A solution to the dynamical inverse problem of EEG generation using spatiotemporal Kalman filtering

Andreas Galka\textsuperscript{1,2}, Tohru Ozaki\textsuperscript{2}, Okito Yamashita\textsuperscript{2}, Rolando Biscay\textsuperscript{3} and Pedro Valdés-Sosa\textsuperscript{4}

\textsuperscript{1} Institute of Experimental and Applied Physics, University of Kiel, 24098 Kiel, Germany
\textsuperscript{2} Institute of Statistical Mathematics (ISM), Minami-Azabu 4-6-7, Tokyo 106-8569, Japan
\textsuperscript{3} University of Havana, Ciudad Habana, Calle 34 No. 304-2, Cuba
\textsuperscript{4} Cuban Neuroscience Center, Ave 25 No. 5202 esquina 158 Cubanacán, POB 6880, 6990, Ciudad Habana, Cuba

Abstract

We present a new approach for estimating approximative solutions of the inverse problem of EEG generation. Unlike with previous approaches, we develop our solution in a dynamical framework; to this end we propose a new extension of Kalman filtering to the case of spatiotemporal dynamics. The inclusion of the temporal aspect of the problem opens up new perspectives for improved accuracy of the estimates and for data-driven design of dynamical models. For the purpose of estimating the parameters of such models from given data we employ a likelihood maximisation approach. The performance of the algorithm is demonstrated by application to simulated EEG time series. It is shown that the availability of appropriate dynamical models is a crucial precondition for obtaining improved inverse solutions.

1 Introduction

Recordings of electromagnetic fields on the scalp surface are well known to provide an important source of information about brain dynamics. Electrical potentials recorded on the scalp surface are very easy to measure at a set of electrodes attached to the skin; as a result multivariate electroencephalographic (EEG) time series are obtained.

It is reasonable to assume that the sources of these electromagnetic fields are electrical currents of charged ions which result from the electrical and chemical activity of the neurons in the gray matter of brain; these currents are sometimes referred to as "primary current density". In order to obtain more direct access to the dynamics governing the activity of these neurons it would be desirable to have direct estimates of these sources. The estimation of these sources from surface recordings has recently become a subject of intense research\textsuperscript{[1, 2, 3, 4]}.\footnote{Here, the superscript numbers represent references to related works.}

Two main classes of source models have been developed: "equivalent current dipole" approaches, in which the sources are represented by a relatively small number of focal sources\textsuperscript{[3]}, and "linear distributed" approaches, in which a large number of
locations in brain is assumed to contribute to the source distribution[3]. In the latter approach a discretisation of space within brain into a rectangular grid consisting of a finite number of "voxels" is usually employed. This paper exclusively deals with the linear distributed model approach.

It is a characteristic problem of distributed source models that a large number of unknown quantities has to be estimated from a much smaller number of measurements; as a consequence of this, we are facing a problem which does not possess a unique solution, known as "inverse problem". The number of measurements given at one instant of time may be as low as 19, if the standard 10-20 system of clinical EEG recordings is employed; by increasing the number of electrodes we may eventually obtain up to a few hundred measurements, but they will fail to provide an equivalent amount of independent information due to strong correlations between adjacent electrodes. On the other hand, the number of voxels will typically be several thousand, and furthermore at each voxel site a full three-dimensional current vector has to be modeled.

In order to identify a unique solution (i.e. an "inverse solution") additional constraints have to be imposed. Certain constraints can be obtained from neurophysiology [4]; as an example, it is reasonable to assume that only voxels within gray matter contribute to the generation of the electromagnetic fields; other constraints refer to the probable direction of local current vectors at specific locations. But such constraints do not suffice to remove the ambiguity of the inverse solution.

For this purpose much more restrictive constraints are needed, such as the minimum-norm constraint suggested by Hämäläinen and Ilmoniemi[6] or the maximum-smoothness constraint suggested by Pascual-Marqui[1]. We will refer to the resulting algorithms for obtaining inverse solutions as "instantaneous" algorithms, since they use only the data which is measured at one specific instant of time.

We do not intend to question the usefulness of these constraints, or of the resulting inverse solutions, but rather we would like to introduce a new approach to obtaining estimates of the distributed source model. Clearly also this approach will have to impose suitable constraints, but unlike with previous methods, these constraints will be expressed and motivated in a fully dynamical framework. By reformulating the inverse problem of EEG generation in a dynamical framework, namely as a spatiotemporal filtering problem, we will be able to exploit at each time point much more information than only given by the measurements recorded at that time point.

