A space–frequency analysis of MEG background processes

Fetsje Bijma a,⁎, Jan C. de Munck b

a VU University, Faculty of Science, Department of Mathematics, De Boelelaan 1081a, 1081 HV Amsterdam, The Netherlands
b VU University Medical Centre, Department of Physics and Medical Technology, Brain Imaging Section, De Boelelaan 1118, 1081 HZ Amsterdam, The Netherlands

A R T I C L E   I N F O

Article history:
Received 21 December 2007
Revised 10 July 2008
Accepted 18 July 2008
Available online 13 August 2008

A B S T R A C T

In MEG source localization the estimated source parameters will be more reliable when the spatiotemporal covariance of the noise and background activity is taken into account. Since this covariance is in general too large to estimate based on the data and to invert efficiently, different parametrizations have been proposed in the literature. These models can be seen as special cases of the general decomposition of the covariance into a sum of Kronecker products of spatial matrices \(X_n\) and temporal matrices \(T_n\) (Van Loan, 2000).

In this study we investigate the assumption of the matrices \(T_n\) being Toeplitz. If so, the covariance matrix in the space–frequency domain will have an approximate block-diagonal structure, facilitating inversion, which is a prerequisite for source localization. In this study we address the question whether the Toeplitz approximation is valid for data sets obtained in visual evoked field, auditory evoked field, somatosensory evoked field experiments and data sets containing spontaneous activity. It turns out that on average 87% is in the block-diagonal of the sample covariance, which is close to the values obtained for real Toeplitz matrices \(T_n\). This implies that the space–frequency domain is very interesting for source localization since the major part of the entire covariance can be incorporated in that domain straightforwardly. Finally, the two major processes in the background activity are characterized in terms of their spatial and frequency patterns, yielding a focal and a non-focal pattern in 8 of 10 data sets analyzed in this study. The focal pattern represents the alpha frequency at parieto-occipital areas, whereas the non-focal pattern is more widespread both in space and in frequency.

© 2008 Elsevier Inc. All rights reserved.

Introduction

The covariance structure of the background activity in MEG is important when performing source localization. When the covariance is incorporated in the source localization cost function the resulting parameter estimators will have lower variance, and, hence, will be better than estimators derived without taking the covariance into account (Lütkenhöner, 1998).

In practice, the dimension of the spatiotemporal covariance matrix is too big to allow direct use in the source localization. Therefore, a parametrization is needed. It follows from Van Loan (2000) that in general the matrix \(\sum\) can be decomposed into a sum of Kronecker products (KP) of lower dimensional matrices \(T_n\) (temporal matrices) and \(X_n\) (spatial matrices), \(n = 1, \ldots, N\):

\[
\sum = \sum_{n=1}^{N} T_n \otimes X_n. \quad (1)
\]

The number of terms in this decomposition, \(N\), depends on the dimensions of \(X_n\) and \(T_n\). Kronecker products are attractive for MEG covariance models, since \((T \otimes X)^{-1} = T^{-1} \otimes X^{-1}\), i.e., inversion of the large matrix can be performed by inverting two lower dimensional matrices. This efficient inversion only applies to a single KP, whereas a sum of KP in general has to be inverted at the larger dimension.

Three different parametrizations that have been proposed, can be regarded as special cases of Eq. (1),

\[
\sum = T \otimes X \quad (2)
\]

\[
\sum = \sum_{n=1}^{2} T_n \otimes X_n \quad (3)
\]

\[
\sum = \sum_{n=1}^{N} T_n \otimes X_n^2 \quad (4)
\]

The first parametrization (Eq. (2)) is a single Kronecker product of a temporal covariance matrix \(T\) and a spatial covariance matrix \(X\) (De Munck et al., 2002; Huizenga et al., 2002; Werner et al., 2008). This model is very efficient in source localization computations, though physiologically it is too elementary: the model assumes that temporal correlations are fixed in space and vice versa, which is violated in measurements. The second model (Eq. (3)) is a sum of Kronecker products of temporal covariance matrices \(T_n\) and spatial covariance matrices \(X_n\) and does allow for a better physiological interpretation. Each term in the sum can be regarded as the covariance structure of an independent background process (Bijma et al., 2005). Nevertheless,
the drawback of this model is its lack of efficient inversion, which is a prerequisite for application in source localization. This problem is solved in the third model (Eq. (4)), where the spatial covariance matrices \( X_p \) are constrained to be rank 1 \( \{X_p = x_1x_1^T, \ldots, x_jx_j^T, \ldots, x_Nx_N^T\} \) and orthogonal or independent (Plis et al., 2007). Despite an efficient inversion of this model, the constraint forces the different terms into spatial patterns that are hard to interpret physiologically, because of the rank 1 constraint.

In the present study, the spatial matrices \( X_p \) are left unconstrained, and instead, we investigate constraints on the temporal covariance matrices \( T_n \) in Eq. (3), in order to obtain constraints that can be inverted efficiently. Moreover, the model should allow for a physiological interpretation in terms of background processes, as in model (3). The constraint on the temporal covariance matrices in the present study is that the \( T_n \) be Toeplitz, that is, constant along sub-diagonals. This allows for efficient inversion, by means of the frequency domain. In general, the temporal covariance matrix of a weakly stationary time series is Toeplitz. Weak stationarity of a time series \( X(t) \) means that \( \text{Cov}(X(t), X(t+u)) = c(u) \), i.e., the covariance only depends on time lag, not on absolute time (see e.g. Brockwell and Davis, 1993). The autocovariance function \( c \) is even, \( c(u) = c(-u) \). Consequently, the temporal covariance matrix is a symmetric Toeplitz matrix. In such a case the covariance matrix of the discrete Fourier transform (DFT) of the signal will asymptotically be real, diagonal and non-negative under certain assumptions on the autocovariance function of the time series (see Appendix A). In other words, different frequencies are asymptotically uncorrelated for stationary time series. Applying this reasoning to all temporal matrices \( T_n \) in the decomposition in Eq. (1), the covariance matrix of the DFT of the MEG signals will be block-diagonal, and can be inverted easily. Subsequently, source localization can be performed either in the frequency domain or, using the inverse DFT, in the time domain.

The goal of the present study is to verify the assumption of Toeplitz temporal covariance matrices in Eq. (1) in the space–time domain or, equivalently, a block-diagonal covariance in the space–frequency domain. This can be regarded as a test for stationarity on covariance level, in stead of signal level (Mäkinen et al., 2005b). The measure used to quantify the block-diagonality is the relative matrix power in the real part of the block-diagonal. For perfectly stationary signals the properties of this measure are known exactly (Appendix B). Furthermore, analogous to model (3), the two major background processes are characterized in the space–frequency domain, yielding the spatial and frequency patterns of these processes.

