Author's Accepted Manuscript

Spatio-temporal reconstruction of brain dynamics from EEG with a Markov prior

Sofie Therese Hansen, Lars Kai Hansen

Cite this article as: Sofie Therese Hansen and Lars Kai Hansen, Spatio-tempora reconstruction of brain dynamics from EEG with a Markov prior, *NeuroImage* http://dx.doi.org/10.1016/j.neuroimage.2016.12.030

Spatio-temporal reconstruction of brain dynamics from EEG with a Markov prior

Sofie Therese Hansen[a,*], Lars Kai Hansen[a]

[a]*Cognitive Systems, Department of Applied Mathematics and Computer Science, Technical University of Denmark, Richard Petersens Plads, Building 324, DK-2800 Kgs. Lyngby*

Abstract

Electroencephalography (EEG) can capture brain dynamics in high temporal resolution. By projecting the scalp EEG signal back to its origin in the brain also high spatial resolution can be achieved. Source localized EEG therefore has potential to be a very powerful tool for understanding the functional dynamics of the brain. Solving the inverse problem of EEG is however highly ill-posed as there are many more potential locations of the EEG generators than EEG measurement points. Several well-known properties of brain dynamics can be exploited to alleviate this problem. More short ranging connections exist in the brain than long ranging, arguing for spatially focal sources. Additionally, recent work (Delorme et al., 2012) argues that EEG can be decomposed into components having sparse source distributions. On the temporal side both short and long term stationarity of brain activation are seen. We summarize these insights in an inverse solver, the so-called "Variational Garrote" (Kappen and Gómez, 2013). Using a Markov prior we can incorporate flexible degrees of temporal stationarity. Through spatial basis functions spatially smooth distributions are obtained. Sparsity of these are inherent to the Variational Garrote solver. We name our method the MarkoVG and demonstrate its ability to adapt to the temporal smoothness and spatial sparsity in simulated EEG data. Finally a benchmark EEG dataset is used to demonstrate MarkoVG's ability to recover non-stationary brain dynamics.

Keywords: Source reconstruction, Bayesian inference, the Variational Garrote, EEG, Inverse problem, Temporal dynamics

*Corresponding author

Email addresses: sofha@dtu.dk (Sofie Therese Hansen), lkai@dtu.dk (Lars Kai Hansen)

1. Introduction

The large body of event-related potential (ERP) studies demonstrates that EEG is a productive tool for detailed and accurate understanding of brain dynamics. While ERP studies are typically based on the native scalp electrode measures, imaging of human
[5] brain dynamics is gaining interest. Imaging by source reconstruction solves one of the main issues with EEG scalp studies, namely the limited spatial specificity due to volume conduction (Nunez et al., 1997). EEG imaging is obtained by solving an inverse problem, where the measured EEG scalp data is used to reconstruct the location and strength of the signal's cortical sources. However, the inverse problem is very ill-posed as the
[10] number of possible source locations exceeds the number of EEG electrodes by orders of magnitude (Hämäläinen and Ilmoniemi, 1994; Pascual-Marqui et al., 1994; Baillet et al., 2001; Hulbert and Adeli, 2013; De Ciantis and Lemieux, 2013). The inverse problem is based on a forward model which describes the mapping from a large set of hypothetical local sources to a smaller number of scalp electrodes. The forward model is constructed
[15] from electrophysiological first principles based on anatomical data and assumed values of conductivities of the various tissues; scull, scalp, etc. Attempts have been made at estimating the forward model from the EEG data, see e.g., (Stahlhut et al., 2011; Akalin Acar et al., 2016; Hansen et al., 2016). However, in the following we will assume the forward model known and focus on solving the inverse problem.