As the main tool for our task we will generalise the well-known Kalman filter (or Kalman-Bucy filter)[7] in such a way, that it can be applied to spatiotemporal filtering problems. We will demonstrate that Kalman filtering provides a natural framework for addressing the inverse problem of EEG generation by combining its spatial and temporal aspects. Through a simulation study it will become evident that the crucial element in this combination is given by the dynamical model according to which the dynamics of the voxel currents is assumed to evolve. If a very simple model is chosen, we will be able to obtain similar solutions as would also result from previous non-dynamical algorithms for solving the inverse problem; if the model contains additional information about the true dynamics, much better solutions are obtained.
2 Theoretical formulation

2.1 EEG inverse problem

We start from a rectangular grid of $N_v$ voxels covering the gray matter parts of the human brain; in this study inverse solutions will be confined to these voxels. In the particular discretisation which we will employ, we have $N_v = 3433$. At each voxel there is a local three-dimensional current vector $\mathbf{j}(v, t) = (j_x(v, t), j_y(v, t), j_z(v, t))^\top$, where $v$ is a voxel label, $t$ denotes time, and $^\top$ denotes matrix transposition. The column vector of all current vectors (i.e., for all gray-matter voxels) will be denoted by $\mathbf{J}(t) = (\mathbf{j}(1, t)^\top, \mathbf{j}(2, t)^\top, \ldots, \mathbf{j}(N_v, t)^\top)^\top$; it represents the dynamical state variable of the entire system.

These currents are mapped to the electroencephalographic signal (EEG), which is recorded at the scalp surface. The EEG at an individual electrode shall be denoted by $y(i, t)$, where $i$ is an electrode label; the $N_e$-dimensional column vector composed of all the electric potentials at all available electrodes shall be denoted by $\mathbf{Y}(t) = (y(1, t), y(2, t), \ldots, y(N_e, t))^\top$. In this study we assume that the 10-20 system is employed, such that $N_e = 19$. Potentials are assumed to refer to average reference, but other choices are possible.

We approximate the mapping from $\mathbf{J}$ to $\mathbf{Y}$ by a linear function whence it can be expressed as

$$\mathbf{Y}(t) = \mathbf{KJ}(t) + \mathbf{e}(t) \; .$$

(1)

Here $\mathbf{K}$ denotes the $N_e \times 3N_v$ transfer matrix, which we will call "lead field matrix". This matrix can approximately be calculated for a three-shell realistic head model and given electrode locations by the "boundary element method" [8, 1, 9].

It will be convenient for later use to define the individual contribution of each voxel to the vector of observations by $\mathbf{K}(v)\mathbf{j}(v, t)$, where $\mathbf{K}(v)$ is the $N_e \times 3$ matrix which results from extracting those three columns out of $\mathbf{K}$, which are multiplied with $\mathbf{j}(v, t)$ in the process of the multiplication of $\mathbf{K}$ and $\mathbf{J}(t)$. From this definition, equation 1 can also be written as

$$\mathbf{Y}(t) = \sum_{v=1}^{N_v} \mathbf{K}(v)\mathbf{j}(v, t) + \mathbf{e}(t) \; .$$

(2)

By $\mathbf{e}(t)$ we denote a vector of observational noise, which we assume to be white and Gaussian with zero mean and covariance matrix $\mathbf{C}_e = E(\mathbf{e}\mathbf{e}^\top)$. We will make the assumption that $\mathbf{C}_e$ has the simplest possible structure, namely

$$\mathbf{C}_e = \sigma_e^2 \mathbf{I}_{N_e} \; ,$$

(3)

where $\mathbf{I}_{N_e}$ denotes the $N_e \times N_e$ identity matrix, i.e., we assume that the observation noise is uncorrelated between all pairs of electrodes and of equal variance for all electrodes.

Equation 1 is part of the standard formulation of the inverse problem of EEG; in this paper we propose to interpret it as an observation equation in the framework of Kalman filtering.

2.2 Dynamical models of voxel currents

The Kalman filter provides the optimum tool for predicting, filtering and smoothing estimates of the state of dynamical systems which cannot be observed directly, but
only through an observation equation containing observational noise[7]. As a pre-condition both the equations governing the dynamics and the observation equation have to be known.

Since for the case of a the dynamics of human brain no well substantiated models for the spatiotemporal dynamics are known yet, we are faced with the problem of estimating suitable models from data. Clearly this constitutes a research task of enormous complexity which reaches far beyond the scope of this paper, therefore we will only be able to explore the very first and simplest approximations to such models.

Having defined a spatial discretisation by using a finite set of voxels, it is advisable to formulate dynamical models also with temporal discretisation; for simplicity we shall regard the basic time unit of this discretisation as equal to the sampling rate of the EEG recording. The corresponding time points will be labeled by \( t = 1, 2, 3, \ldots, N_t \).

In general form the dynamics of a set of \( N_v \) voxels may be described by nonlinear multivariate autoregressions given as

\[
\mathbf{J}(t) = \mathcal{F}(\mathbf{J}(t-1), \mathbf{J}(t-2), \ldots, \mathbf{J}(t-p) | \vartheta) + \eta(t),
\]

where \( p \) denotes the positive integer model order and \( \eta(t) \) denotes dynamical noise, which we assume to be white and Gaussian with zero mean and covariance matrix \( \mathbb{C}_\eta = E(\eta(t) \eta(t)^\top) \). When fitting such models to given data, \( \eta(t) \) represents a time series of innovations, i.e. components of the data which cannot be explained from the dynamics itself. It is desirable to find dynamical models which produce a white innovation time series, such that the process of modeling can be regarded as "temporal whitening".