## Methods

### Block-diagonality of the covariance in space–frequency

Denote the background MEG signal on channel \( h \), in time sample \( j \) by \( y_{hj} \), \( h = 1, \ldots, H \) and \( j = 1, \ldots, J \). These signals are viewed as stochastic signals with \( E[y_{hj}] = 0 \) for all \( h, j \), i.e., as signals with zero mean. Collecting all signals in one vector \( y' = (y_{1,1}, \ldots, y_{1,J}, \ldots, y_{H,1}, \ldots, y_{H,J})' \), a stochastic vector \( y' \in \mathbb{C}^{HJ \times 1} \) is obtained with \( E[y'] = 0 \). The spatiotemporal covariance matrix of the background MEG signals is the covariance matrix of \( y' \),

\[
\Sigma_y = E \left( (y - E(y'))(y - E(y'))' \right) = E(yy') \in \mathbb{C}^{HJ \times HJ},
\]

where \( A' \) denotes the transpose of a matrix \( A \). (The subscript \( y \) in \( \Sigma_y \) is inserted in order to distinguish this matrix from \( \Sigma_x \), defined below.)

In the present study we investigate the covariance matrix of \( x' \), the DFT of \( y' \). This discrete Fourier transform is defined channelwise as (see e.g. Brockwell and Davis, 1993)

\[
z_{hk} = \frac{1}{\sqrt{J}} \sum_{j=1}^{J} y_{hj} e^{-2\pi i kj/J} \quad \text{for} \quad k = 1, \ldots, J, \quad \text{and} \quad h = 1, \ldots, H
\]

where \( J = 2 \). Similar to \( y' \), the stochastic vector \( z' \in \mathbb{C}^{HJ \times 1} \) is defined as \( z' = (z_{1,1}, \ldots, z_{1,J}, \ldots, z_{H,1}, \ldots, z_{H,J})' \). \( E[z'] = 0 \) and, hence, denoting complex conjugation by an asterisk, the covariance matrix of \( z' \) is

\[
\Sigma_z = E \left( (z' - E(z'))(z' - E(z'))' \right) = E(z'z')' \in \mathbb{C}^{HJ \times HJ}.
\]

The structure of \( \Sigma_z \) depends on the structure of \( \Sigma_y \). The decomposition in Eq. (1) for \( \Sigma_y \) is a consequence of the more general result in Van Loan (2000) that any \( n \times m \) matrix \( A \) can precisely be expressed as a sum of \( \min(nm, pq) \) Kronecker products of \( n \times m \) matrices \( T_n \) and \( p \times q \) matrices \( X_p \). Applying this to \( A = \Sigma_y \) with \( n = m = H \) and \( p = q = 1 \) one finds

\[
\Sigma_y = \sum_{n=1}^{\min(H^2, J^2)} T_n \otimes X_n.
\]

The decomposition in Eq. (8) is not unique. To obtain uniqueness within each term \( T_n \otimes X_n \), the normalization constraint \( \|X_ny\| = 1 \) is imposed for \( n = 1, \ldots, N. \) In other words, the Frobenius matrix norm, defined by

\[
|A|^2 = \text{tr}(A^*A) = \sum_{i,j}|a_{ij}|^2
\]

of the spatial covariance matrices is set to 1. In Eq. (9) \( \text{tr}(M) \) indicates the trace of a matrix \( M \) and the elements of matrix \( A \) are denoted as \( a_{ij} \). With this constraint, the decomposition in Eq. (8) is still not unique. The next section and Appendix C discuss how to find a representation of the sum that can be interpreted as a sum of KP of covariance matrices.

Using the expression in Eq. (8) for \( \Sigma_y \), the matrix \( \Sigma_z \) can be expressed as a sum of Kronecker products as well. Appendix A shows that Toeplitz matrices \( T_n \) lead asymptotically to \( \Sigma_z \) having a block-diagonal structure with real blocks on the diagonal and zeroes elsewhere, where ‘asymptotically’ indicates the limit for \( J \rightarrow \infty \). Hence, for \( J \) large enough,

\[
\Sigma_z = \sum_{n=1}^{\min(H^2, J^2)} D_n \otimes X_n,
\]

where the \( D_n \) are real, diagonal and positive definite. Moreover

\[
(D_n)_{kk} = \sum_{u=1}^{J-1} c_u(u)e^{-2\pi inu/J},
\]

with \( c_u(u) \) the autocovariance function corresponding to \( T_n \). See Appendix A. The constraint of unit Frobenius norm of the \( X_n \) is also applied in Eq. (10).

The right hand side in Eq. (10) is appealing, since it can be inverted block by block. The inverse of the \( K^2 \) block matrix is equal to

\[
\left( \sum_{n=1}^{\min(H^2, J^2)} (D_n)_{kk} \otimes X_n \right)^{-1}.
\]

This efficient inversion implies that the right hand side in Eq. (10) is useful for source localization. The question now is, whether the assumptions made on the way leading to this appealing expression are valid. If the time series are weakly stationary and a large amount of data is available, then the approximate relation (10) must hold. Unfortunately, verifying either of these conditions in real, finite data is not straightforward. It is easier to verify the right hand side of Eq. (10) immediately. For that reason, the complex-valued space–frequency sample covariance matrix is computed and its block-diagonality is investigated.
The sample covariance matrix of \( z \), denoted as \( T_z \), is given by

\[
T_z = \frac{1}{L-1} \sum_{l=1}^{L-1} \left( z_l - \bar{z} \right) \left( \left( z_l - \bar{z} \right)^T \right)
\]

where \( z_l \) is the DFT of \( y_l \), \( \bar{z} \) is the measured data vector in trial \( l \), for \( l = 1, \ldots, L \) and \( z \) is the average of the \( z_l \). The blocks \( T_{ik} \) all have dimension \( H \times H \). The vectors \( \bar{z} \) are considered to be independent realizations of \( y(z) \) for \( l = 1, \ldots, L \). The matrix \( T_z \) is an unbiased estimator for \( \Sigma_z \). If the approximate relation (10) holds for \( \Sigma_z \), then a similar relation will hold for its estimator \( \tilde{T}_z \)

\[
\tilde{T}_z = \begin{pmatrix}
\Re(T_{11}) & 0 & \ldots & 0 \\
0 & \Re(T_{22}) & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & \Re(T_{jj})
\end{pmatrix}
\]

Since \( y \) is real, there is dependency amongst the Fourier transforms: \( 2\delta_m \tilde{z}_n \). Hence, it is common to consider only part of the block-diagonal matrix in the right hand side of Eq. (14). The blocks with indices \( k = 0, \ldots, \frac{L}{2} \) provide all information \( J \) is assumed to be even), where index \( k = 0 \) corresponds to the case where \( k = \frac{L}{2} \). Define \( K \) as the frequency index \( k \) run from 0 to \( K \). In the sequel, when the number of trials is big, whereas \( \Sigma_z(H^H + H^H)^2 \) matrix \( T_z \) is considered.