[20] Although no gold standard EEG inverse solver has been established, the field is converging on methods that employ spatial sparsity (Gorodnitsky and Rao, 1997; Wipf and Rao, 2007; Vega-Hernández et al., 2008; Friston et al., 2008; Zhang and Rao, 2011; Stahlhut et al., 2011; Montoya-Martinez et al., 2012; Gramfort et al., 2013; Hansen et al., 2013c; Hansen and Hansen, 2014; Andersen et al., 2014). Evidence was presented, in re-
[25] cent work (Delorme et al., 2012) that the instantaneous independent components of EEG signals are dipolar and localized. In particular it was shown that the residual variance after a dipole fit to the component scalp maps is less than 5% for large fractions of the independent components. ICA can thus provide sparse source distributions supporting the search for sets of localized sources, and motivates reconstruction algorithms that empha-
[30] size sparsity in contrast to the distributed spatial source patterns promoted in classical alternatives (Pascual-Marqui et al., 1994, 2002). The connectivity of the brain tissue

2

speaks in favor of focal and sparse solutions in general as there exists more short than long ranging cortical connections (Schüz and Braitenberg, 2002; Markov et al., 2011), enabling local coordination at typical EEG time scales. Sparsity can furthermore result
35 from averaging repetitions of stimuli leaving only focal or a sparse network of activity, as in ERP studies.

Imaging strategies can in general be divided into two categories, each having their own limitations. Dipole fits assume the number of active dipoles to be fixed and estimate their locations (Scherg and Von Cramon, 1985). Meaningful solutions thus rely on a qual-
40 ified guess at the number of active dipoles. In contrast, distributed imaging approaches estimate the source strength in a large number of source locations (Gorodnitsky et al., 1995; Friston et al., 2008). These methods thus avoid making subjective assumptions, but do render the EEG inverse problem underdetermined. Constraints are therefore needed to obtain unique solutions. These can, however, be formulated based on physiological
45 assumptions (Haufe et al., 2008) and spatial priors obtained from other neuroimaging modalities can be incorporated (Henson et al., 2010).

As we are interested in brain dynamics, the goal is to reconstruct not only sources at a given moment in time, but rather the spatio-temporal source distribution from a sequence of scalp measurements. To stabilize the solution it is useful to impose some level
50 of temporal smoothness. A basic scheme is to enforce that the locations of activity are fixed throughout an analysis window (Wipf and Rao, 2007; Friston et al., 2008; Ou et al., 2009; Zhang and Rao, 2011; Hansen et al., 2013c). While useful for short time windows, this may be less appropriate for more extended and non-stationary settings. Recently proposed methods enforce temporal coherency while also allowing for dynamic activation
55 patterns (Montoya-Martinez et al., 2012; Gramfort et al., 2013). These methods model the temporal dynamics more realistically by assuming brain areas to be sequentially or simultaneously activated during, e.g. a stimulus after which the activity returns to a baseline level (Gramfort et al., 2013). Both methods employ a mixed-norm scheme to recover what is hypothesized to be a structured sparsity pattern across time, see also
60 (Haufe et al., 2008; Gramfort et al., 2012). These types of convex relaxation schemes are very interesting and are frequently applied to solve inverse problems in general (Vega-Hernández et al., 2008). We have started investigations in a recent alternative for sparse

recovery proposed in (Kappen, 2011; Kappen and Gómez, 2013). The approach, called the Variational Garrote (VG), solves the sparse recovery problem directly without resorting

⁶⁵ to convex relaxation. In addition VG enables separation of the variables encoding the location and source magnitude estimation, which is relevant when a given dipole is active for an extended period (i.e., location is smooth in time) in which the activation magnitude involves high-frequency changes. Finally, a Bayesian inference scheme leads to a relatively low-complexity set of non-linear equations that are iterated towards the solution.

⁷⁰ The contribution of the present paper is to advance our understanding of this new algorithm. VG has been applied to EEG brain imaging, where it was extended to the spatio-temporal setting by assuming a fixed sparsity profile in time windows (Hansen et al., 2013c,b). In this presentation our aim is to replace the fixed sparsity model with a more flexible Markov prior, which in a preliminary unpublished workshop note was named

⁷⁵ "MarkoVG" (Hansen and Hansen, 2013). Here we further develop this model by including spatial basis functions to obtain focal smooth sources, and improve on the optimization scheme. In the following sections we analyze our proposed inverse solver and show how the model's degree of spatial and temporal sparsity can be adapted to fit the data. Finally we demonstrate MarkoVG's application to the spatio-temporal reconstruction of the EEG

⁸⁰ response to a face perception task.

1.1. Notation

In the following we have defined $\mathbf{X}^{\top}$ and X_{ij} as the transpose and the scalar element in row i and column j of the matrix $\mathbf{X}$, respectively. Capital bold thus indicates a matrix, a vector is in lower case and in bold font, while a scalar is in normal font, either in lower

⁸⁵ or upper case. The L_2-norm of $\mathbf{x}$ is denoted by $||\mathbf{x}||_2$.