\( \mathcal{F} \) denotes a function describing the deterministic part of the dynamics; it depends on a vector of parameters \( \vartheta \). This function may contain enormous internal complexity and a huge number of parameters (described by \( \vartheta \)), since it maps an input space of dimensionality \( 3N_v p \) to an output space of dimensionality \( 3N_v \).

The simplest non-trivial example of the class of autoregressions described by equation 4 is a linear multivariate autoregressive model of first order (AR(1)):

\[
\mathbf{J}(t) = A \mathbf{J}(t-1) + \eta(t).
\]

The parameter matrix \( A \) is of size \( (3N_v) \times (3N_v) \), which in our case is of the order of \( 10^9 \). This large number of parameters is still far too high to be estimated from real data, therefore we need additional reductions of model complexity. Also the practical application of Kalman filtering requires a simplified model structure. As an obviously reasonable simplification we propose that each voxel will interact only with its direct spatial neighbours; in a rectangular grid of voxels there will 6 direct neighbours for each voxel, except for those at the boundaries of the gray-matter parts of brain. Most elements of \( A \) become zero by this proposition. The dynamical model for each voxel becomes

\[
\mathbf{j}(v, t) = A_v \mathbf{j}(v, t-1) + \frac{1}{6} \mathbf{B}_v \sum_{v' \in \mathcal{N}(v)} \mathbf{j}(v', t-1) + \eta(t),
\]

where \( A_v \) and \( \mathbf{B}_v \) now are the autoregressive parameter \( 3 \times 3 \) matrices for self-interaction and nearest-neighbour interaction, respectively, and \( \mathcal{N}(v) \) denotes the set of labels of gray-matter voxels that are direct neighbours of voxel \( v \). By furthermore assuming total spatial homogeneity and isotropy for all pairs of neighbouring voxels, and also
for the three projections of local current vectors, we can ultimately reduce the number of parameters to two. The dynamical model for each voxel becomes

$$j(v,t) = a_1 l_3 j(v,t - 1) + \frac{b_1}{6} l_3 \sum_{v' \in \mathcal{N}(v)} j(v',t - 1) + \eta(t) \ ,$$  

(7)

where $l_3$ denotes the $3 \times 3$ identity matrix; now $a_1$ and $b_1$ are scalar autoregressive parameters. This model implements complete symmetry between voxels and also between projections of local currents; but clearly, this symmetry is not preserved by multiplication with the lead field matrix $K$. As a consequence of this, inverse solutions also will display non-symmetric behaviour with respect to voxels and projections.

It should be stressed that equation 7 corresponds to the specific choice for the parameter matrix $A$ that can be expressed as

$$A = (a_1 + b_1) l_{3N_v} - b_1 L \ ,$$  

(8)

where $L$ denotes a discrete spatial Laplacian operator defined by

$$L = \left(l_{N_v} - \frac{1}{6} N \right) \otimes l_3 \ .$$  

(9)

Here $N$ denotes a $N_v \times N_v$ matrix having $N_{v,v'} = 1$, if $v' \in \mathcal{N}(v)$ (i.e. voxels $v$ and $v'$ are neighbours) and 0 otherwise. By the symbol $\otimes$ Kronecker multiplication of matrices is denoted.

2.3 Spatial whitening

Application of Kalman filtering to the full spatiotemporal model as given by equation 5 would be infeasible in terms of computational time and memory demands due to the huge size of the parameter matrix $A$, i.e. if interactions between all pairs of voxels have to be considered. Only by decomposing the dynamics into a collection of small-scale dynamics centred at each voxel, as described by equation 6, this spatiotemporal filtering problem becomes tractable.

In order to apply this decomposition to the dynamics it is also necessary that the dynamical noise covariance matrix $C_\eta$ be a diagonal matrix, as assumed in equation 3 for the case of the observational noise covariance matrix $C_v$. But in the case of $C_\eta$ such assumption is much more problematical and will typically not be justified. Therefore we need an instantaneous data transformation

$$\tilde{J} = TJ \ ,$$  

(10)

such that in the corresponding dynamical model

$$\tilde{J}(t) = \tilde{A} \tilde{J}(t - 1) + \tilde{\eta}(t) \ .$$  

(11)

the dynamical noise covariance matrix $C_\tilde{\eta}$ becomes diagonal:

$$C_\tilde{\eta} = \sigma^2_\tilde{\eta} l_{3N_v} \ .$$  

(12)

In order to find a simple but efficient transformation we propose to extend the concept of temporal whitening to the spatial domain. A simple whitening approach in
temporal domain is given by differentiating the time series; we can perform a spatial differentiating step by applying the discrete Laplacian as defined by equation 9 to the dynamical state $\mathbf{J}$. At first sight this choice might seem somewhat arbitrary, albeit suggestive, but its justification will ultimately be given by successful performance in practical applications. If we choose $T = L$, equation 11 yields

$$\mathbf{J}(t) = L^{-1} \hat{\mathbf{A}} L \mathbf{J}(t-1) + L^{-1} \hat{\eta}(t). \quad (13)$$