In the space–time domain the goodness-of-fit of model (3) was measured by the relative matrix power explained in the model (Bijma et al., 2005). This same measure can be used in order to verify the “realness” and “block-diagonality” of \( T_z \). The matrix power in the real part of the block-diagonal of \( T_z \) relative to the matrix power of the entire matrix \( T_z \), is

\[
P_{\text{rel}} = \frac{1}{\|T_z\|^2} \sum_{k = 0}^{K} \left( \left\| \Re(T_{kk}) \right\|^2 \right) \times 100\%.
\]

When this fraction is (much) smaller than 1, it can either be due to large off-block-diagonal elements or to large imaginary parts in the block-diagonal elements or both. Therefore, the fractions \( P_{\text{off}} \) and \( p_{\text{im}} \), quantifying these two types of deviations respectively, are of interest too:

\[
P_{\text{off}} = \frac{1}{\|T_z\|^2} \sum_{k,k'=0 \atop k \neq k'} \left( \left\| \left( T_{kk'} \right) \right\|^2 \right) \times 100\%.
\]

\[
p_{\text{im}} = \frac{1}{\|T_z\|^2} \sum_{k = 0}^{K} \left( \left\| \Im(T_{kk}) \right\|^2 \right) \times 100\%.
\]

One can show that

\[
P_{\text{rel}} + p_{\text{off}} + p_{\text{im}} \geq 100\%.
\]

The matrices \( T_{ij} \) are close to Toeplitz, \( P_{\text{rel}} \) should be big, whereas \( P_{\text{off}} \) and \( p_{\text{im}} \) should be small.

In the case of perfect stationary data (e.g., white noise) the covariance is (very close to) block-diagonal when \( K \) is large enough. This means that the expectation of the off-diagonal elements in \( T_z \) is (very close to) zero. However, the deviation from the expected value, i.e., the variance, will depend on the number of trials \( (L) \) taken into account. Since the squared absolute values of the off-diagonal elements are summed in \( P_{\text{rel}} \), the fraction \( p_{\text{rel}} \) will also depend on \( L \). Moreover, since the number of off-diagonal elements is \( O(K^2) \) and the number of diagonal elements is \( O(K) \), the value of \( p_{\text{rel}} \) decreases with \( K \) for a fixed number of trials. This is explained in Appendix B.

Finding background processes in space–frequency

In Bijma et al. (2005) the sum of KP model (8) was applied to the sample covariance of \( y \). A sum of \( N=2 \) Kronecker products was fitted to the sample covariance in the space–time domain and it appeared that 67% to 93% of the matrix power was accounted for by two KP terms. These two terms each represent the spatiotemporal pattern of a background process. Applying an analogous method in the space–frequency domain, one can find the dominant background processes in terms of their space–frequency patterns. This is performed by fitting the right hand side of Eq. (10) to the sample covariance matrix, \( T_z \), i.e., by minimizing

\[
\tilde{C}(T_D, X_1, \ldots, D_n, X_N) = \| \tilde{T}_z - \sum_{n=1}^{N} D_n \otimes X_n \|_F^2.
\]

subject to the constraints that the matrices \( X_n \) and \( D_n \) are positive definite and the matrices \( D_n \) are diagonal for \( n=1, \ldots, N \). Both the imaginary part and the off-block-diagonal elements of \( T_z \) cannot be explained by the real matrices \( X_n \) and \( D_n \) that minimize \( C \) in Eq. (19). Consequently, these matrices can as well be found by approximating only the real part of the block-diagonal of \( T_z \), which was defined as \( \tilde{T}_z \) in Eq. (14),

\[
\tilde{C}(T_D, X_1, \ldots, D_n, X_N) = \| \tilde{T}_z - \sum_{n=1}^{N} D_n \otimes X_n \|_F^2.
\]

When we first disregard the constraint of positivity of the matrices \( X_n \) and \( D_n \), the cost function in Eq. (20) can be minimized using the rearrangement operator S for matrices presented in Van Loan (2000)

\[
\tilde{C}(T_D, X_1, \ldots, D_n, X_N) = \| S(\tilde{T}_z) - \sum_{n=1}^{N} \text{vec}(X_n)^T \text{vec}(D_n) \|_F^2
\]

with \( S(\tilde{T}_z) \in \mathbb{R}^{(H^{2}+K^{2}) \times N} \), containing the elements of \( \tilde{T}_z \) in a rectangular ordering. The vectors \( \text{vec}(D_n) \) and \( \text{vec}(X_n) \), \( n=1, \ldots, N \), that minimize the cost function are found from the singular value decomposition (SVD) of \( S(\tilde{T}_z) \). Advantage is taken from the diagonality of the \( D_n \) leading to many columns being zero in \( \text{vec}(X_n) \) and, thus, reducing the dimensionality of the model. Because of this dimension reduction from \( H^2 \times (K+1)^2 \) to \( H^2 \times (K+1) \), computing the SVD is not a problem in terms of computation time. In fact, the computation time is dominated by the time needed to compute \( T_{ij} \) and the fractions \( P_{\text{rel}}, P_{\text{off}} \) and \( p_{\text{im}} \).

Due to the orthogonality of the singular vectors in the SVD of \( S(\tilde{T}_z) \), \( \text{vec}(X_n) \) and \( \text{vec}(D_n) \) are perpendicular for \( n \neq n' \), and, hence, the matrices \( X_n \) and \( X_n' \) cannot both be positive (semi-)definite (Bijma et al., 2005). A similar reasoning holds true for the matrices \( D_n \). This means that the constraint of positivity of the matrices \( X_n \) and \( D_n \) is not fulfilled. Hence, these matrices cannot be interpreted as covariance matrices, since for that interpretation they should be positive (semi-)definite. This problem is solved using the non-uniqueness of a general rank \( N \) expression: the sum \( \sum_n \text{vec}(X_n)^T \text{vec}(D_n) \) that is found by minimizing Eq. (21) is the best rank \( N \) approximation to \( S(\tilde{T}_z) \), and, therefore, this sum can be expressed in many ways. This non-uniqueness is directly comparable to the non-uniqueness of the representation of a matrix as a sum of KP, as in Eq. (8), through \( S^T \). It appears that it is possible to express the obtained sum in such a way that all matrices obey an interpretation as covariance matrix. This procedure is discussed for \( N=2 \) in Appendix C. Thanks to the diagonality of the \( D_n \) this rewriting procedure is more efficient in the space–frequency domain than it is in the space–time domain in Bijma et al. (2005). The reader is referred to that study for more details. In the sequel, \( \tilde{X}_n \) and \( D_n \) indicate the matrices that appear in the final positive (semi-)definite sum, which can be interpreted as covariance matrices.
In this study N=2 will be used. Since a sum of 2 real valued Kronecker products may not describe the entire \( T_2 \), let alone the entire \( T_2 \), the fraction that is explained by the estimated sum of 2 Kronecker products is computed in the applications, as follows:

\[
p_T = \frac{1-ar{C}(D_1, \hat{X}_1, D_2, \hat{X}_2)}{\| T_2 \|_F^2} \times 100\% \tag{22}
\]

\[
p_T = p_T \times \frac{\| T_2 \|^2}{\| T_2 \|_F^2} \times 100\% = P_T \times p_{re}. \tag{23}
\]

