MEG/EEG Spatiotemporal Noise Covariance Estimation in the Least Squares Sense

Minkyu Ahn* and Sung C. Jun†
* Gwangju Institute of Science and Technology, Gwangju, South Korea
E-mail : frerap@gist.ac.kr
† Gwangju Institute of Science and Technology, Gwangju, South Korea
E-mail : scjun@gist.ac.kr

***Abstract*—A new spatiotemporal noise covariance model were recenlty proposed to consist of the multi-pair Kronecker products of spatial covariance matrices of rank 1 and their corresponding full temporal covariance matrices. Optimized estimators for parameters within this model were derived through the maximum likelihood method and their inversions were accomplished through simple closed formulas, thereby allowing the method to be easily incorporated into parametric source localization methods. However, these maximum likelihood estimators were derived under the assumption that collected spatiotemporal noise samples are Gaussian distributed, which is generally not true for such non-averaged (or single trial) MEG/EEG signals. In this work, an unbiased estimators of spatiotemporal noise covariance in the least squares sense is proposed with using the same multi-pair Kronecker product model without assuming a Gaussian distribution for noise in the data. We found that the least squares covariance estimator for orthogonal spatial components is the same (only differing by constant factors) as the maximum likelihood estimator. However, for independent spatial components the least squares estimator is different from the maximum likelihood estimator.**

I. Introduction

In MEG/EEG experiments, it is common to encounter very low signal-to-noise ratio (SNR) signals due to low signal strength coupled with noise from the system, the brain, external environment and other sources. Brain noise, from ongoing or non-stimulus correlated activity, is often the dominating noise in well shielded MEG/EEG systems and has a complicated structure, with correlations in both time and space. Due to the central limit theorem, the noise of averaged data is expected to be Gaussian distributed and parameterized with a noise covariance matrix (2nd order statistic). This Gaussian noise model is a basic component of most statistically based MEG/EEG source localization approaches and the proper estimation of noise covariance is of importance. It has been reported that noise covariance estimation has a great effect on localization performance in both spatial-only analyses [1] as well as spatiotemporal analyses ([2], [3], [4]). Thus it is crucial to estimate accurately the spatiotemporal noise covariance for good localization performance.

The most general form for the spatiotemporal noise covariance matrix has too many parameters to estimate in practice, while the simple and most commonly used noise model (ignoring the correlation that is present in the background noise) consists of sensor variances without both spatial and temporal correlations. De Munck *et. al* [2] and Huizenga *et. al* [5] proposed new noise covariance models that use a Kronecker product (KP) of the full spatial covariance matrix and full temporal covariance matrix under the assumption that spatial and temporal noise structures are separable. Bijma *et. al* [6] investigated a multi-pair KP model with full spatial covariance matrices and full temporal covariance matrices. However, its inversion is not achievable with ease and thus it is not tractable in source localization methods. Recently, Plis *et. al* [4] proposed the spatiotemporal noise model consisting of multi-pair KP of spatial covariance matrices of rank 1 and full temporal covariance matrices. Invertibility of this model was assured and the model provided better explanations of noise structure than the one pair KP model in studies with simulated and empirical data. However, Plis derived the noise covariance estimators through the maximum likelihood technique under the assumption that non-averaged (single trial) spatiotemporal noise samples are Gaussian distributed. We have ample evidence that single trial or non-averaged spatiotemporal noise samples are likely to be non-Gaussian.

In this work we use the same multi-pair KP model as [4] and derive spatiotemporal noise covariances using the least squares technique. Our derivations do not require the assumption that non-averaged (or single trial) spatiotemporal noise samples are Gaussian distributed.

II. KP Noise Covariance models

In this section, two KP noise covariance models are described briefly and discussed. We assume the following experimental setup:

- The same stimulus is repeated M times.
- Trial-to-trial noise variations of the response are used to estimate covariance.
- Spatiotemporal measurement noise (before stimulus or long after stimulus) at trial number m is $\mathbf{E}_m$ ($L \times K$ matrix) and mean of these noise samples is zero. L is the number of channels and K is the number of time samples.

This work was supported by NRF-2010-32A-B00283, NRF-2009-0071225, and the BioImaging Research Center at GIST.

