A generalized spatiotemporal covariance model
for stationary background in analysis of MEG data

Plis S.M., *Student Member, IEEE*, Schmidt D.M., Jun S.C., *Member, IEEE*, Ranken D.M.

Abstract— Using a noise covariance model based on a single Kronecker product of spatial and temporal covariance in the spatiotemporal analysis of MEG data was demonstrated to provide improvement in the results over that of the commonly used diagonal noise covariance model. In this paper we present a model that is a generalization of all of the above models. It describes models based on a single Kronecker product of spatial and temporal covariance as well as more complicated multi-pair models together with any intermediate form expressed as a sum of Kronecker products of spatial component matrices of reduced rank and their corresponding temporal covariance matrices. The model provides a framework for controlling the tradeoff between the described complexity of the background and computational demand for the analysis using this model. Ways to estimate the value of the parameter controlling this tradeoff are also discussed.

I. INTRODUCTION

Given that MEG/EEG is measured in M trials, on L sensors and in C time samples, let $\mathbf{E}_m$ be the L by C single trial noise matrix at trial m. And to simplify our notation further we assume that the background in $\mathbf{E}_m$ is zero-mean. In this case, the sample spatiotemporal covariance matrix of dimension $N = LC$ for the averaged over trials noise is

$$\mathbf{C_s} = \frac{1}{M-1} \sum_{m=1}^{M} vec(\mathbf{E}_m) vec(\mathbf{E}_m)^T, \qquad (1)$$

where $vec(\mathbf{E})$ is all the columns of $\mathbf{E}$ stacked in a vector.

Due to the tremendously large number of parameters ($O(L^2C^2)$) compared to the relative sparsity of background data typically available many attempts to model the covariance has been made. In order to be useful for source analysis a model of spatiotemporal covariance should capture the structure of the background with as few parameters as possible. Finally to be useful the model should be invertible, so it can be used in a likelihood formulation.

To capture the structure in the brain background signal we note that there is evidence that the background activity of the brain has a stationary spatial distribution over the time of interest in the neuroscientific experiments [1]. This

observation provides for the current approach in exploratory fMRI analysis to represent four dimensional spatiotemporal data as a two dimensional matrix $\mathbf{X}$, which is decomposed into a sum of Kronecker products contaminated by normal noise [2]:

$$\mathbf{X} = \sum_{r}^{R} \mathbf{a_r} \otimes \mathbf{b_r} + \mathbf{E}. \qquad (2)$$

Relying on such evidence, we can make assumptions about the background activity in the brain in the MEG (magnetoencephalography) measurements:

1) The measured background is a superposition of R spatially fixed sources with independent temporal behavior.
2) Each spatial source produces correlated Gaussian noise.

These are the basis for modeling spatiotemporal covariance adopted in this work. The final goal is to find spatial components according to a justified criteria and then estimate respective temporal covariances for each component. For L sensors performing the measurement we can extract at most L spatial components from the measured background, which gives us the starting form for the spatiotemporal covariance matrix:

$$\sum_{l=1}^{L} \boldsymbol{\mathcal{T}}_l \otimes \boldsymbol{\mathcal{S}}_l, \qquad (3)$$

where $\boldsymbol{\mathcal{S}}_l = v_l v_l^T$ is a spatial component represented as a singular matrix of rank 1.

In what follows we demonstrate how existing spatiotemporal covariance models based on the Kronecker product can be described in the proposed framework. This includes the models based on one Kronecker product suggested in [3], [4] and on a series of Kronecker products introduced by us [5]. Finally we extend the framework by making it span the range of models between one pair Kronecker models and L component models.

II. ONE-PAIR KRONECKER PRODUCT

In [3] was proposed a model that casts spatiotemporal covariance in the form that uses the Kronecker product:

$$\mathbf{COV} \approx \mathbf{T} \otimes \mathbf{S}. \qquad (4)$$

Temporal covariance is a $C \times C$ matrix $\mathbf{T}$ and spatial covariance is a $L \times L$ matrix $\mathbf{S}$, where C is the number of time samples and L is the number of sensors.

An important feature of this model is its computationally fast inverse calculation $(\mathbf{T} \otimes \mathbf{S})^{-1} = \mathbf{T}^{-1} \otimes \mathbf{S}^{-1}$. This feature

is required of all covariance models to make them useful in the analysis. All models presented in this paper are invertible and manageable within reasonable time.

