_b = \boldsymbol{U}_B \boldsymbol{V}_C$ for sets $\boldsymbol{A}$ and $\boldsymbol{B}$ respectively.

The matrices $\boldsymbol{U}_a$ and $\boldsymbol{U}_b$ are each orthogonal, and the columns comprise the ordered sets of principal vectors for matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ respectively. If both matrices are of the same subspace dimension, the measure $\sqrt{1 - s_r^2} = \sin\theta_r$ is called the *distance* between spaces $\boldsymbol{A}$ and $\boldsymbol{B}$ [6]. When the distance is zero, we see that $\boldsymbol{A}$ and $\boldsymbol{B}$ are parallel subspaces. A maximum distance of unity ($s_r = 0$)indicates at least one basis of $\boldsymbol{A}$ is orthogonal to $\boldsymbol{B}$; if the principal correlation $s_1 = 0$,

then all bases are orthogonal. We see that minimizing the distance is equivalent to maximizing the minimum subspace correlation between $\boldsymbol{A}$ and $\boldsymbol{B}$.

We may also readily compute the specific linear combinations of $\boldsymbol{A}$ and $\boldsymbol{B}$ that yielded these principal vectors and angles. By construction, we know that $\boldsymbol{AX} = \boldsymbol{U}_a$ for some $\boldsymbol{X}$, and $\boldsymbol{X}$ can be simply found using the pseudoinverse of $\boldsymbol{A}$. If we have used the SVD to decompose $\boldsymbol{A}$, then the calculation of $\boldsymbol{X}$ reduces to $\boldsymbol{X} = \boldsymbol{V}_A \Sigma_A^{-1} \boldsymbol{U}_C$; similarly, we compute $\boldsymbol{Y} = \boldsymbol{V}_B \Sigma_B^{-1} \boldsymbol{V}_C$.

The best way to linearly combine the columns of $\boldsymbol{A}$ (i.e. the combination that minimizes the angle of the resulting vector with $\boldsymbol{B}$) is found in the first column of $\boldsymbol{X} \equiv [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r]$ (similarly define $\boldsymbol{Y}$), $\boldsymbol{a}_1 = \boldsymbol{A}\boldsymbol{x}_1$, which is best correlated with $\boldsymbol{B}$ when it is arranged as $\boldsymbol{b}_1 = \boldsymbol{B}\boldsymbol{y}_1$. In other words, there is no other $\boldsymbol{x}$ (excepting a scale factor of $\boldsymbol{x}_1$) for which a corresponding best fitting $\boldsymbol{y}$ will yield a better correlation between $\boldsymbol{a}$ and $\boldsymbol{b}$. The first columns of $\boldsymbol{U}_a$ and $\boldsymbol{U}_b$ are $\boldsymbol{a}_1$ and $\boldsymbol{b}_1$.

Similarly, the worst way to linearly combine $\boldsymbol{A}$ is $\boldsymbol{a}_r = \boldsymbol{A}\boldsymbol{x}_r$. The best fit to this particular $\boldsymbol{x}$ is $\boldsymbol{b}_r = \boldsymbol{B}\boldsymbol{y}_r$, with a correlation of only s_r. No other $\boldsymbol{x}$ will yield a *best* fitting $\boldsymbol{y}$ such that the correlation is *lower*. Hence $\boldsymbol{x}_r$ and $\boldsymbol{y}_r$ yield the "minimum maximum" (minimax) correlation, $s_r = min\{subcorr\{\boldsymbol{A}, \boldsymbol{B}\}\}$.

If two correlations are identical, for instance $s_1 = s_2 = 1$, then the two corresponding vectors $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are themselves arbitrary, but they form a plane such that any linear combination of the two vectors yields a vector whose corresponding correlation is $s_1 = s_2$. By extension, repeated singular values equal to zero have corresponding orientation vectors that are individually arbitrary, but the set forms a subspace in which any vector is orthogonal to $\boldsymbol{B}$.