Thus, the unbiased sample averaged noise covariance can be estimated by

$$\mathbf{C_s} = \frac{1}{M(M-1)} \sum_m vec(\mathbf{E}_m) vec(\mathbf{E}_m)'. \tag{1}$$

Here $vec(\mathbf{E})$ is a column vector stacked over all column vectors of a matrix $\mathbf{E}$, and $'$ denotes the transposition. This sample covariance has $LK(LK+1)/2$ degrees of freedom (DOF)[1]. This requires a large amount of memory to store it and its inversion is very time-consuming, thus it is sometimes intractable on a common workstation. These motivated one to develop simplified noise covariance models which have far less DOF and are inverted efficiently.

Based on the assumption that full rank $K \times K$ temporal ($\mathbf{T}$) and full rank $L \times L$ spatial ($\mathbf{S}$) covariances of background noise are separable, a single pair KP model for averaged noise covariance was proposed [2], [5] :

$$\mathbf{C_{OP}} := \mathbf{T} \otimes \mathbf{S}. \tag{2}$$

The parameters within this model can be estimated through the maximum likelihood method under the assumption that single trial spatiotemporal noise samples are Gaussian distributed. The maximum likelihood method is to find the optimal noise covariance estimator that maximizes the following Gaussian probability density distribution :

$$\max_{\hat{\mathbf{T}},\hat{\mathbf{S}}} \frac{e^{-\frac{1}{2}\sum_m \mathrm{Tr}\{\mathbf{E}_m'\hat{\mathbf{S}}\mathbf{E}_m\hat{\mathbf{T}}\}}}{(2\pi)^{KLM/2}\det(\hat{\mathbf{T}})^{MK/2}\det(\hat{\mathbf{S}})^{ML/2}}. \tag{3}$$

The above optimization yields

$$\hat{\mathbf{S}} = \frac{1}{K}\left(\frac{1}{M}\sum_m \mathbf{E}_m\hat{\mathbf{T}}^{-1}\mathbf{E}_m'\right), \tag{4}$$

$$\hat{\mathbf{T}} = \frac{1}{L}\left(\frac{1}{M}\sum_m \mathbf{E}_m'\hat{\mathbf{S}}^{-1}\mathbf{E}_m\right). \tag{5}$$

Here $\otimes$, Tr$(\cdot)$, and det$(\cdot)$ denote the Kronecker product, trace, and determinant, respectively. These coupled equations (4) and (5) are solved in an iterative way [2].

This one pair KP model (2) is easily inverted by

$$[\mathbf{T} \otimes \mathbf{S}]^{-1} = \mathbf{T}^{-1} \otimes \mathbf{S}^{-1}. \tag{6}$$

Furthermore, this one-pair KP product covariance model tremendously reduces the degrees of freedom (DOF) of spatiotemporal noise covariance, thereby enabling us to estimate a noise covariance in an efficient way when reasonably sufficient noise information is available[2]. Nevertheless, this model has an important problem. Due to its rigidness, it can not well account for some phenomena such unequally distributed alpha activity over the head [6], and trial-to-trial variation [8].

To better explain physiological background activity, capture noise structure as well as to ensure its efficient inversion, another model (multi-pair KP) consisting of a sum of KPs was introduced [4] :

$$\mathbf{C_{MP}} := \sum_{l=1}^{L} \mathbf{T}(l) \otimes \mathbf{S}(l). \tag{7}$$

Here spatial components $\mathbf{S}(l)$ are $L \times L$ matrices of rank 1 and their corresponding temporal matrices $\mathbf{T}(l)$ are full rank $K \times K$ matrices. Comparing to the one-pair KP model, this multi-pair model can better account for different temporal structures over the head. Even though trial-to-trial variations have not been well understood and thus it is not known whether or not they are likely to be statistically stationary, we believe this multi-pair model may better consider trial-to-trial variations.

When spatial components (matrices of rank 1) are independent, that is, all generating vectors $\{v_1, v_2, \cdots, v_L\}$ (spatial component is generated by matrix product of a vector and its transposition) are independent, they can be expressed into $\mathbf{S}(l) = v_l v_l'$ for the l-th column vector v_l of an invertible $L \times L$ matrix $\mathbf{V}$. Thus, due to independence of spatial components, the multi-pair KP model can be inverted as follows [4]:

$$\left(\sum_{l=1}^{L} \mathbf{T}(l) \otimes \mathbf{S}(l)\right)^{-1} = \sum_{l=1}^{L} \mathbf{T}(l)^{-1} \otimes (\mathbf{P}^{-1}\mathbf{S}(l)\mathbf{P}^{-1}). \tag{8}$$

Here $\mathbf{P} := \mathbf{V}\mathbf{V}'$. We remark that linear independence of spatial components of rank 1 ensures the efficient inversion of the multi-pair KP models.