Since the background is assumed to have a Gaussian distribution the model can be estimated in the maximum likelihood framework as in [3].

$$\mathbf{S} = \frac{1}{C}\left(\frac{1}{M}\sum_{m=1}^{M}\boldsymbol{E}_m\mathbf{T}^{-1}\boldsymbol{E}_m^T\right), \tag{5}$$

$$\mathbf{T} = \frac{1}{L}\left(\frac{1}{M}\sum_{m=1}^{M}\boldsymbol{E}_m\mathbf{S}^{-1}\boldsymbol{E}_m^T\right).$$

Here M is the number of available $L \times C$ spatiotemporal measurements $\boldsymbol{E}_m$.

In essence, the main assumption of the one-pair Kronecker product model is that all spatial components of the background have one and the same temporal covariance structure. To demonstrate the validity of this statement we perform spectral decomposition of the spatial covariance:

$$\mathbf{T} \otimes \sum_{l=1}^{L}\sigma_l\boldsymbol{S}_l, \tag{6}$$

where $\boldsymbol{S}_l = v_l v_l^T$ is an orthonormal basis component represented as a singular matrix. Using the identity $\mathbf{A}\otimes(\mathbf{B}+\mathbf{C}) = \mathbf{A}\otimes\mathbf{B}+\mathbf{A}\otimes\mathbf{C}$, the expression (6) can be represented as:

$$\sum_{l=1}^{L}\mathbf{T}\otimes\sigma_l\boldsymbol{S}_l = \sum_{l=1}^{L}\sigma_l\mathbf{T}\otimes\boldsymbol{S}_l. \tag{7}$$

The Left Hand Side of the equation (7) makes it obvious that each spatial component has the same temporal covariance. The contribution of each temporal covariance is weighted by the variance of the corresponding orthonormal spatial component, as seen from the Right Hand Side of (7). When we set $\sigma_l\mathbf{T} = \boldsymbol{\mathcal{T}}_l$ we see how our general model from (3) subsumes the one pair model.

A parameterized model of the same form as in (4) was introduced in order to achieve a further reduction in the number of parameters [4]. This model is also a special case of (7). To see that it is enough to follow the same logic as we just did for the unparameterized model.

III. Multi-pair models

In a more realistic case it is easy to picture a situation where several noise generators belonging to separate spatial components also have different temporal structures.

Two models covering such case were introduces in [6] one in which spatial components $\boldsymbol{S}_l$ are components of an orthonormal basis and the other with spatial components from an independent basis. The following is assumed:

1) Spatiotemporal noise is generated by L spatially orthogonal (or independent) generators which do not change their location during the period of interest.
2) Each spatial component is uncorrelated with other components.
3) Measured signal has a Gaussian distribution with zero mean.

Both models have the form as in (3) and any set of orthogonal or independent components can be used. As demonstrated in [6] estimation of orthogonal components through singular value decomposition and of independent components through an independent component analysis algorithm work well.

For given orthogonal components $\boldsymbol{S}_l = v_l v_l^T$ with v_l being a column of an orthogonal matrix $\boldsymbol{V}$, Maximal Likelihood estimate of the model is:

$$\mathbf{COV} \approx \widetilde{\mathbf{COV}} \equiv \sum_{l=1}^{L}\boldsymbol{\mathcal{T}}_l\otimes\boldsymbol{S}_l, \tag{8}$$

$$\boldsymbol{\mathcal{T}}_l = \frac{1}{M}\sum_{m=1}^{M}\mathbf{E}_m^T\boldsymbol{S}_l\mathbf{E}_m.$$

For given independent components $\boldsymbol{\mathcal{R}}_l = w_l w_l^T$ with w_l being a row of a full rank matrix $\boldsymbol{W}^{-1}$, Maximal Likelihood estimate of the model is:

$$\mathbf{COV} \approx \widetilde{\mathbf{COV}} \equiv \sum_{l=1}^{L}\boldsymbol{\mathcal{T}}_l\otimes\boldsymbol{\mathcal{R}}_l, \tag{9}$$

$$\boldsymbol{\mathcal{T}}_l = \frac{1}{M}\sum_{m=1}^{M}\mathbf{E}_m^T\mathbf{W}\mathbf{W}^T\boldsymbol{\mathcal{R}}_l\mathbf{W}\mathbf{W}^T\mathbf{E}_m.$$