But due to equation 8 (which by definition also describes the structure of $\hat{\mathbf{A}}$) the matrices $\hat{\mathbf{A}}$ and $L$ commute, such that by comparison with equation 5 we find that the choice $T = L$ corresponds to $\eta(t) = L^{-1} \hat{\eta}(t)$, and our assumption for the non-diagonal dynamical noise covariance matrix becomes

$$C_\eta = L^{-1} \mathbb{E}(\hat{\eta} \hat{\eta}^\dagger) (L^{-1})^\dagger = \sigma_\eta^2 (L^\dagger L)^{-1}. \quad (14)$$

If other dynamical models than described by equation 7 are assumed, $\hat{\mathbf{A}}$ and $L$ will generally not commute. Other transformations than the plain Laplacian $L$ may then be needed for perfect spatial whitening. But even in this case $L$ can be expected to make $C_\eta$ "more diagonal", and for this reason we will continue to employ it as spatial whitening transformation, until future research may identify better whitening transformations.

For the actual application of spatiotemporal Kalman filtering we will exclusively express the dynamics as $\hat{\mathbf{J}}(t)$, i.e. using the spatially whitened version. From now on we will omit the tilde.

2.4 Spatiotemporal Kalman filtering

Given the observation equation (equation 1) and the dynamical equation (in the case of linear first-order autoregression equation 11) we could in principle apply Kalman filtering according to its usual form, however, due to the very high dimensionality of the state variable $\mathbf{J}$ this will be infeasible. But by appropriate design of some modifications and extensions of the standard filtering procedure it is possible to decompose the very high-dimensional filtering problem into a set of coupled low-dimensional problems; this set is labeled by the voxel label $v$, i.e. it represents the spatial dimension of the problem. These modifications are not trivial, and we will defer a detailed discussion to a later paper. Here we will only briefly mention the crucial points.

Let $\hat{\mathbf{j}}(v, t-1 \mid t-1)$ denote the estimate of the local current vector at voxel $v$ at time $t-1$, i.e. the local state estimate, and $\mathbf{P}(v, t-1 \mid t-1)$ the corresponding estimate of the local error covariance matrix (i.e. a $3 \times 3$ matrix). The notation $\hat{\mathbf{j}}(t_1 \mid t_2)$ (where $t_1 \geq t_2$) represents the estimate of $\mathbf{j}$ at time $t_1$ which is based on all information having become available until (and including) time $t_2$. For each voxel the local state prediction is then given by

$$\hat{\mathbf{j}}(v, t \mid t-1) = \mathbf{A}_1 \hat{\mathbf{j}}(v, t-1 \mid t-1) + \frac{1}{6} \mathbf{B}_1 \sum_{v' \in N(v)} \hat{\mathbf{j}}(v', t-1 \mid t-1), \quad (15)$$

and the corresponding local prediction error covariance matrix can be approximated by

$$\mathbf{P}(v, t \mid t-1) = \mathbf{A}_1 \mathbf{P}(v, t-1 \mid t-1) \mathbf{A}^\dagger_1 + \sigma_\eta^2 \mathbf{I}_3. \quad (16)$$
Here we have assumed that the second term on the rhs of equation 15 behaves like an exogeneous variable, i.e. without contributing significantly to the prediction error covariance. The local state predictions $\hat{\mathbf{J}}(v, t | t - 1)$ for all voxels form the overall state prediction $\hat{\mathbf{J}}(t | t - 1)$, from which the observation prediction (for all electrodes) follows as

$$\hat{\mathbf{Y}}(t | t - 1) = \mathbf{R}(v) \hat{\mathbf{J}}(v, t | t - 1) .$$

(17)

The symbol $\mathbf{R}$ stands for the product $\mathbf{K} \mathbf{L}^{-1}$; multiplication by the inverse of the Laplacian is needed due to the spatial whitening approach described in section 2.3. The actual observation at time $t$ is $\mathbf{Y}(t)$, and the observation prediction error results as

$$\Delta \mathbf{Y}(t) = \mathbf{Y}(t) - \hat{\mathbf{Y}}(t | t - 1) .$$

(18)

We note that this multivariate time series represents the innovations of the actual observations, as opposed to the innovations of the (unobservable) system states, which have been denoted by $\eta(t)$ in equation 4. The corresponding observation prediction error covariance matrix can be approximated by

$$\mathbf{R}(t | t - 1) = \sum_{v=1}^{N_v} \mathbf{K}(v) \mathbf{P}(v, t | t - 1) \mathbf{K}(v) + \sigma^2 \mathbf{I}_{N_v} .$$

(19)