It was reported in [4] that this multi-pair KP model was estimated through the maximum likelihood method under the assumption that noise samples are Gaussian distributed. For the given invertible matrix $\mathbf{V}$ consisting of column vectors $v_1, v_2, \cdots, v_L$, the maximum likelihood covariance estimator yields

$$\begin{aligned} \hat{\mathbf{S}}(l) &= v_l v_l' \\ \hat{\mathbf{T}}(l) &= \frac{1}{M^2}\sum_m \mathbf{E}_m'\mathbf{P}\hat{\mathbf{S}}(l)\mathbf{P}\mathbf{E}_m. \end{aligned} \tag{9}$$

These temporal estimators may account for possibly different temporal characteristics coming from spatially different head regions.

Regarding estimation of spatial components, Plis *et. al* [4] proposed two ways in their recent work. For the spatial orthogonal components, they applied singular value decomposition (SVD) to all accumulated spatial noise data $\mathbf{A_{noise}}$ (having KM kinds of L-dimensional spatial samples):

$$\begin{pmatrix} \mathbf{E}_1' \\ \mathbf{E}_2' \\ \vdots \\ \mathbf{E}_m' \\ \vdots \\ \mathbf{E}_M' \end{pmatrix} \equiv \mathbf{A_{noise}} = \mathbf{U_O}\mathbf{\Sigma}\mathbf{V_O}'. \tag{10}$$

The orthogonal matrix $\mathbf{V_O}$ consists of L orthogonal spatial feature column vectors representing spatial characteristics of

[1] For 100 sensors and 100 time points, it is roughly 10^8.

[2] In general, single trial noise collection without any stimuli for a certain time is required. For very limited noise data, refer to [7].

KM samples. $\mathbf{U_O}$ consists of orthogonal temporal sequence column vectors. Σ is a diagonal matrix having singular values, each of which represents how much of the spatial characteristics its corresponding orthogonal spatial feature vector can account for. In essence, SVD here is a principal component analysis (PCA), that extracts orthogonal spatial features from $\mathbf{A_{noise}}$.

For independent spatial components (mostly unorthogonal), the SOBI (second-order blind identification) ICA (independent component analysis) technique was proposed to extract L spatial features among KM spatial samples as follows:

$$\mathbf{A_{noise}} = \mathbf{U_I}\mathbf{V_I}'. \tag{11}$$

Here $\mathbf{V_I}$ is an $L \times L$ mixing matrix consisting of spatial feature column vectors (linearly independent) and $\mathbf{U_I}$ is a $MK \times L$ matrix having statistically independent temporal sequence column vectors. This mixing matrix $\mathbf{V_I}$ was used to estimate the independent spatial components. We note that the SOBI algorithm was chosen among several ICA methods in this work because it showed good performance on MEG data [9]. Mixing matrices from ICA techniques other than SOBI can be applicable in this derivation.

Just like the one-pair KP model, these multi-pair Kronecker products covariance models significantly reduce the degrees of freedom (DOF) of spatiotemporal noise covariance and enable us to estimate their inversion without significant computational cost. However, the necessary assumption (that non-averaged single trial noise is Gaussian distributed) is generally not realistic for MEG/EEG signals. This motivates us to derive least squares multi-pair KP models without this assumption. We note that Bijma *et. al* [6] investigated the least squares one-pair KP estimator.