These models are more general than the model of Section II and the independent basis multi-pair model of (9) is the most general in terms of having the largest number of degrees of freedom. This increased complexity brings advantages in being able to more accurately model the complex empirical noise covariance, which lead to improved source localization performance as demonstrated in [6]. Furthermore, inversion is computationally efficient due to the built in structure. For the orthogonal basis model:

$$\widetilde{\mathbf{COV}}^{-1} = \sum_{l=1}^{L}(\boldsymbol{\mathcal{T}}_l)^{-1}\otimes\boldsymbol{S}_l. \tag{10}$$

And for the independent basis model:

$$\widetilde{\mathbf{COV}}^{-1} = \sum_{l=1}^{L}(\boldsymbol{\mathcal{T}}_l)^{-1}\otimes[\mathbf{W}\mathbf{W}^T\boldsymbol{\mathcal{R}}_l\mathbf{W}\mathbf{W}^T]. \tag{11}$$

Details of the derivation can be found in [6].

IV. Generalized multi-pair model

Even though the models in Section III are more descriptive and yield better results than single-pair models, the problem with them is the computational complexity of each iteration in the analysis where the likelihood needs to be computed. The negative log-likelihood function in the case of using a multipair model looks like:

$$\sum_{l=1}^{L}(\boldsymbol{E}_b - \boldsymbol{E}_\mathcal{M})^T\boldsymbol{S}_l(\boldsymbol{E}_b - \boldsymbol{E}_\mathcal{M})\boldsymbol{\mathcal{T}}_l^{-1}. \tag{12}$$

Here $\boldsymbol{E}_b$ and $\boldsymbol{E}_\mathcal{M}$ are $L \times C$ matrices of spatiotemporal measurements and the current prediction made by the model, respectively. Calculation of (12) takes $O(L(L^2C + LC^2 +$

C^3)) operations and it is L times bigger than for one-pair models.

Yet, notice that multi-pair models utilize all spatial components which in the case of the orthogonal basis model also includes those components that have small singular values. Without loss of generality we consider first only the orthogonal basis multi-pair model. Separating the first $r \leq L$ most significant singular values, the model can be expressed as

$$\sum_{l=1}^{r} \boldsymbol{T}_l \otimes \boldsymbol{S}_l + \sum_{k=r+1}^{L} \boldsymbol{T}_k \otimes \boldsymbol{S}_k. \qquad (13)$$

The second term in expression (13) contains small variance relative to the first one. Conventional dimensionality reduction techniques like PCA would eliminate this term from further consideration. Though this would make the model smaller in terms of storage requirements it will at the same time render it useless in the analysis. The inversion following the lines of (10) would become impossible. At the same time if we set temporal covariance of all spatial components with small singular values to be the same as it is assumed in the one-pair model (7) we obtain the following expression:

$$\sum_{l=1}^{r} \boldsymbol{T}_l \otimes \boldsymbol{S}_l + \boldsymbol{T} \otimes \sum_{k=r+1}^{L} \boldsymbol{S}_k. \qquad (14)$$

Expression (14) is a general form describing spatiotemporal covariance models based on the assumption that the brain has stationary spatial distribution of the background activity over the time of interest [1]. It can describe the one-pair model when r is set to zero and multi-pair models when $r = L$.

A feature absolutely required from this model is that it be invertible. Otherwise, the model cannot be used in the Likelihood function (12) for source localization. Indeed, this model has a computationally manageable inverse with calculation time depending on r and ranging between the times for the one-pair model and the multi-pair models. For the orthogonal basis multi-pair model the inverse is:

$$\sum_{l=1}^{r} \boldsymbol{T}_l^{-1} \otimes \boldsymbol{S}_l + \boldsymbol{T}^{-1} \otimes \sum_{k=r+1}^{L} \boldsymbol{S}_k. \qquad (15)$$

To prove this claim it suffices to show that:

$$\left(\sum_{l=1}^{r} \boldsymbol{T}_l^{-1} \otimes \boldsymbol{S}_l + \boldsymbol{T}^{-1} \otimes \sum_{k=r+1}^{L} \boldsymbol{S}_k \right) \qquad (16)$$
$$\times \left(\sum_{l=1}^{r} \boldsymbol{T}_l \otimes \boldsymbol{S}_l + \boldsymbol{T} \otimes \sum_{k=r+1}^{L} \boldsymbol{S}_k \right) = \boldsymbol{I}.$$

To proceed we need the following properties of $\boldsymbol{S}_l$:

- property (1)

$$\begin{aligned}
(\boldsymbol{S}_l)^2 &= (v_l v_l^T)(v_l v_l^T) = v_l(v_l^T v_l)v_l^T \\
&= v_l v_l^T \quad (v_l^T v_l = 1 \text{ by orthogonality}) \\
&= \boldsymbol{S}_l.
\end{aligned}$$

- property (2) : $\boldsymbol{I} = \mathbf{V} \, \mathbf{V}^T = \sum_l v_l v_l^T$.
- property (3) : $\boldsymbol{S}_l \boldsymbol{S}_{l'} = \delta(l, l')\boldsymbol{S}_l$, where $\delta(l, l')$ is the Kronecker delta.