Here the direct summation over voxels seems to provide the appropriate generalisation of the standard expression to the spatiotemporal case. The matrices $\mathbf{K}(v)$ have been defined in section 2.1; again the bar refers to the fact that due to spatial whitening we have to replace $\mathbf{K}$ by $\mathbf{K} \mathbf{L}^{-1}$, before extracting the columns refering to voxel $v$. The Kalman gain matrix for voxel $v$ follows as

$$\mathbf{G}(v, t) = \mathbf{P}(v, t | t - 1) \mathbf{K}(v)^\dagger \mathbf{R}(t | t - 1)^{-1} ,$$

(20)

and finally the local state estimation and the corresponding local estimation error covariance matrix are given by

$$\hat{\mathbf{J}}(v, t | t) = \hat{\mathbf{J}}(v, t | t - 1) + \mathbf{G}(v, t) \Delta \mathbf{Y}(t)$$

(21)

and

$$\mathbf{P}(v, t | t) = (\mathbf{I}_3 - \mathbf{G}(v, t) \mathbf{K}(v)) \mathbf{P}(v, t | t - 1) ,$$

(22)

respectively. Equations 19 and 22 again rely on the validity of similar approximations as also employed in the case of equation 16.

This set of equations constitutes the new spatiotemporal Kalman filter. It should be stressed that equations 15, 16, 20, 21 and 22 are applied locally to each voxel, whereas only equation 17 requires a large-scale multiplication of the lead-field matrix $\mathbf{K}$ with the full $(3N_v)$-dimensional state vector $\mathbf{J}$.

For practical application of this filter to time series data initial values $\mathbf{J}(v, 0 | 0)$ and $\mathbf{P}(v, 0 | 0)$ are needed. As initial state estimates $\mathbf{J}(v, 0 | 0)$ we propose to use those values provided by a standard instantaneous inverse solution; in particular, we use the regularised maximum-smoothness algorithm as proposed by Pascual-Marqui[1] (also known as "LORETA"). The regularisation parameter is obtained by a generalised cross-validation (GCV) approach. According to our experience, the choice of initial values for $\mathbf{P}(v, 0 | 0)$ is rather non-critical; unity matrices can be used. In a more refined analysis the choice of initial values should be evaluated by the Maximum-Likelihood technique (see next section).
2.5 Parameter estimation

The general autoregressive model described by equation 4 depends on a parameter vector \( \theta \); in the largely simplified model given by equation 7 we have \( \theta = (a_1, b_1) \). Usually there will be no \textit{a priori} estimates of these parameters available. Furthermore we need estimates for the variances \( \sigma_t^2 \) and \( \sigma_n^2 \), as defined by equations 3 and 14.

Estimates for these dynamical parameters and variances should be obtained preferably from actual data. This can be accomplished within the framework of spatiotemporal Kalman filtering by likelihood maximisation. So far no successful applications of the principle of likelihood maximisation to the field of inverse problems have been reported; recently, Phillips \textit{et al.}[4] have presented an approach involving restricted Maximum Likelihood, but their approach does not involve dynamical modeling.

Assume that an EEG time series \( Y(t) \) is given, where \( t = 1, 2, \ldots, N_t \). At each time point the Kalman filter provides an observation prediction \( \hat{Y}(t) \), given by equation 17, and hence also observation innovations \( \Delta Y(t) \); if for these a multivariate Gaussian distribution with mean \( \bar{Y}(t) \) and covariance matrix \( R(t | t - 1) \) is assumed, the logarithm of the likelihood (i.e.\, log-likelihood, which is the proper function for numerical optimisation) immediately results as

\[
\log L = -\frac{1}{2} \sum_{t=1}^{N_t} \left( \log |R(t | t - 1)| + \Delta Y(t)^\top R(t | t - 1)^{-1} \Delta Y(t) \right) + \text{const.} \tag{23}
\]

Here \(|.|\) denotes matrix determinant.

Minimisation of \(-2 \log L\) as a function of model parameters and noise variances provides a well justified and efficient tool for obtaining estimates for unknown parameters. Also the effect of changing the structure of the model can be evaluated by monitoring the change of the log-likelihood.

3 Practical application

3.1 Spatial discretisation

This study employs a discretisation of brain into voxels which is based on a grid of \( 27 \times 23 \times 27 \) voxels (left/right \times up/down \times front/back). Out of these 16767 grid positions 8723 represent voxels actually covering the brain, out of which 3433 are gray-matter voxels. For the underlying brain geometry and the identification of the gray-matter voxels an averaged brain model was used which was derived from the average Probabilistic MRI Atlas produced by the Montreal Neurological Institute[10]. More details on this model can be found in [11] and references cited therein.

3.2 Design of simulation

We shall now present some results of applying the spatiotemporal Kalman filter, as presented in the first part of this paper, to time series data. It is a well-known problem of all algorithms providing inverse solutions that it is difficult to perform meaningful evaluations of the results and the performance, since for such evaluation we would need to know the true sources.

Inverse solutions obtained from real EEG time series typically display fluctuating spatiotemporal structures, but it is usually not possible to ascertain \textit{a posteriori} to
which extent these structures describe the true brain dynamics which was present during the recording. For this reason it is necessary to work with simulated data. If the primary currents at the gray-matter voxel sites are simulated, the corresponding EEG observations can be computed simply by multiplication with the lead field matrix, and consequently both the EEG time series and its true sources are known. But clearly these “true” sources will not represent any realistic brain dynamics. Nevertheless, we will now design a very simple simulated brain dynamics for the purpose of evaluating and comparing different inverse solutions.