III. Multi-pair KP Covariance Estimator in the Least Squares Sense

We assume that $\{v_1, v_2, \cdots, v_L\}$ are linearly independent, that is, v_l is the l-th column vector of an invertible matrix $\mathbf{V}$ of size $L \times L$. For the given spatial components $\hat{\mathbf{S}}^{\mathbf{LS}}(l) = v_l v_l'$, we attempt to estimate the parameters of a multi-pair KP model by fitting it to the sample covariance matrix under the conventional Frobenius norm as follows:

$$\min_{\substack{\text{symmetric } \hat{\mathbf{T}}^{\mathbf{LS}}(l),\\ l=1,\cdots,L}} \left\| \left[\sum_{l=1}^{L} \hat{\mathbf{T}}^{\mathbf{LS}}(l) \otimes \hat{\mathbf{S}}^{\mathbf{LS}}(l) - \mathbf{C_s} \right] \right\|_F^2 \tag{12}$$

Here $vec(\mathbf{E})$ is a column vector stacked over all column vectors of a matrix $\mathbf{E}$. Due to symmetry of covariance matrices, the above Frobenius norm can be expressed in the following:

$$\min_{\substack{\text{symmetric } \hat{\mathbf{T}}^{\mathbf{LS}}(l),\\ l=1,\cdots,L}} \mathrm{Tr}\left(\left[\sum_{l=1}^{L} \hat{\mathbf{T}}^{\mathbf{LS}}(l) \otimes \hat{\mathbf{S}}^{\mathbf{LS}}(l) - \mathbf{C_s} \right]^2 \right) \tag{13}$$

Here $\mathrm{Tr}(\cdot)$ is a trace operator. We seek the optimal $\hat{\mathbf{T}}^{\mathbf{LS}}(l)$ for $l = 1, \cdots, L$ such that (12) is minimized. Due to symmetry of $\hat{\mathbf{T}}^{\mathbf{LS}}(l)$ and $\hat{\mathbf{S}}^{\mathbf{LS}}(l)$, rearranging (12) for $\hat{\mathbf{T}}^{\mathbf{LS}}(l)$ yields

$$\begin{aligned} &\left\| \sum_{l=1}^{L} \hat{\mathbf{T}}^{\mathbf{LS}}(l) \otimes \hat{\mathbf{S}}^{\mathbf{LS}}(l) - \mathbf{C_s} \right\|_F^2 \\ &= \sum_{l=1}^{L} \mathrm{Tr}(\hat{\mathbf{T}}^{\mathbf{LS}}(l)^2)\mathrm{Tr}(\hat{\mathbf{S}}^{\mathbf{LS}}(l)^2) + \mathrm{Tr}(\mathbf{C_s}^2) \\ &- \frac{2}{M(M-1)} \sum_{m,l} \mathrm{Tr}(\mathbf{E}_m' \hat{\mathbf{S}}^{\mathbf{LS}}(l) \mathbf{E}_m \hat{\mathbf{T}}^{\mathbf{LS}}(l)). \end{aligned} \tag{14}$$

Using symmetries of $\hat{\mathbf{S}}^{\mathbf{LS}}(l)$ yields

$$\begin{aligned} \mathrm{Tr}(\hat{\mathbf{S}}^{\mathbf{LS}}(l_1)\hat{\mathbf{S}}^{\mathbf{LS}}(l_2)) &= \mathrm{Tr}(\mathbf{V}e_{l_1}e_{l_1}'\mathbf{V}'\mathbf{V}e_{l_2}e_{l_2}'\mathbf{V}') \\ &= \mathrm{Tr}(\mathbf{V}'\mathbf{V}e_{l_1}e_{l_1}\mathbf{V}'\mathbf{V}e_{l_2}e_{l_2}) \\ &= ((\mathbf{V}'\mathbf{V})_{l_1,l_2})^2 = (\mathbf{Q}_{l_1,l_2})^2, \\ &\quad (\mathbf{Q} := \mathbf{V}'\mathbf{V}) \end{aligned}$$

and differentiation of (14) over $\hat{\mathbf{T}}^{\mathbf{LS}}(l), l = 1, \cdots, L$ gives

$$\begin{aligned} &2\textstyle\sum_{l_1=1}^{L} \hat{\mathbf{T}}^{\mathbf{LS}}(l_1)(\mathbf{Q}_{l,l_1})^2 \\ &\qquad - \frac{2}{M(M-1)} \sum_m \mathbf{E}_m' \hat{\mathbf{S}}^{\mathbf{LS}}(l) \mathbf{E}_m. \end{aligned} \tag{15}$$