**Subjects and experiments**

The space–frequency sample covariance matrix was calculated for 10 data sets: 2 data sets containing spontaneous activity (SPON), 3 data sets containing visual evoked fields (VEF), 2 data sets containing somatosensory evoked fields (SEF) and 3 data sets containing auditory evoked fields (AEF). The spontaneous data were recorded during rest condition, with eyes closed, lying in a horizontal position at a sample rate of 312.5 Hz. In the VEF experiment the stimuli consisted of 6' checkerboard pattern onset shown full field, and data were recorded at 625 Hz. The SEF data were recorded using electric stimulation of the left wrist, with an inter stimulus interval uniformly distributed between 800 and 1200 ms at a sample rate of 2038 Hz. Offline these data were downsampled by a factor of 2, resulting in a sample rate of 297.57 Hz. Finally, the auditory stimuli consisted of 500 Hz tones at 625 Hz. The SEF data were recorded using electric stimulation of the scalp, at 625 Hz. In the VEF experiment the stimuli consisted of 6 1st order sine waves with a frequency of 1 to 50 Hz. In this study the different numbers of epochs (\( J \)) and the length of the different windows for each data set are stated in Table 1. We did not strive for uniformity amongst data sets, since we wanted to investigate the different data sets and are stated in Table 1. We did not strive for uniformity amongst data sets, since we wanted to investigate the different numbers of epochs for many different settings. The number of epochs in the first column contains the value of \( L \). Columns 2 to 3, 4 to 5, 6 to 7 and 8 to 9 correspond to epoch lengths of 256, 128, 64 and 32 samples respectively, as is indicated in the top row. The last column contains the data set name corresponding to the value of \( L \). The values shown are averages over 25 simulations of \( L \) epochs of length \( J \). The variance amongst the 25 values per group was in all cases very small, (10<sup>-5</sup>). The value \( p_{re} \) is zero in the case of perfectly stationary data in one channel.

therefore such a baseline correction would introduce artifactual non-stationarities in the covariance structure, due to alpha activity (Bijma et al., 2003). Therefore, in these data sets a baseline correction over 100 ms pre-stimulus was applied. The length 100 ms is taken, since this length can be taken as one alpha period in general. For the SEF and SPON epochs consisting of 32 samples the baseline correction was applied over the entire epoch, since 1075 and 1025 ms for SEF and SPON respectively is rather close to one alpha period. The baseline correction together with the subtraction of the average over epochs results in zero averages both within epochs and over epochs for each channel, i.e.,

\[
\frac{1}{L} \sum_{i=1}^{L} y_{h,j} = \frac{1}{L} \sum_{i=1}^{L} y_{h,j} - \frac{1}{L} \sum_{i=1}^{L} y_{h,j} = 0
\]

where \( y_{h,j} \) is the signal in channel \( h \), time sample \( j \) in epoch \( l \). The DFT (Eq. [6]) was computed for these residual signals.

**Results**

In Table 2 values for \( p_{re} \) and \( p_{off} \) are shown for different values of \( J \) and \( L \), all based on simulated, stationary white noise. The values in this table are simulated values for \( p_{re} \) and \( p_{off} \) based on white noise in one channel, as described in Appendix B.

The first column identifies the data set. Columns 2 to 4, 5 to 7, 8 to 10 and 11 to 13 correspond to epoch lengths of 256, 128, 64 and 32 samples respectively, as is indicated in the top row. \( p_{re} \) gives the matrix power in the real part of the block-diagonal relative to the entire matrix (Eq. [15]), \( p_{off} \) gives the relative matrix power in the off-block-diagonal matrix elements (Eq. [16]) and \( p_{off} \rangle \) gives the relative matrix power in the imaginary part of the block-diagonal (Eq. [17]). All values shown are percentages. The last two rows show the average and standard deviation of each column.

### Table 1

<table>
<thead>
<tr>
<th>Dataset</th>
<th>256</th>
<th>128</th>
<th>64</th>
<th>32</th>
</tr>
</thead>
<tbody>
<tr>
<td>AEF1</td>
<td>289</td>
<td>410</td>
<td>205</td>
<td>102.5</td>
</tr>
<tr>
<td>AEF2</td>
<td>217</td>
<td>410</td>
<td>205</td>
<td>102.5</td>
</tr>
<tr>
<td>AEF3</td>
<td>268</td>
<td>410</td>
<td>205</td>
<td>102.5</td>
</tr>
<tr>
<td>SEF1</td>
<td>480</td>
<td>860</td>
<td>430</td>
<td>215</td>
</tr>
<tr>
<td>SEF2</td>
<td>465</td>
<td>860</td>
<td>430</td>
<td>215</td>
</tr>
<tr>
<td>SPON1</td>
<td>268</td>
<td>820</td>
<td>410</td>
<td>205</td>
</tr>
<tr>
<td>SPON2</td>
<td>268</td>
<td>820</td>
<td>410</td>
<td>205</td>
</tr>
<tr>
<td>VEF1</td>
<td>400</td>
<td>410</td>
<td>205</td>
<td>102.5</td>
</tr>
<tr>
<td>VEF2</td>
<td>400</td>
<td>410</td>
<td>205</td>
<td>102.5</td>
</tr>
<tr>
<td>VEF3</td>
<td>400</td>
<td>410</td>
<td>205</td>
<td>102.5</td>
</tr>
</tbody>
</table>

The first column identifies the data set. The columns 256, 128, 64 and 32 give the epoch length in ms for the \( J = 256, 128, 64 \) and 32, respectively.

### Table 2

<table>
<thead>
<tr>
<th>( J )</th>
<th>256</th>
<th>128</th>
<th>64</th>
<th>32</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( p_{re} )</td>
<td>( p_{off} )</td>
<td>( p_{re} )</td>
<td>( p_{off} )</td>
</tr>
<tr>
<td>217</td>
<td>62.9</td>
<td>37.1</td>
<td>77.2</td>
<td>22.8</td>
</tr>
<tr>
<td>268</td>
<td>67.6</td>
<td>32.4</td>
<td>80.7</td>
<td>19.3</td>
</tr>
<tr>
<td>289</td>
<td>69.4</td>
<td>30.6</td>
<td>82.0</td>
<td>18.0</td>
</tr>
<tr>
<td>400</td>
<td>75.7</td>
<td>24.3</td>
<td>86.1</td>
<td>13.9</td>
</tr>
<tr>
<td>465</td>
<td>78.4</td>
<td>21.6</td>
<td>87.9</td>
<td>12.1</td>
</tr>
<tr>
<td>480</td>
<td>79.0</td>
<td>21.0</td>
<td>88.3</td>
<td>11.7</td>
</tr>
</tbody>
</table>

### Table 3

<table>
<thead>
<tr>
<th>( J )</th>
<th>256</th>
<th>128</th>
<th>64</th>
<th>32</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( p_{re} )</td>
<td>( p_{off} )</td>
<td>( p_{re} )</td>
<td>( p_{off} )</td>
</tr>
</tbody>
</table>

The first column identifies the data set. Columns 2 to 4, 5 to 7, 8 to 10 and 11 to 13 correspond to epoch lengths of 256, 128, 64 and 32 samples respectively, as is indicated in the top row. \( p_{re} \) gives the matrix power in the real part of the block-diagonal relative to the entire matrix (Eq. [15]), \( p_{off} \rangle \) gives the relative matrix power in the off-block-diagonal matrix elements (Eq. [16]) and \( p_{off} \rangle \) gives the relative matrix power in the imaginary part of the block-diagonal (Eq. [17]). All values shown are percentages. The last two rows show the average and standard deviation of each column.
globally decreases with epoch length in Fig. 1(a), which is in line with the stationary noise simulation results in Table 2. Figs. 1(b) and (c) show a global increase with epoch length in the values for $p_{\text{re}}$ and $p_{\text{im}}$, though the $p_{\text{off}}$ values do not increase monotonically.