Using these properties and performing the multiplication in the LHS of (16):

$$\sum_{l=1}^{r} \boldsymbol{I} \otimes \boldsymbol{S}_l + \boldsymbol{I} \otimes \sum_{k=r+1}^{L} \boldsymbol{S}_k \qquad (17)$$
$$= \boldsymbol{I} \otimes \sum_{l=1}^{r} \boldsymbol{S}_l + \boldsymbol{I} \otimes \sum_{k=r+1}^{L} \boldsymbol{S}_k$$
$$= \boldsymbol{I} \otimes \sum_{l=1}^{L} \boldsymbol{S}_l$$

and using property (2)

$$= \boldsymbol{I} \otimes \boldsymbol{I} = \boldsymbol{I}.$$

For the independent basis multi-pair model if we replace orthogonal components $\boldsymbol{S}_l$ by independent components $\mathcal{R}_l$ the inverse is:

$$\sum_{l=1}^{r} \boldsymbol{T}_l^{-1} \otimes \boldsymbol{W}\boldsymbol{W}^T \mathcal{R}_l \boldsymbol{W}\boldsymbol{W}^T \qquad (18)$$
$$+ \boldsymbol{T}^{-1} \otimes \sum_{k=r+1}^{L} \boldsymbol{W}\boldsymbol{W}^T \mathcal{R}_k \boldsymbol{W}\boldsymbol{W}^T.$$

The proof proceeds in the similar fashion as for the orthogonal basis model.

Computational efficiency of this new generalized spatiotemporal covariance model is now a function of r and not fixed to the number of sensors, as in the case of models from Section III. Here again, without loss of generality we demonstrate results using the orthogonal basis model. We redefine $\sum_{k=r+1}^{L} \boldsymbol{S}_k = \boldsymbol{S}$. The negative log-likelihood function with the use of generalized covariance is then expressed as:

$$\sum_{l=1}^{r} (\boldsymbol{E}_b - \boldsymbol{E}_{\mathcal{M}})^T \boldsymbol{S}_l (\boldsymbol{E}_b - \boldsymbol{E}_{\mathcal{M}}) \boldsymbol{T}_l^{-1} \qquad (19)$$
$$+ (\boldsymbol{E}_b - \boldsymbol{E}_{\mathcal{M}})^T \boldsymbol{S} (\boldsymbol{E}_b - \boldsymbol{E}_{\mathcal{M}}) \boldsymbol{T}^{-1}.$$

The computational complexity of this calculation is $O((r + 1)(L^2C + LC^2 + C^3))$ and is flexible since it is parameterized by r. It can be as large as the one of (12) when the background is diverse enough to need a complex representation and can be as small as that of the one-pair model.

The generalized orthogonal basis model from (14) can be estimated in a relatively straightforward manner based on the nature of PCA. The number of single Kronecker product pairs of our model $r + 1$ is obtained by choosing the most significant components provided by singular value decomposition. This is a well justified approach and it follows the conventional method of dimensionality reduction in PCA. After that the model can be estimated in the Maximal Likelihood framework. Before proceeding with its derivation we need to calculate the determinant of (14). It can be found

from the determinant of the orthogonal basis multipair model derived in [6]:

$$| \sum \boldsymbol{T}_l \otimes \boldsymbol{S}_l | = \prod_l^L |\boldsymbol{T}_l|. \tag{20}$$

Noticing that the last $J = L - r$ temporal covariances in the product are the same, and are $\boldsymbol{T}$, we get:

$$| \sum_{l=1}^r \boldsymbol{T}_l \otimes \boldsymbol{S}_l + \boldsymbol{T} \otimes \boldsymbol{S} | = \boldsymbol{T}^J \prod_l^r |\boldsymbol{T}_l|. \tag{21}$$