Here e_l is a canonical column vector having 1 at the l-th entry and 0 at other entries and $\mathbf{Q}_{l_1,l_2}$ denotes the (l_1, l_2)-th element of the matrix $\mathbf{Q}$. We used the following identities to derive (15):

$$\frac{d\mathrm{Tr}(\mathbf{T}^2)}{d\mathbf{T}} = 2\mathbf{T}', \quad \frac{d\mathrm{Tr}(\mathbf{AT})}{d\mathbf{T}} = \mathbf{A}'. \tag{16}$$

Setting (15) to zero yields

$$\begin{aligned} &\begin{pmatrix} \omega_{1,1}\mathbf{I} & \omega_{1,2}\mathbf{I} & \cdots & \omega_{1,L}\mathbf{I} \\ \omega_{2,1}\mathbf{I} & \omega_{2,2}\mathbf{I} & \cdots & \omega_{2,L}\mathbf{I} \\ \vdots & \vdots & \ddots & \vdots \\ \omega_{L,1}\mathbf{I} & \omega_{L,2} & \cdots & \omega_{L,L}\mathbf{I} \end{pmatrix} \begin{pmatrix} \hat{\mathbf{T}}^{\mathbf{LS}}(1) \\ \hat{\mathbf{T}}^{\mathbf{LS}}(2) \\ \vdots \\ \hat{\mathbf{T}}^{\mathbf{LS}}(L) \end{pmatrix} \\ &= [(\mathbf{Q} \circ \mathbf{Q}) \otimes \mathbf{I}] \begin{pmatrix} \hat{\mathbf{T}}^{\mathbf{LS}}(1) \\ \hat{\mathbf{T}}^{\mathbf{LS}}(2) \\ \vdots \\ \hat{\mathbf{T}}^{\mathbf{LS}}(L) \end{pmatrix} = \begin{pmatrix} \mathbf{R}(1) \\ \mathbf{R}(2) \\ \vdots \\ \mathbf{R}(L) \end{pmatrix} \end{aligned} \tag{17}$$

Here ω_{l_1,l_2} is the element of Hadamard product $\mathbf{Q} \circ \mathbf{Q}$ and $\mathbf{R}(l) := \frac{1}{M(M-1)} \sum_m \mathbf{E}_m' \hat{\mathbf{S}}^{\mathbf{LS}}(l) \mathbf{E}_m$. Solving (17) for $\hat{\mathbf{T}}^{\mathbf{LS}}(l)$ yields

$$\hat{\mathbf{T}}^{\mathbf{LS}}(l) = \sum_{k=1}^{L} d_{l,k}\mathbf{R}(k). \tag{18}$$

Here $d_{l,k}$ is the (l,k)-th element of $(\mathbf{Q} \circ \mathbf{Q})^{-1}$.

In summary, for the given invertible matrix $\mathbf{V}$ consisting of column vectors $v_1, v_2, \cdots, v_L$, the least squares covariance estimator yields

$$\hat{\mathbf{S}}^{\mathbf{LS}}(l) = v_l v_l'$$
$$\hat{\mathbf{T}}^{\mathbf{LS}}(l) = \frac{1}{M(M-1)} \sum_m \mathbf{E}_m' \left[\sum_{k=1}^{L} d_{l,k} \hat{\mathbf{S}}^{\mathbf{LS}}(k) \right] \mathbf{E}_m. \tag{19}$$

Interestingly, this yields an unbiased estimator because the unbiased sample covariance $\mathbf{C_s}$ was used in the LS estimator derivation, however, the maximum likelihood estimator (9) is an inherently biased one. When $\mathbf{V}$ is orthogonal, $\mathbf{P}$ and $\mathbf{Q}$ come to all identities. Thus, these estimators only differ in constant factors $\frac{1}{M-1}$ and $\frac{1}{M}$. This ensures that the maximum likelihood estimator would be good for a large number of trials M regardless of the validity of the Gaussian distribution assumption of noise samples.

For a general invertible matrix $\mathbf{V}$ (independent spatial components), these estimators as shown in (9) and (19) are quite different. However, we can observe that as $\mathbf{Q}$ approaches a diagonal matrix, then the LS estimator approaches the maximum likelihood estimator. In general, seeking the optimal (in some sense) $\mathbf{V}$ would be an interesting problem. We expect that various ICA techniques are likely to yield suboptimal mixing matrices $\mathbf{V}$. That is, the optimality of $\mathbf{V}$ would be related to the criteria of ICA variants.