Over all analyses $p_{\text{re}} = 83.8 \pm 7.4\%$, $p_{\text{off}} = 12.7 \pm 6.9\%$ and $p_{\text{im}} = 4.1 \pm 3.0\%$, indicating averages and standard deviations. Since the values of $p_{\text{re}}$ are rather high, it makes sense to see what background processes dominate this main body of the data. Note that data set SPON2 is an outlier, in having higher $p_{\text{im}}$ values and lower $p_{\text{re}}$ values than other data sets.

To find the two dominant background processes, the sum of two KP was fitted. It appeared that in the total of 40 analyses the average $p_{\text{T}}$ in Eq. (22) was 98.7%, ranging from 94.53% to 99.97%, see Table 4. This means that setting $N=2$ in stead of a general, larger $N$ in Eq. (21) is no serious restriction. On the other hand, the selection of the real block-diagonal $\hat{T}$ of $\mathbf{T}$ is of course a true restriction and is quantified in $p_{\text{re}}$. The fraction $p_{\text{T}} = p_{\text{T}} \times p_{\text{re}}$ was in all cases very close to $p_{\text{re}}$ in Table 3. Finally, the last possible limitation of fitting a sum of two covariance KP terms lies in the positivity constraint of the matrices $X_n$ and $D_n$: the negative parts of the final matrices are discarded (see Appendix C). In 14 of the 40 cases, the rewritten sum was fully positive definite and this limitation did not apply. Over all 40 analyses the positivity of the rewritten matrices was on average 99.97%, ranging from 95.87% to 100% (results not shown). Hence, the explained matrix power that is lost by discarding the negative parts in order to get covariance interpretations is negligible. In short, the sum of only 2 real valued positive (semi)-definite Kronecker products explains $p_{\text{T}}$ part of the background characteristics.

Each term in the so obtained sum of two Kronecker products represents a background process in the measured MEG signals in terms of its frequency and spatial patterns. The diagonal elements of $\hat{D_n}$ contain the power of the different frequencies in the process, and $\hat{X_n}$ contains the spatial covariance pattern for the process, for $n = 1, 2, \ldots$. The fewer samples the epochs contain, the worse the frequency resolution in these patterns. Since the epochs consisting of 256 samples are the most interesting in terms of space–frequency patterns, only results for this epoch length are shown. For this epoch length the value for $p_{\text{re}}$ was between 69% and 85% for all data sets except for SPON2, which showed a $p_{\text{re}}$ value of only 45.2%. In Table 2 values between 62% and 79% are reported for $p_{\text{re}}$ for perfectly weakly stationary signals and similar values of $L$. This shows that the off-block-diagonal elements in the estimated covariance matrices are not substantially larger than those in the stationary case. The results are illustrated in Figs. 2–5 for 4 of the 10 data sets. For each data set three figures are shown. The two figures at the top present the spatial patterns in a top Mercator view of the MEG helmet. The figure at the bottom shows the frequency patterns. In the spatial patterns, the diagonal elements of $\hat{X}_1$ and $\hat{D}_2$, that is, the variance in all channels, is

![Diagram](image)

**Fig. 1.** (a) $p_{\text{re}}$, (b) $p_{\text{off}}$, (c) $p_{\text{im}}$. The percentages $p_{\text{re}}$, $p_{\text{off}}$ and $p_{\text{im}}$ from Table 3 as functions of epoch length, shown for each subject. Note the logarithmic scale on the horizontal axis, and the different scales on the vertical axes in the different figures.

<table>
<thead>
<tr>
<th>Set</th>
<th>256</th>
<th>128</th>
<th>64</th>
<th>32</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_{\text{T}}$</td>
<td>$p_{\text{T}}$</td>
<td>$p_{\text{T}}$</td>
<td>$p_{\text{T}}$</td>
<td>$p_{\text{T}}$</td>
</tr>
<tr>
<td>AEF1</td>
<td>96.81</td>
<td>71.22</td>
<td>98.96</td>
<td>85.48</td>
</tr>
<tr>
<td>AEF2</td>
<td>98.97</td>
<td>83.53</td>
<td>99.20</td>
<td>89.02</td>
</tr>
<tr>
<td>AEF3</td>
<td>98.06</td>
<td>80.57</td>
<td>99.51</td>
<td>89.70</td>
</tr>
<tr>
<td>SEF1</td>
<td>94.53</td>
<td>73.44</td>
<td>96.85</td>
<td>74.57</td>
</tr>
<tr>
<td>SEF2</td>
<td>96.73</td>
<td>71.18</td>
<td>97.57</td>
<td>81.72</td>
</tr>
<tr>
<td>SPON1</td>
<td>97.97</td>
<td>68.02</td>
<td>97.95</td>
<td>65.80</td>
</tr>
<tr>
<td>SPON2</td>
<td>98.48</td>
<td>44.54</td>
<td>99.52</td>
<td>68.30</td>
</tr>
<tr>
<td>VEF1</td>
<td>98.77</td>
<td>78.37</td>
<td>99.43</td>
<td>83.14</td>
</tr>
<tr>
<td>VEF2</td>
<td>96.22</td>
<td>79.24</td>
<td>97.30</td>
<td>85.26</td>
</tr>
<tr>
<td>VEF3</td>
<td>97.32</td>
<td>77.29</td>
<td>99.60</td>
<td>86.41</td>
</tr>
</tbody>
</table>

The first column identifies the data set. All values shown are percentages.
plotted. For each data set, the color scale is auto-scaled in order to keep the contrast at an adequate level. The scaling factor is kept constant within one data set. Nevertheless, the magnitude of a space–frequency pattern is fully accounted for by the $\hat{D}_n$, since the $\hat{X}_n$ are normalized. In the frequency patterns, the diagonal elements of $\hat{D}_1$ and $\hat{D}_2$ are plotted as function of frequency. The blue line corresponds to $\hat{D}_1$ and figure (a), and the red line corresponds to $\hat{D}_2$ and figure (b). The percentages stated below figures (a) and (b) denote the relative matrix power of $\hat{X}_n$ accounted for by the first and second space–frequency pattern, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
and $\hat{D}_2$ are plotted as function of frequency. Blue (red) lines correspond to $\hat{D}_1$ ($\hat{D}_2$). For each term in the presented sum of 2 Kronecker products the relative matrix power of $\Upsilon_z$ accounted for by that term is computed and shown in the captions of Figs. 2–5. These two percentages sum to $p_z$ given in Table 4 (apart from some rounding errors).