A typical phenomenon of human brain dynamics is the presence of strong oscillations within local neighbourhoods, e.g. alpha activity in the visual cortex. If we regard the ”simple” dynamical model described by equation 7, where the parameters $a_1$ and $b_1$ are constant, as a device to generate simulated brain dynamics, driven by Gaussian white noise, we find that it will not produce such oscillations. In order to have oscillating behaviour in linear autoregressive models, a model order of at least $p = 2$ is needed. Alternatively the desired oscillation can be generated separately and imposed onto the brain dynamics through modulation of the system parameters. We will make use of this second alternative now.

By considering equations 5 and 8 it can be seen that the stability condition for the dynamical model described by equation 7 is approximately given by

$$|a_1 + b_1| < 1 .$$

(24)

If we choose to keep $b_1$ constant and let $a_1$ depend explicitly on time by

$$a_1(t) = a_c(1 + a_s \sin(2\pi ft))$$

(25)

and choose the parameters $b_1$, $a_c$ and $a_s$ such that $a_1 + b_1$ will repeatedly become larger than unity, we have defined a transiently unstable system. If this modulation of the parameter $a_1$ is confined only to those gray-matter voxels within a limited area of brain, this area will become a source of oscillations which spread out into neighbouring voxels. In this simulation we do not add dynamical noise, i.e. we are employing a linear, deterministic, explicitly time-dependent model. Alternatively the periodicity could also be generated by introducing additional state variables.

We define two areas in brain as centres for the generation of alpha-style oscillations, one in frontal brain and one in occipital brain. Each area is spherical and contains about 100 voxels; despite using equal radii the number is slightly different in both areas due to different content of non-gray-matter voxels. We choose the parameters $a_c$, $a_s$, $b_1$ and $f$ differently for both areas: The occipital oscillation has $f = 10.65$ Hz, and the frontal oscillation has $f = 8.05$ Hz (assuming a sampling rate of 256 Hz). Careful choice of these parameters is necessary in order to obtain an at least approximately stable dynamics of the simulated system. We choose for the occipital oscillation $a_c = 0.7$, $a_s = 0.75$ and $b_1 = 0.3$, and for the frontal oscillation $a_c = 0.9$, $a_s = 0.5$ and $b_1 = 0.1$. In this simulation the orientation of all vectors is the $(1,1,1)$ direction, and there is no interaction between different projections of vectors (as also assumed in equation 7). When simulating the system, an initial transient is discarded, and a multivariate time series of $N_t = 512$ points length is recorded. It represents the spatiotemporal dynamics of this simulation, i.e. for each of the $N_v = 3433$ voxels a 3-variate time series for the local current vector is recorded.

By multiplication by the lead field according to equation 1 we create artificial EEG recordings from this simulated dynamics; we assume a standard recording according
to the 10-20 system, average reference and a sampling rate of 256 Hz, i.e. with a length of the simulated time series of \( N_t = 512 \) two seconds of EEG can be represented. A small amount of Gaussian white noise is added to the pure EEG data (signal-to-noise ratio 100:1 in terms of standard deviations). The resulting EEG time series are shown in figure 1. As can be seen, they clearly display the two oscillations with their different frequencies, but in a quite blurred fashion.

### 3.3 Calculation of inverse solutions

For the EEG data shown in figure 1 we compute three inverse solutions: A "timeframe-by-timeframe" instantaneous inverse solution (using regularised LORETA); a dynamical inverse solution by using the spatiotemporal Kalman filter, as described in the first part of this paper, employing the simplest possible dynamical model, which is given by equation 7; and a dynamical inverse solution using the spatiotemporal Kalman filter, but employing the correct dynamical model, i.e. a "perfect" model.

For the application of the Kalman filter in the case of the simplest model, four parameters \((a_1, b_1, \sigma^2_\epsilon, \sigma^2_n)\) have to be chosen according to the principle of Maximum Likelihood, i.e. by minimising equation 23. This optimisation poses no particular problems, apart from being somewhat time-consuming, it even turns out that the likelihood as a function of these parameters behaves very smoothly, as long as the Kalman filter itself remains stable. It is also necessary to allow for a transient of the Kalman filter itself to die out, before the likelihood can be evaluated. In the case of our simulation the optimisation yields the values of the autoregressive parameters \(a_1 = 0.8117\) and \(b_1 = 0.1983\). These values are to be compared with the correct values for \(a_\epsilon\), \(a_n\) and \(b_1\) given in the previous subsection.

In this setting, the dynamical model is ignoring two important aspects of the true dynamics, namely the fact that the autoregressive parameter \(a_1\) is behaving differently for different groups of voxels, and that it shows explicit dependence on time for two groups of voxels. The dynamical model given by equation 7 is very primitive, and therefore it provides almost no additional information which could be used for the purpose of estimating an improved inverse solution.