IV. Source localization - Formulation

The source localization (for the fixed dipole model) is now done by incorporating the estimated multi-pair KP noise covariance $\hat{\mathbf{C}}_{\mathbf{MP}}$ into the generalized negative log-likelihood function (called cost function):

$$\begin{aligned} c(\theta, \mathbf{J}) &= [vec(\mathbf{F}(\theta)\mathbf{J}) - vec(\mathbf{Y})]' \hat{\mathbf{C}}_{\mathbf{MP}}^{-1} [vec(\mathbf{F}(\theta)\mathbf{J}) \\ &\qquad -vec(\mathbf{Y})] \\ &\quad \left(\hat{\mathbf{C}}_{\mathbf{MP}}^{-1} := \sum_{l=1}^{L} \hat{\mathbf{T}}^{\mathbf{LS}}(l)^{-1} \otimes \mathbf{P}^{-1} \hat{\mathbf{S}}^{\mathbf{LS}}(l) \mathbf{P}^{-1} \right) \\ &= \sum_{l=1}^{L} \mathrm{Tr} \Big([\mathbf{F}(\theta)\mathbf{J} - \mathbf{Y}]' \mathbf{P}^{-1} \hat{\mathbf{S}}^{\mathbf{LS}}(l) \mathbf{P}^{-1} [\mathbf{F}(\theta)\mathbf{J} \\ &\qquad -\mathbf{Y}] \hat{\mathbf{T}}^{\mathbf{LS}}(l)^{-1} \Big). \end{aligned} \tag{22}$$

Here $\theta, \mathbf{J}$, $\mathbf{F}(\theta)$, and $\mathbf{Y}$ are source location and its orientation, a matrix ($N \times K$) representing source time course, the forward lead field matrix, and a spatiotemporal measurement matrix ($L \times K$), respectively. N, K, and L are the expected number of sources, the number of time points in the temporal window, and the number of sensors, respectively.

To minimize the above cost function, differentiating (22) over $\mathbf{J}$ and setting it to zero yields

$$\sum_{l=1}^{L} \Big(\mathbf{F}(\theta)' \mathbf{P}^{-1} \hat{\mathbf{S}}^{\mathbf{LS}}(l) \mathbf{P}^{-1} \mathbf{F}(\theta) \mathbf{J} \hat{\mathbf{T}}^{\mathbf{LS}}(l)^{-1} - \mathbf{F}(\theta)' \mathbf{P}^{-1} \hat{\mathbf{S}}^{\mathbf{LS}}(l) \mathbf{P}^{-1} \mathbf{Y} \hat{\mathbf{T}}^{\mathbf{LS}}(l)^{-1} \Big) = 0. \tag{23}$$

Solving for $\mathbf{J}$ gives the optimal $\mathbf{J^{opt}}$ (20). Here we used the identity

$$vec(\mathbf{ABC}) = (\mathbf{C}' \otimes \mathbf{A}) vec(\mathbf{B}). \tag{24}$$

Inserting (20) into (22) yields (21). While the one-pair KP model requires the inversion of the $N \times N$ matrix [5], this evaluation requires the inversion of a $KN \times KN$ matrix, which depends on the number of time points K. This can be tractable for common multi-dipole source problems (several dipole sources and up to several hundreds time points) on common workstations.

The covariance matrix of the source parameters $\psi = \{\theta, \mathbf{J}\}$ at the solution $\hat{\psi}$ can be estimated ([10], [5]) by

$$\mathbf{C}_{\psi} = 2\sigma^2 \left[E(\mathbf{H}) \right]^{-1}. \tag{25}$$

Here $\sigma = c(\hat{\psi})/(LK - 6N - NK)$. $\mathbf{H}$ is the Hessian matrix with second-order partial derivatives of (22). $E[\cdot]$ means the expectation. Taking the expectation of the Hessian $E(\mathbf{H})$, we obtain $E(\mathbf{H})_{i,j}$ as

$$2 \,\mathrm{Tr} \left(\sum_{l=1}^{L} \hat{\mathbf{T}}^{\mathbf{LS}}(l)^{-1} \left[\frac{\partial \mathbf{F}(\hat{\theta}) \hat{\mathbf{J}}}{\partial \psi_i} \right]' \mathbf{P}^{-1} \hat{\mathbf{S}}^{\mathbf{LS}}(l) \mathbf{P}^{-1} \left[\frac{\partial \mathbf{F}(\hat{\theta}) \hat{\mathbf{J}}}{\partial \psi_j} \right] \right). \tag{26}$$

This is a generalization of the result in [5].