The log-likelihood function $\mathcal{L}$ is then written as:

$$const - \frac{M}{2} \left(J \ln |\boldsymbol{T}| + \sum_{l=1}^r \ln |\boldsymbol{T}_l| \right) \tag{22}$$

$$- \frac{1}{2} tr \left(\sum_{m=1}^M \left(\sum_{l=1}^r \mathbf{E}_m^T \boldsymbol{S}_l \mathbf{E}_m (\boldsymbol{T}_l)^{-1} + \mathbf{E}_m^T \boldsymbol{S} \mathbf{E}_m \boldsymbol{T}^{-1} \right) \right).$$

Differentiating with respect to $\boldsymbol{T}$ results in:

$$d\mathcal{L} = - \frac{MJ}{2} tr((\boldsymbol{T})^{-1} d\boldsymbol{T}) \tag{23}$$

$$+ \frac{1}{2} tr \left(\sum_{m=1}^M \mathbf{E}_m^T \boldsymbol{S} \mathbf{E}_m (\boldsymbol{T})^{-1} d\boldsymbol{T} (\boldsymbol{T})^{-1} \right)$$

$$= - \frac{MJ}{2} tr((\boldsymbol{T})^{-1} d\boldsymbol{T})$$

$$+ \frac{1}{2} tr \left(\sum_{m=1}^M (\boldsymbol{T})^{-1} \mathbf{E}_m^T \boldsymbol{S} \mathbf{E}_m (\boldsymbol{T})^{-1} d\boldsymbol{T} \right)$$

$$= - \frac{MJ}{2}$$

$$tr \left(\left[(\boldsymbol{T}^l)^{-1} - \frac{1}{M} \sum_{m=1}^M (\boldsymbol{T})^{-1} \mathbf{E}_m^T \boldsymbol{S} \mathbf{E}_m (\boldsymbol{T})^{-1} \right] d\boldsymbol{T} \right)$$

And the final result is:

$$\boldsymbol{T} = \frac{1}{MJ} \sum_{m=1}^M \mathbf{E}_m^T \boldsymbol{S} \mathbf{E}_m. \tag{24}$$

Together with the equation (8) we can estimate the generalized model as a sum of r Kronecker products of rank one spatial component matrices with their corresponding temporal covariance matrices plus a Kronecker product of a spatial component matrix of rank J with its corresponding temporal covariance matrix.

This approach will not work for the independent basis model. There are no singular values that can be used for thresholding. However, we can use the variance of the data projected into each independent spatial component to perform thresholding based on its value. Then a similar Maximal Likelihood estimation as in (9) can be used for the generalized independent basis model. This would add asymmetry to the way these two models are estimated.

Another possible approach can be based on observations made in [5]. Many temporal covariances of different spatial components were found to be similar. By discovering r clusters of similar temporal covariances we can obtain r Kronecker products of the generalized multi-pair model. This approach is applicable both for orthogonal and independent basis models. A similarity criterion could be developed for that case, such that it provides clustering of components in a manner that balances optimizing source localization with reducing computational time. In this second approach, the resulting model will be different from the one obtained by thresholding of significant components. The model is then described as a sum of Kronecker products of spatial component matrices of incomplete rank and their corresponding temporal covariances. The difference is that there is no requirement of having rank one component matrices and having only one spatial component matrix of a bigger rank as in (14).

It is to be determined which of these two approaches shows better localization performance. In future work we plan to answer this question and investigate which clustering algorithm to use for obtaining summands in the second approach.

REFERENCES

[1] 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." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 2, pp. 676 – 682, 2001.

[2] C. F. Beckmann and S. M. Smith, "Tensorial extensions of independent component analysis for multisubject fmri analysis." *NeuroImage*, vol. 25, no. 1, pp. 294 – 311, 2005.

[3] 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 Transaction on Signal Processing*, vol. 50, no. 7, pp. 1565–1572, July 2002.

[4] H. M. Huizenga, J. C. De Munck, L. J. Waldorp, and R. P. Grasman, "Spatiotemporal EEG/MEG source analysis based on a parametric noise covariance model," *IEEE Trans Biomed. Engin.*, vol. 50, pp. 533–539, 2002.

[5] S. Plis, J. S. George, S. C. Jun, J. Pare-Blagoev, D. M. Ranken, D. M. Schmidt, and C. C. Wood, "Realistic spatiotemporal noise covariance model for MEG/EEG source analysis," *Brain Topography*, vol. 16, 2004.

[6] S. M. Plis, J. S. George, S. C. Jun, J. Pare-Blagoev, D. M. Ranken, D. M. Schmidt, and C. C. Wood, "Modeling spatiotemporal noise covariance for MEG/EEG source analysis," 2005. [Online]. Available: http://www.arXiv.org:physics/0503063