As a contrast to this, we are also providing the same spatiotemporal Kalman filter with perfect knowledge about the true dynamics, i.e. not only does the filter know the correct values of the parameters as used in generating the simulated data, but also the information about the explicit time-dependence of \(a_1(t)\) for the two oscillating areas and the correct assignment of the voxels to these areas are given to the filter. Nevertheless, also in this case a Maximum Likelihood step is necessary in order to obtain estimates for \(\sigma^2_\epsilon, \sigma^2_n\).

### 3.4 Comparison of inverse solutions

The three different inverse solutions which we have obtained are given as functions of space and time: \(\mathbf{j}(v, t)\). We can compare them with the true solution (i.e. the simulated dynamics) \(\mathbf{j}(v, t)\) by forming a RMS error according to

\[
E = \sqrt{\frac{1}{N_v N_t} \sum_v \sum_t (\mathbf{j}(v, t) - \mathbf{j}(v, t))^2}.
\] (26)
This comparison yields $E = 2.2209$ for the instantaneous inverse solution using regularised LORETA (iIS), $E = 2.0856$ for the dynamical inverse solution with simplest model (dISs) and $E = 0.4949$ for the dynamical inverse solution with perfect model (dISp). These result indicate that compared to iIS, dISs achieves only a very small improvement, if any, but using dISp, i.e. knowing the perfect model, a much better estimation of the currents becomes possible.

In figures 2, 3 and 4 we present some graphical illustrations of the inverse solutions obtained in this simulation. Figure 2 shows the spatial appearance of true currents and inverse solutions at a fixed moment in time by displaying the maximum-intensity projection of the absolute values of the local current vectors by a gray-scale coding. Directions of vectors are not shown, since they were not allowed to vary in the simulation. For each case three projections of the spatial field of currents are shown. In subfigures A1, A2 and A3 the true currents from the simulated dynamics are shown. The two centres of simulated alpha activity can be seen clearly; most other areas of the brain remain inactive. The frontal centre shows a certain tendency to produce two neighbouring maxima of activity.

Subfigures B1, B2, B3 show the estimated currents according to iIS. It can be seen that the locations of the two main centres of activity are correctly reconstructed, but these two centres are much less focussed than in subfigures A1, A2, A3, rather does the active area spread out over most parts of the brain; in particular we see spurious structure extending into the temporal lobes.

Subfigures C1, C2, C3 show the estimated currents according to dISs. These results resemble very much those obtained by iIS (subfigures B1, B2, B3). This very close similarity, which also holds for the temporal domain (as shown in figure 4), is remarkable, since these two inverse solutions were obtained by completely different approaches, penalised Least Square (which is the essence of the regularised LORETA approach) in the case of iIS and spatiotemporal Kalman filtering in the case of dISs. Only the initial state estimates for the Kalman filter are borrowed from iIS. Nevertheless the resulting inverse solutions turn out to be very similar, as long as no specific information on the underlying dynamics is provided. The same phenomenon was also observed for inverse solutions obtained from various real EEG data sets.

Subfigures D1, D2, D3 show the estimated currents according to dISp. These results are much more similar to the true currents (subfigures A1, A2, A3) than to dISs or iIS: the two centres of activity are well focussed, and even a detail such as the presence of two maxima in the frontal centre is reproduced. By this result it is illustrated again that this technique has achieved a very good estimation of the currents. This success was obtained on the basis of knowing only the (simulated) EEG recording from $N_e = 19$ electrodes (as shown in figure 1) and the correct dynamical model, including the information about the time-dependency of the autoregressive parameter $a_1$ for certain groups of voxels.

Figure 3 shows absolute values of true currents and inverse solutions for all gray-matter voxels along a horizontal anterior-posterior line through brain which traverses both areas of simulated alpha activity. This line is situated approximately between the hemispheres and passes from the medial frontal gyri through the cingulate region and precuneus to the cuneus region; all these regions are gray-matter regions. Whereas figure 2 shows the situation at one fixed moment in time, here we have averaged the absolute value at each voxel with respect to time. As in figure 2, the letters A, B, C and D refer to true currents, iIS, dISs and dISp, respectively. In all figures the two centres of simulated alpha activity can be discerned clearly; it also becomes
evident that voxels between these two centres also display non-vanishing activity. This figure confirms in a more quantitative way the close similarity between iIS and dISs (subfigures B and C) on the one hand, and true currents and dISp (subfigures A and D) on the other hand.

Figure 4 shows the time series of absolute values of true currents and inverse solutions for two selected voxels, namely a voxel in the right medial frontal gyrus, situated in the middle of the frontal centre of alpha activity (left column of subfigures), and a voxel in the left superior frontal gyrus (right column of subfigures). This second voxel corresponds to the voxel labeled "11" in figure 3, so it is well outside of the centres of alpha activity. Again, the letters A, B, C and D refer to true currents, iIS, dISs and dISp, respectively. In subfigure A1 we see the strong alpha oscillation of this voxel. It is reproduced by iIS and dISs (subfigures B1 and C1), but its amplitude is significantly underestimated. In contrast to this, dISp (subfigure D1) reproduces the correct amplitude of this oscillation very well; in this subfigure we can also see the convergence of the Kalman filter during the first part of the data set.