V. Discussion

According to the experiments published in [11], there is evidence that the background activity of the brain has a stationary spatial distribution over the time of interest. Relying on such evidence, we made assumptions about the background activity in the brain:

- The measured background is a superposition of some spatially fixed sources with independent temporal behavior.
- Each spatial source produces correlated Gaussian noise.

These are the bases for our multi-pair KP spatiotemporal noise covariance modeling. In order to extract spatially fixed noise generators we find spatial orthogonal components of rank 1. The sum of the KP of these components with their corresponding temporal covariances constitutes our multi-pair model. However, higher rank spatial components can be considered to maintain the invertibility of the covariance model, which can be a generalization of multi-pair KP models [12]. Given this reasoning, in MEG/EEG source localization problems it may be unreasonable to consider another multi-pair KP model consisting of the temporal components of rank 1 and their corresponding full spatial matrices even though it may make sense mathematically.

Each ML temporal covariance estimator should be inherently a sample noise covariance computed from projected noise samples onto its corresponding spatial component space. This holds for any spatial components (orthogonal or independent). We found LS temporal covariance for orthogonal spatial components was such a case, but it was not for independent spatial components. The difference between ML and LS estimators has motivated us to conduct a localization

$$vec(\mathbf{J^{opt}}) = \left(\sum_l \hat{\mathbf{T}}^{\mathbf{LS}}(l)^{-1} \otimes \left[\mathbf{F}(\theta)'\mathbf{P}^{-1}\hat{\mathbf{S}}^{\mathbf{LS}}(l)\mathbf{P}^{-1}\mathbf{F}(\theta)\right]\right)^{-1} vec\left(\sum_l \mathbf{F}(\theta)'\mathbf{P}^{-1}\hat{\mathbf{S}}^{\mathbf{LS}}(l)\mathbf{P}^{-1}\mathbf{Y}\hat{\mathbf{T}}^{\mathbf{LS}}(l)^{-1}\right), \quad (20)$$

$$c(\theta, \mathbf{J^{opt}}) = \mathrm{Tr}\left[\sum_l \mathbf{P}^{-1}\hat{\mathbf{S}}^{\mathbf{LS}}(l)\mathbf{P}^{-1}\mathbf{Y}\hat{\mathbf{T}}^{\mathbf{LS}}(l)^{-1}\mathbf{Y}'\right] - vec\left(\sum_l \mathbf{F}(\theta)'\mathbf{P}^{-1}\hat{\mathbf{S}}^{\mathbf{LS}}(l)\mathbf{P}^{-1}\mathbf{Y}\hat{\mathbf{T}}^{\mathbf{LS}}(l)^{-1}\right)'$$
$$\times \left(\sum_l \hat{\mathbf{T}}^{\mathbf{LS}}(l)^{-1} \otimes \left[\mathbf{F}(\theta)'\mathbf{P}^{-1}\hat{\mathbf{S}}^{\mathbf{LS}}(l)\mathbf{P}^{-1}\mathbf{F}(\theta)\right]\right)^{-1} vec\left(\sum_l \mathbf{F}(\theta)'\mathbf{P}^{-1}\hat{\mathbf{S}}^{\mathbf{LS}}(l)\mathbf{P}^{-1}\mathbf{Y}\hat{\mathbf{T}}^{\mathbf{LS}}(l)^{-1}\right). \quad (21)$$

performance comparison on both simulated and empirical data, currently in progress.

In our derivations of LS estimator, the spatial components (coming from $\mathbf{V}$) are assumed to be given. This independent (invertible) matrix ($\mathbf{V}$) can be different for different optimality criteria. For orthogonal matrix V, SVD is used to estimate it. ICA methods like SOBI may be applicable to estimate independent matrix V. Particularly, regarding ICA variants, it would be interesting to investigate which ICA variant yields the best (in some sense) LS estimator.