It appears that in most data sets, one pattern is located in parieto-occipital regions with a clear peak in the frequency pattern at around 10 Hz. This pattern is presenting mainly alpha activity. The second pattern is less focal in space, though for most data sets it covers the temporal regions and is rather symmetric in left–right direction. Also in frequency, there is no specific pattern for this second background pattern.
process. This separation between alpha activity and other background activity is not manifest in 2 of the 10 data sets; in these data sets (e.g., AEF3 in Fig. 2) the patterns show overlap in either space or frequencies or both.

Discussion and conclusions

The proposed model combines the advantages of the models in Eqs. (2) to (4), namely an efficient inversion and the information on different processes in the background activity. Whereas the model given in Eq. (2) is efficient in inversion, the results in this study show that one Kronecker product is not sufficient to describe the background processes: two distinct processes are found in 8 of the 10 data sets analyzed.

The relative power in the real part of the block-diagonal of \( T_p \) was 83.8% on average over different types of data sets and epoch lengths. In fact, for localization of stationary dipole sources in the frequency domain, one does not necessarily have to take the real part of the block-diagonal, though one can take the entire, complex-valued block-diagonal. The relative power of the complex-valued block-diagonal equals \( 100\% - \rho_{re} \), and was on average \( 87.3 \pm 6.9\% \). In order to compare the performance in source localization of the model described in Eq. (2) and the model proposed in this study, an extensive (simulation) study is needed, taking into account different epoch lengths and residual signals of different types of experiments with different levels of stationarity. This is beyond the scope of the present study. Nevertheless, the percentages reported for \( \rho_{re} \) for all 40 analyses in Table 3 do not deviate substantially from those reported for perfectly stationary data in Table 2. Based on these values, the proposed space–frequency covariance model is expected to be a good alternative for the space–time covariance models in source localization.

The asymptotic property of uncorrelated frequencies is sometimes taken as a starting point in MEG studies, e.g., Grasman et al. (2004) and Pham et al. (1987), while it is a priori not evident that the stationarity of the MEG signals in finite time is high enough for assuming the asymptotic state. An investigation of the weak stationarity of MEG signals on covariance level in space–time was performed in Bijma et al. (2003), where the single KP model (2) was used. In that study it appeared that 99% of the so obtained temporal covariance matrix could be explained by a stationary model. However, it is not precisely clear what influence model (2) has on the stationarity of the estimated matrix \( T \); it may be that the part of the sample covariance matrix that is not accounted for by the single KP model contains non-stationarities. Consequently, for MEG signals it is not a priori clear whether different frequencies are (close to) uncorrelated. This study shows that, on covariance level in space–frequency, the results for MEG signals do not deviate substantially from the results for perfectly stationary signals.

Characterizing the two major background processes in space–frequency patterns leads roughly to an alpha process and a more widespread, non-focal process. The findings in the current study are in line with earlier results in Bijma et al. (2005). In most data sets, we found one alpha pattern, localized mainly in the parieto-occipital area with a clear peak around 10 Hz in the frequency pattern and a more widespread pattern. The separation between the two processes was not revealed in all data sets. This has to do with the non-uniqueness of writing a rank-2 sum; the sum can be split into two processes in several ways. The criterion we used here is that of a possible covariance interpretation (i.e. positive (semi)-definiteness), and this has led to an alpha pattern and a remainder pattern in most data sets. Moreover, this criterion causes a small notch in the second frequency pattern of most data sets opposite to the peak in the alpha frequency pattern. Apparently, these patterns maximize the positivity of the corresponding spatial matrices. In Bijma et al. (2005) the frequency content was not quantified, while in the present study this is quantified by the matrices \( D_p \). In this way, the entire block-diagonal matrix is summarized in only two space–frequency patterns. The current method can be extended to more than two processes, leading to the characterization of multiple background processes. These patterns may be interpreted as processes found in resting state or default networks (Fox et al., 2005; Greicius et al., 2003; Ioannides, 2007; Raichle et al., 2001).

In the present study, different types of data (AEF, SEF, VEF and spontaneous) were considered. No clear differences are present in the values of \( \rho_{re} \) or in the space–frequency patterns for these different data types, although the two spontaneous data sets showed \( \rho_{re} \) values in the lower range for all epoch lengths (Table 3). Nevertheless, it may well be that the stationarity of the residual signals in different types of data varies. Since the responses in the evoked field signals were removed by a simple average subtraction, one would expect that the residual signals contain non-stationarities due to possible trial-to-trial variability of the response (Mäkinen et al., 2005a; Truccolo et al., 2002) and possible modulation of the alpha activity (Makeig et al., 2002). Applying advanced methods for estimating single trial responses (e.g., De Munck et al., 2004; Bénar et al., 2007) will facilitate the separation between response and residual and will lead to a more precise covariance estimation in the space–time domain as well as in the space–frequency domain.

In sum, the assumption of Toeplitz \( T_n \) in Eq. (3) leads to a block-diagonal structure in the space–frequency covariance matrix. This in turn, leaves the space–frequency domain promising for source localization, since a large amount of the noise covariance can be easily incorporated. Furthermore, from the block-diagonal of this matrix the space–frequency characteristics of multiple background patterns can be found yielding insight into the different background processes in the brain. Determining the two major processes separates alpha activity from remaining background activity.

Acknowledgments

The authors are grateful to the anonymous reviewers for their helpful suggestions. This had led to new insights and improvements of the method.

Appendix A. Properties of the covariance matrix in space–frequency domain

In this appendix the structure of the covariance matrix of \( z \) is derived under the assumption that the time series \( y_{hj} \) are stationary for all channels \( h \). No periodic extension of the signal is assumed, since such an extension introduces an artificial covariance structure. By extending the signal periodically, there will be a correlation of 1 between time samples \( j=1 \) and \( j=J+1 \), since these signals are identical, whereas the correlation between time samples \( j=1 \) and \( j=J \) will be close to 0, since the time lag between the latter two instants is large. Therefore, the periodic extension introduces discontinuities and artificially large correlations over long time lags. By not assuming a periodic extension one regards the recorded signal as a selection of an infinitely ongoing time series, which is assumed to be weakly stationary throughout this appendix. In textbooks like Brillinger (1975), Brockwell and Davis (1993) and Priestley (1981) the asymptotic zero correlation between different frequencies is not proven explicitly, though stronger results about asymptotic independence amongst frequencies are proven. Such stronger results require more conditions than only weak stationarity, whereas uncorrelated frequencies only require certain conditions on the autocovariance, as shown below.