In E/MEG MUSIC processing, we may compute the subspace correlations between a dipole model and the signal subspace, e.g., $subcorr\{\boldsymbol{G}(\boldsymbol{r}_q), \Phi_s\}$. In this case, the orientations in $\boldsymbol{X}$ represent the dipole orientations. By scaling the first orientation to unity, $\boldsymbol{u}_1 \equiv \boldsymbol{x}_1 / \|\boldsymbol{x}_1\|$, we obtain the unit dipole orientation that best correlates the dipolar source at $\boldsymbol{r}_q$ with the signal subspace. For a two-dipolar topography, $subcorr\{[\boldsymbol{G}(\boldsymbol{r}_{q1}), \boldsymbol{G}(\boldsymbol{r}_{q2})], \Phi_s\}$, then $\boldsymbol{u}_1$ represents the concatenation of the two dipole orientations, $\boldsymbol{u}_1^T = [\boldsymbol{q}_1^T, \boldsymbol{q}_2^T]$, such that the two-dipolar topography

$$[\boldsymbol{G}(\boldsymbol{r}_{q1}), \boldsymbol{G}(\boldsymbol{r}_{q2})]\boldsymbol{u}_1 = \boldsymbol{G}(\boldsymbol{r}_{q1})\boldsymbol{q}_1 + \boldsymbol{G}(\boldsymbol{r}_{q2})\boldsymbol{q}_2 \qquad (50)$$

best correlates with the signal subspace. Consistent with our SPIT model description, we note that the dipole orientations $\boldsymbol{q}_1$ and $\boldsymbol{q}_2$ in *(50)* are themselves not unit vectors, but that their concatenation $\boldsymbol{u}_1$ is.

See [13] for further discussions on subspace correlations and examples of applying them to the problem of E/MEG head modeling.

10.0 References

[1] Achim, A, "Signal detection in averaged evoked potentials: Monte Carlo comparison of the sensitivity of different methods," *Electroenceph. and clin. Neurophys.* Vol. 96:574–584, 1995.

[2] Achim, A, "Cerebral source localization paradigms: spatiotemporal source modeling," *Brain and Cognition,* 27, pp. 256–287, 1995.

[3] Achim A, Richer F, and Saint-Hilaire J, "Methods for separating temporally overlapping sources of neuroelectric data," *Brain Topography,* Vol. 1, no. 1, pp. 22–28, 1988.

[4] Brenner, D., Lipton, J., Kaufman, L., and Williamson, S. J., "Somatically Evoked Magnetic Fields of the Human Brain," *Science*, 199: 81-83, 1978,

[5] Ferrara E, Parks T, "Direction finding with an array of antennas having diverse polarizations," *IEEE Trans. Anten. Prop.* Vol. AP-31, pp. 231–236, Mar. 1983.

[6] Golub GH, Van Loan CF, *Matrix Computations,* second edition, Johns Hopkins University Press, 1984.

[7] Krim H, Viberg M, "Two decades of signal processing: The parametric approach," *IEEE Signal Processing Magazine,* July 1996, Vol. 13, No. 4, pp. 67–94.

[8] Maier J, Dagnelie G, Spekreijse H, and van Dijk B, "Principal components analysis for source localization of VEPs in man," *Vision Research,* vol. 27, no. 2, pp. 165–177, 1987.

[9] Mocks J, Verleger R, "Principal component analysis of event-related potentials: A note on misallocation of variance," *Elec. and clin. Neuro.,* vol. 65, pp. 393–398, 1986.

[10] Mosher JC, Lewis PS, and Leahy RM, "Multiple dipole modeling and localization from spatio-temporal MEG data," *IEEE Trans. Biomedical Eng*, Jun 1992, Vol. 39, pp. 541 – 557.

[11] Mosher JC, Spencer ME, Leahy RM, Lewis PS, "Error bounds for EEG and MEG dipole source localization," *Electroenceph. and clin. Neurophys.* Vol. 86:303–321, June 1993.