We first formulated the cost functions incorporating multi-pair KP covariance models. Comparing the one-pair KP model in [5], the evaluations of the cost functions incorporating the multi-pair models are more computationally expensive than evaluation of the cost function for the one-pair model because they require the $KN \times KN$ matrix inversion. The rapid growth of computational power may enable such computations even for complicated problems like those with tens of dipoles and thousands of time points. Although more complex and time consuming, inclusion of the multi-pair KP noise covariance model in MEG/EEG least squares inverse analyses will likely provide better source localization and time course fitting than analyses using simpler noise covariance estimators.

Comparative study between ML estimator and LS estimator should be investigated in terms of source localization perfor-mace. Such a study is under investigation and will be reported in the subsequent paper.

ACKNOWLEDGMENT

We thank Sergey Plis (MIND Institute), David Schmidt, John George and Petr L. Volegov (LANL) for their valuable discussions and support in initiating this work.

REFERENCES

[1] S. C. Jun, B. A. Pearlmutter, and G. Nolte, “Fast accurate meg source localization using a multilayer perceptron trained with real brain noise,” *Physics in Medicine and Biology*, vol. 47, no. 14, p. 2547, 2002.

[2] J. C. De Munck, H. M. Huizenga, L. J. Waldorp, and R. Heethaar, “Estimating stationary dipoles from MEG/EEG data contaminated with spatially and temporally correlated background noise,” *IEEE Trans Biomed. Engin.*, vol. 50, pp. 1565–1572, 2002.

[3] S. C. Jun, J. S. George, W. Kim, J. Pare-Blagoev, S. Plis, D. M. Ranken, and D. M. Schmidt, “Bayesian brain source imaging based on combined MEG/EEG and fMRI using MCMC,” *NeuroImage*, vol. 40, no. 4, pp. 1581 – 1594, 2008.

[4] S. M. Plis, J. S. George, S. C. Jun, J. Pare-Blagoev, D. M. Ranken, D. M. Schmidt, and C. C. Wood, “Modeling spatiotemporal covariance for MEG/EEG source analysis,” *Phys. Rev. E*, vol. 75, p. 011928, 2007.

[5] H. M. Huizenga, J. C. de Munck, L. J. Waldorp, and R. P. P. P. Grasman, “Spatiotemporal EEG/MEG source analysis based on a parametric noise covariance model,” *IEEE Trans. Biomed. Engin.*, vol. 49, pp. 533–539, 2002.

[6] F. Bijma, J. C. de Munck, and R. M. Heethaar, “The spatiotemporal MEG covariance matrix modeled as a sum of kronecker products,” *NeuroImage*, vol. 27, pp. 402–415, 2005.

[7] S. C. Jun, S. M. Plis, D. M. Ranken, and D. M. Schmidt, “Spatiotemporal noise covariance estimation from limited empirical MEG data,” *Physics in Medicine and Biology*, vol. 51, pp. 5549–5564, 2006.

[8] J. C. de Munck, F. Bijma, P. Gaura, C. A. Sieluzycki, M. I. Branco, , and R. M. Heethaar, “A maximum-likelihood estimator for trial-to-trial variations in noisy MEG/EEG data sets,” *IEEE Trans Biomed. Engin.*, vol. 51, pp. 2123–2128, 2004.

[9] A. Tang, B. Pearlmutter, N. Malaszenko, D. Phung, and B. Reeb, "Independent components of magnetoencephalography: localization." *Neural Computation*, vol. 14, no. 8, pp. 1827–58, 2002.

[10] H. M. Huizenga and P. C. M. Molenaar, "Ordinary least squares dipole localization is influenced by the reference," *Electroenceph. clin. Neurophysiol.*, vol. 99, pp. 562–567, 1996.

[11] M. E. Raichle, A. M. MacLeod, A. Z. Snyder, W. J. Powers, D. A. Gusnard, and G. L. Shulman, "A default mode of brain function," *PNAS*, vol. 98, pp. 676–682, 2001.

[12] S. Plis, D. Schmidt, S. Jun, and D. Ranken, "A generalized spatiotemporal covariance model for stationary background in analysis of MEG data," in *Engineering in Medicine and Biology Society, 2006. EMBS '06. 28th Annual International Conference of the IEEE*, 30 2006-sept. 3 2006, pp. 3680 –3683.