In subfigure A2 we see the true current of a voxel which does not take part in any pronounced oscillation. The slight decrease of the current with time still is a transient behaviour resulting from the deterministic stable autoregressive dynamics, as employed in this simulation. Subfigures B2 and C2 show, that iIS and dISs incorrectly assign a spurious oscillation to this voxel, whereas dISp succeeds in approximately retrieving the correct dynamics. This result again illustrates the much sharper localisation which is achieved by dISp.

4 Conclusion

In this paper we have presented a new approach for approximately solving the inverse problem of EEG generation. We have demonstrated how the standard Kalman filter can be extended to the case of spatiotemporal dynamics; an essential precondition for this was the concept of spatial whitening which renders it possible to decompose a very high-dimensional filtering problem into a set of coupled low-dimensional problems. The application of Kalman filtering has the additional benefit of enabling the use of likelihood maximisation for the purpose of parameter estimation from given data.

We have demonstrated through numerical simulation that this new dynamical approach to estimating inverse solutions requires the availability of a suitable dynamical model. If only a very simple model is given, the quality of the resulting inverse solution will typically be very similar to the result of the standard instantaneous technique (regularised LORETA); at best, the dynamical inverse solution may be slightly better. On the other hand, a perfect model of the underlying dynamics enables the estimation of inverse solutions which are very similar to the true currents.

Perfect models will typically be unavailable in the analysis of real EEG data, but we are confident that by using the powerful tool of likelihood maximisation it will be possible to adapt initially simple models to given data, such that considerably improved models can be obtained. At the same time, these models themselves will be highly useful for various purposes of research and clinical diagnosis.

As another advantage of the dynamical approach to estimating inverse solutions we would like to mention the possibility to calculate spatial innovation maps by forming for each voxel and each vector projection the differences between predicted states (given by equations such as 15) and estimated states (given by equation 21). These innovations describe those components of the spatiotemporal dynamics which could
not be predicted by the given dynamical model, therefore they contain information about forces and processes driving the dynamics. Such information cannot be obtained by instantaneous techniques.

Altogether we expect that by reinterpreting the inverse problem of EEG generation as a spatiotemporal filtering problem it will become possible to obtain much more useful information from the analysis of EEG data than it was previously possible. Also for the combination of EEG data with other techniques for measuring temporally resolved information related to brain dynamics, such as MEG or fMRI, our work will open up new perspectives.

Acknowledgements

AG gratefully acknowledges support by the Deutsche Forschungsgemeinschaft (DFG) under contract GA 673/1-1.

References


Figure captions

Fig. 1: Simulated EEG recording for 19 standard electrodes according to the 10-20 system; electrode abbreviations are given on the vertical axis. The EEG potential is measured in arbitrary units, time is measured in seconds, assuming a sampling rate of the simulation of 256 Hz.

Fig. 2: Gray-scale coded representation of maximum-intensity projection of the three-dimensional field of absolute values of local currents for the gray-matter voxels of a model brain at a fixed point in time, using frontal projection (left column), top projection (middle column) and right projection (right column). Subfigures A1, A2, A3 show the original currents used in the simulation; subfigures B1, B2, B3 show the estimated currents according to the instantaneous inverse solution (regularised LORETA), subfigures C1, C2, C3 show the estimated currents according to the dynamical inverse solution using the simplest dynamical model, and subfigures D1, D2, D3 show the estimated currents according to the dynamical inverse solution using the perfect dynamical model.

Fig. 3: Absolute value of local currents for all gray-matter voxels along a horizontal anterior-posterior line through brain which traverses both areas of simulated alpha activity. Values at each voxel were averaged over time. Subfigure A shows values from the original currents used in the simulation, subfigure B shows results from estimated currents according to the instantaneous inverse solution (regularised LORETA), subfigure C1 shows results from estimated currents according to the dynamical inverse solution using the simplest dynamical model, and subfigure D1 shows results from estimated currents according to the dynamical inverse solution using the perfect dynamical model.

Fig. 4: Absolute value of local currents for a voxel in right medial frontal gyrus (left column of subfigures) and for a voxel in left superior frontal gyrus (right column of subfigures) versus time. Subfigures A1, A2 show values from the original currents used in the simulation, subfigures B1, B2 show results from estimated currents according to the instantaneous inverse solution (regularised LORETA), subfigures C1, C2 show results from estimated currents according to the dynamical inverse solution using the simplest dynamical model, and subfigures D1, D2 show results from estimated currents according to the dynamical inverse solution using the perfect dynamical model. Note the different scale on the vertical axis for left and right columns.
Figure 1
Figure 2
Figure 3

A

B

C

D

abs(current vector)

voxel number (in x-direction)
Figure 4

A1

B1

C1

D1

A2

B2

C2

D2