[12] Mosher JC, Lewis PS, Leahy RM, "Coherence and MUSIC in biomagnetic source localization" In C. Baumgartner, L, Deecke, G. Stroink, and S. J. Williamson (Eds.), *Biomagnetism: Fundamental Research and Clinical Applications,* Elsevier/IOS Press, Amsterdam, pp. 330–334, 1995.

[13] Mosher, JC, "Subspace Angles: A Metric for Comparisons in EEG and MEG," In Aine, C.J., Flynn, E.R., Okada, Y., Stroink, G., Swithenby, S.J., and Wood, C.C. (Eds.) *Biomag96: Advances in Biomagnetism Research*, Springer-Verlag, New York, 1997.

[14] Mosher, JC, Leahy, RM, Lewis, PS, "Matrix kernels for EEG and MEG source modeling," submitted to *IEEE Trans. Biomedical Eng*, June 1996.

[15] Oh, SK, Un, CK, "A sequential estimation approach for performance improvement of eigenstructure-based methods in array processing," *IEEE Trans. Signal Processing,* Jan. 1993, Vol. 41, No. 1, pp. 457–463.

[16] Scherg M, "Fundamentals of dipole source potential analysis," in *Auditory Evoked Magnetic Fields and Potentials,* vol. 6 (M. Hoke, F. Grandori, and G.L. Romani, eds.) Basel, Karger, 1989.

[17] Scherg M, von Cramon D, "Two bilateral sources of the late AEP as identified by a spatio-temporal dipole model," *Elec. and clin. Neuro.,* vol. 62, pp. 32–44, 1985.

[18] Schmidt, RO "Multiple emitter location and signal parameter estimation," *IEEE Trans. on Ant. and Prop.* vol. AP-34, pp. 276–280, March 1986. Reprint of the original 1979 paper from the *RADC Spectrum Estimation Workshop.*

[19] Shaw, J.C., Roth, M., "Potential Distribution Analysis I: A New Technique for the Analysis of Electrophysiological Phenomena," *Electroencephalography and clinical Neurophysiology,* 273-284, 1955.

[20] Sorenson, H., *Parameter Estimation, Principles and Problems*, Marcel Dekker, Inc., New York, 1980.

[21] Stoica P, Sharman KC, "Maximum likelihood methods for direction-of-arrival estimation," *IEEE Trans. Signal Processing,* July 1990, Vol. 38, No. 7, pp. 1132–1143.

[22] Stoica, P, Handel, P, Nehorai, A, "Improved sequential MUSIC," *IEEE Trans. Aero. Elect. Sys,* Oct. 1995, Vol. 31, No. 4, pp. 1230–1239.

[23] Supek S, Aine CJ, "Simulation studies of multiple dipole neuromagnetic source localization: Model order and limits of source resolution," *IEEE Trans. Biomedical Eng*, Jun 1993, Vol. 40, pp. 529–540.

[24] Tripp, JH, "Physical concepts and mathematical models," in *Biomagnetism: An interdisciplinary approach,* (S.J. Williamson, ed.) pp. 101–149, Plenum Press, 1982.

[25] Viberg M, Ottersten B, "Sensor array processing based on subspace fitting," *IEEE Trans. Signal Processing,* May 1991, Vol. 39, No. 5, pp. 1110–1121.

[26] Viberg M, Ottersten B, Kailath T, "Detection and estimation in sensor arrays using weighted subspace fitting," *IEEE Trans. Signal Processing,* Nov. 1991, Vol. 39, No. 11, pp. 2436–2449.

[27] Viberg M, Swindlehurst AL, "Analysis of the combined effects of finite samples and model errors on array processing performance," *IEEE Trans. Signal Processing,* Nov. 1994, Vol. 42, No. 11, pp. 3073–3083.

[28] Wood CC, "Application of dipole localization methods to source identification of human evoked potentials," *Annals New York Academy Science,* vol. 388, pp. 139–155, 1982.

[29] Wood CC, McCarthy G, "Principal components analysis of event-related potentials: simulation studies demonstrate misallocation of variance across components," *Elec. and clin. Neuro.,* vol. 59, pp. 249–260, 1984.