Consider first the single KP model (2) with a symmetric Toeplitz \( T \) matrix. Then

\[
\text{Cov}(y_{hj}, y_{hj'}) = \mathbb{E}(y_{hj}y_{hj'}) = T_{jj'}X_{hh'} = c(j-j')X_{hh'}.
\]

where \( c \) denotes the even autocovariance function of the time series as function of the time lag, denoted by the time sample index lag for simplicity.
The entries of $\sum_z$ are given by

$$\text{Cov}(z_{nk}, z_{n'k'}) = E(z_{nk} z_{n'k'}^*)$$

(26)

for $k, k' = 0, \ldots, K$. Inserting the definition $z_{nk} = \frac{1}{\sqrt{J}} \sum_{j=1}^J y_{nj} e^{-2\pi i j k / J}$ yields

$$\text{Cov}(z_{nk}, z_{n'k'}) = \frac{1}{J} \sum_{j=1}^J \sum_{j'=1}^J E(y_{nj} y_{n'j'}) e^{-2\pi i j (k-k') / J}$$

(27)

where Eq. (25) was used for the last equality. Rewriting this double sum over $j$ and $j'$ as a double sum over $u = j - j'$ and $j'$ yields

$$\text{Cov}(z_{nk}, z_{n'k'}) = \frac{X_{hh}}{J} \left( \sum_{u=-J}^J \sum_{u=-J}^J (j + u) + \sum_{u=-J}^J (j-u) \right) c(u) e^{-2\pi i k u / J}$$

(28)

If $k = k'$, then $e^{-2\pi i k (k-k') / J} = 1$ and, therefore,

$$\text{Cov}(z_{nk}, z_{n'k'}) = \frac{X_{hh}}{J} \left( \sum_{u=-J}^J (j + u) + \sum_{u=-J}^J (j-u) \right) c(u) e^{-2\pi i k u / J}$$

$$= X_{hh} \sum_{u=-J}^J \left( 1 - \frac{|u|}{J} \right) c(u) e^{-2\pi i k u / J}$$

(29)

For the asymptotic result, consider a sequence $k_0 \in \mathbb{Z}$ such that $\lambda_n = 2\pi k_0 / J \rightarrow \lambda$ converges for $n \rightarrow \infty$ and assume

$$\sum_{u=-\infty}^\infty |c(u)| < \infty$$

(30)

Then by the dominated convergence theorem, the covariance $\text{Cov}(\text{zh}_{nk}, \text{zh}_{n'k'})$ converges for $n \rightarrow \infty$ to

$$X_{hh} \sum_{u=-\infty}^\infty c(u) e^{-2\pi i k_0 u / J}$$

(31)

which is the Fourier transform of the autocovariance function $c(u)$ multiplied by the spatial covariance term $X_{hh}$. For $k \neq k'$, consider a second sequence $k_0' \in \mathbb{Z}$ such that $k_0 \neq k_0'$ for all $n$ and $\lambda_n = 2\pi k_0 / J \rightarrow \lambda' \neq \lambda$ converges for $n \rightarrow \infty$. Then the covariance $\text{Cov}(\text{zh}_{nk}, \text{zh}_{n'k'})$ tends to zero for $n \rightarrow \infty$. This follows from the following reasoning.

$$\text{Cov}(z_{nhk}, z_{nhk'}) = \frac{X_{hh}}{J} \left( \sum_{u=-J}^J \sum_{u=-J}^J (j + u) + \sum_{u=-J}^J (j-u) \right) c(u) e^{-2\pi i k_0 u / J} e^{-2\pi i k_0' u / J}$$

$$= X_{hh} \left( \sum_{u=-J}^J \sum_{u=-J}^J (j + u) + \sum_{u=-J}^J (j-u) \right) c(u) e^{-2\pi i k_0 u / J} e^{-2\pi i k_0' u / J}$$

$$\times c(u) e^{-2\pi i k_0 u / J} e^{-2\pi i k_0' u / J}$$

$$= X_{hh} \left( \sum_{u=-J}^J \sum_{u=-J}^J (j + u) + \sum_{u=-J}^J (j-u) \right) c(u) e^{-2\pi i k_0 u / J} e^{-2\pi i k_0' u / J}$$

(32)

where the last identity holds because $\sum_{j=1}^J e^{-2\pi i k_0 j / J} = 0$ for $k_0 \neq k_0'$. This remaining term is bounded by

$$\text{Cov}(z_{nhk}, z_{nhk'}) \leq \frac{|X_{hh}|}{J} \left( \sum_{u=-J}^J \sum_{u=-J}^J \sum_{j=-J}^J \sum_{j=-J}^J c(u) e^{-2\pi i k_0 u / J} e^{-2\pi i k_0' u / J} \right)$$

$$= \frac{|X_{hh}|}{J} \sum_{u=-J}^J \sum_{u=-J}^J |c(u)||c(u)| \leq M$$

(33)

for some finite $M$ if

$$\sum_{u=-\infty}^\infty |c(u)| < \infty$$

(34)

Hence, the covariance between different frequencies vanishes for $n \rightarrow \infty$. The condition (34) holds, for example, when $c(u)$ is zero for $|u| > K$ for some $K < \infty$. If Eq. (34) holds, then Eq. (30) holds as well.

In sum, one can write in matrix notation for large $J$

$$\sum_n \mathcal{F}(T) \otimes X$$

(35)

where

$$\mathcal{F}(T)_{hk} = \delta_{h,k'} \sum_{u=-J}^J c(u) e^{-2\pi i k u / J}$$

(36)

the finite Fourier transform of the autocovariance corresponding to $T$. This matrix $\mathcal{F}(T)$ is diagonal. Since $c(u) = c(-u)$, it is also Hermitian and, hence, it is real. Finally, the diagonal elements of $\sum_n$ are non-negative since $\text{Cov}(z_{nhk}, z_{nhk'})$ is equal to the expectation of $z_{nhk} z_{nhk'}^*$, that is $|z_{nhk}|^2 \geq 0$ in expression (27). Since $X_{hh} > 0$, the diagonal elements of $\mathcal{F}(T)$ must be non-negative too. In sum, the matrix $\mathcal{F}(T)$ is real, diagonal and positive semi-definite. The diagonal elements of $\mathcal{F}(T)$ contain the power in the different frequencies.

When a sum of KP is considered, $N > 1$, the covariance in the space-time domain becomes

$$\text{Cov}(y_{nj}, y_{n'j'}) = \sum_{n=1}^N \sum_{n=1}^N \text{Cov}(X_{nh}, X_{nh'})$$

(37)

where $c_n(u)$ is the autocovariance function corresponding to the $n^{\text{th}}$ temporal covariance matrix. The above derivation now holds for each term in this sum, given that condition (34) holds for each autocovariance function $c_n$. Hence, we get for $J$ large

$$\sum_n \mathcal{F}(T_n) \otimes X_n$$

(38)

with $\mathcal{F}(T_n)$ for $n = 1, \ldots, N$ real, diagonal positive semi-definite matrices. Consequently, $\sum_n$ is approximately real and block-diagonal.

Appendix B. The case of perfectly stationary data: White noise

In this appendix the estimated covariance matrix $T_x$ of the DFT of white noise signals, that is, perfectly weakly stationary signals, is considered. More specifically, the dependence of $p_{\text{re}}$ and $p_{\text{off}}$ on the dimension $(K + 1)$ of the matrix is investigated. Consider the following simulated signal for one channel ($H=1$)

$$y_j \sim N(0, 1)$$

(39)

where the $y_j$ are independently identically distributed for $j = 1, \ldots, J$ and $l = 1, \ldots, L$. The Fourier transform $z_k$ is computed according to Eq.
Combining Eqs. (45) and (46) yields $T_{kk'} = \frac{1}{2} \left( z_{kk'} - z_{k'}k' \right)$.

Since $T_z$ is an unbiased estimator, the expectation of $T_{kk'}$ goes to zero for $J \to \infty$ if $k \neq k'$, see Appendix A. However, in the fraction $p_{off}$ in Eq. (16) the sum of squared absolute values, $|T_{kk'}|^2$, $k \neq k'$, of the off-diagonal elements is computed, and these do not expect equal to zero. Assuming that the expectation of $T_{kk'}$ for $k = k'$ is zero, the squared absolute value is the variance. Inserting the definition (6) of $z_{kk'}$ in Eq. (40) and assuming that $\zeta$ can be neglected for $J$ large, yields

$$E|T_{kk'}|^2 = \frac{1}{(L-1)^2} \sum_{j=1}^{L-1} \sum_{j'=1}^{L-1} \left| y_l^j e^{2\pi l j/k} e^{2\pi l j'/k} \right|^2,$$

where the last identity follows from writing the squared absolute value as the sum of the squared real and imaginary parts. Computing the expectations of the two squares in the last expression, using the expectation of the fourth moment of the standard normal distribution, yields for $k = k'$

$$E|T_{kk'}|^2 = \frac{1}{(L-1)^2} \frac{2K}{(L-1)^2} \left( \frac{L^2}{L^2} L + \frac{L^2}{L} \right) = \frac{1}{L-1} \left( \frac{L^2}{L} \right).$$

and for $k \neq k'$

$$E|T_{kk'}|^2 = \frac{1}{(L-1)^2} \frac{2KL}{(L-1)^2} \left( \frac{L^2}{L} L + \frac{L^2}{L} \right) = \frac{1}{L-1} \left( \frac{L^2}{L} \right).$$

Using these quantities and considering that $T_z$ is a $(K+1) \times (K+1)$ matrix, the matrix power of the entire matrix $T_z$ becomes

$$||T_z||^2 = (K+1) \times (K+1) \times \left( \frac{K+1}{L} \right).$$

The values of $p_{re}$ and $p_{off}$ now become approximately equal to

$$p_{re} = \frac{(K+1) \times (K+1) \times (K+1) \times (K+1)}{L} = \frac{1}{L-1} \left( \frac{L^2}{L} \right)$$

and

$$p_{off} = \frac{(K+1) \times (K+1) \times (K+1) \times (K+1)}{L} = \frac{1}{L-1} \left( \frac{L^2}{L} \right).$$

Combining Eqs. (45) and (46) yields $p_{re} + p_{off} = 1$, which also follows from Eq. (18) since in the case of white noise in one channel $p_{in} = 0$.

These computations show that, once the number of trials $L$ is fixed, the value for $p_{re}$ decreases when $K$ increases, or equivalently, when longer epoch length is chosen. The variance of the white noise does not have any influence, it is a scaling factor in both the numerator and the denominator of $p_{re}$ and $p_{off}$. The values reported in Table 2, which are based on white noise simulations, are slightly larger than values computed in Eqs. (45) and (46). This is due to the approximations made in Eqs. (42) and (43).

Appendix C. Rewriting a sum of 2 KP in space–frequency domain

The first estimated term, $D_1 \otimes X_n$, corresponding to the best rank 1 approximation $\text{vec}(X_n)\text{vec}(D_1)'$ of $S(\hat{T}_z)$ is (semi)-definite, since the matrix $T_z$ is positive (semi)-definite. Higher order terms $(n>1)$ contain indefinite matrices $X_n$ and $D_n$ which lack a covariance interpretation. In Bijma et al. (2005), the estimated sum of 2 KP was rewritten to a sum consisting of positive semi-definite matrices, which do obey a covariance interpretation. In the frequency domain a similar rewriting can be performed, although it can be done more efficient because of the diagonality of the matrices $D_n$. In the sequel, only the case $N=2$ is considered. For rewriting the sum of 2 KP, consider

$$\sum_{n=1}^{2} \text{vec}(X_n)\text{vec}(D_n)' = \text{vec}(X_1)\text{vec}(D_1)' + \text{vec}(X_2)\text{vec}(D_2)'$$

given that the determinant $ad-bc 
eq 0$. The rewritten matrices are indicated by tildes,

$$\tilde{X}_1 = a\text{vec}(X_1) + c\text{vec}(X_2)$$
$$\tilde{X}_2 = b\text{vec}(X_1) + d\text{vec}(X_2)$$
$$\tilde{D}_1 = \frac{1}{ad-bc} (d\text{vec}(D_1) - b\text{vec}(D_2))$$
$$\tilde{D}_2 = \frac{1}{ad-bc} (-c\text{vec}(D_1) + a\text{vec}(D_2)).$$

After rewriting, the normalization should still be intact, that is, the vectors $\text{vec}(\tilde{X}_1)$ and $\text{vec}(\tilde{X}_2)$ should still have length 1, yielding conditions

$$a^2 + c^2 = 1$$
$$b^2 + d^2 = 1.$$
Once $M$ and $m$ are determined, first a coarse search for $a$ and $b$ is performed, using increments of 0.05, over the ranges in Eqs. (50) and (51). Secondly, a fine search over a range of length 0.1 symmetrically around the initially found values of $a$ and $b$ is performed, using increments of 0.01. The values for $a$ and $b$ that maximize the total positivity in Eq. (52), found by this second search, are used as the optimal values, maximizing the positivity of the rewritten sum. In some cases multiple values for $a$ and $b$ lead to the maximal positivity of 400% in Eq. (52). This can happen, since a rank 2 expression can be expressed in many ways. In such a case, the smallest values for $a$ and $b$ that maximize the positivity are taken. It is also possible that the maximal positivity of 400% is not reached. In that case, the rewritten matrices are not strictly positive definite, though have some tiny negative eigenvalues. In those cases, the negative parts of $D_1$, $D_2$, $X_1$ and $X_2$ are set to zero, and the remaining matrices, $D_1$, $D_2$, $X_1$ and $X_2$, are positive semi-definite. The relative contribution of the two rewritten terms in the resulting positive semi-definite sum is calculated analogously to the method in Bijma et al. (2005).

References