2. Methods

2.1. The Variational Garrote

At the frequencies relevant for EEG acquisition the scalp EEG can be considered as a linear combination of the underlying brain activity (Baillet et al., 2001). The VG is therefore immediately applicable as it provides a framework to solve a linear inverse problem by imposing a "spike-and-slab" like representation (Ishwaran and Rao, 2005).

For T time samples the linear relation between N possible brain sources, $\mathbf{X} \in \mathbb{R}^{N \times T}$, and K EEG recordings, $\mathbf{Y} \in \mathbb{R}^{K \times T}$, is given by the forward model, $\mathbf{A} \in \mathbb{R}^{K \times N}$. This relation is modified in VG (Kappen, 2011) by introducing binary variables $\mathbf{S} \in \{0, 1\}^{N \times T}$ that dictate the spatio-temporal activation states (inactive or active), i.e.

$$Y_{kt} = \sum_{n=1}^{N} A_{kn} S_{nt} X_{nt} + E_{kt}. \tag{1}$$

We assume the noise, E_{kt}, to be i.i.d. with zero mean and normally distributed with scalar unknown precision β. As seen in eq. (1) there will for each dipolar location n and time sample t be an estimate of its state S_{nt} *and* its dipolar strength X_{nt}. The VG therefore supplies a framework which is highly flexible for including different priors into the solver.

2.2. Temporal coherence

We now impose temporal coherence through the binary variable S_{nt}. In a previous study we suggested to enforce a strict prior on the temporal smoothness by keeping S_{nt} fixed for each source for a given time window, while allowing for the activity strength, X_{nt}, to vary (Hansen et al., 2013c). Here we adapt the VG to provide a data-driven flexible degree of temporal smoothness of S_{nt} by imposing a Markov prior on this variable (Hansen and Hansen, 2013; Hansen et al., 2013a). For dipole location n the transition probabilities of S_{nt} thus depend on the activation state at time sample $t-1$ and is given by $\Gamma_{ji} = \mathrm{P}(S_{nt} = j | S_{n,t-1} = i)$, where $i, j = 0, 1$. The full transition matrix can be described by two parameters, as $\Gamma_{00} + \Gamma_{10} = 1$ and $\Gamma_{01} + \Gamma_{11} = 1$, and is given by

$$\mathbf{\Gamma} = \begin{bmatrix} \Gamma_{00} & \Gamma_{01} \\ \Gamma_{10} & \Gamma_{11} \end{bmatrix} = \begin{bmatrix} 1 - \Gamma_{10} & \Gamma_{01} \\ \Gamma_{10} & 1 - \Gamma_{01} \end{bmatrix}. \tag{2}$$

Through different combinations of Γ_{10} and Γ_{01} the Markov prior thus enables flexibility in both temporal smoothness and spatial sparsity. Temporal smoothness is for example achieved by having large probabilities of staying in a state, i.e. large Γ_{00} and Γ_{11}, while spatial sparsity is achieved by large probabilities of staying in or switching to an inactive state, i.e. large Γ_{00} and Γ_{01}.

Lucka et al. suggested to use hierarchical Bayesian methods to solve the inverse problem of EEG using fully Bayesian inference methods (Lucka et al., 2012). Practically,

Lucka et al. showed these methods' strengths in the single-measurements setup for especially deeply located sources. Here we follow Kappen et al., and instead solve the inverse problem by turning to approximate variational Bayesian inference (Kappen, 2011; Kappen and Gómez, 2013). First we define the posterior

$$p(\mathbf{S}, \mathbf{X}, \beta | \mathbf{D}, \mathbf{\Gamma}) = \frac{p(\mathbf{X}, \beta)p(\mathbf{S}|\mathbf{\Gamma})p(\mathbf{D}|\mathbf{X}, \mathbf{S}, \beta)}{p(\mathbf{D}|\mathbf{\Gamma})}, \tag{3}$$

where $\mathbf{D}$ is the data. Since we intend to optimize the posterior with respect to the source dipole activations we can ignore the denominator of eq. (3). Next we follow Kappen et al. by 1) assuming a flat prior on $\mathbf{X}$ and β, 2) marginalizing over S_{nt}, and 3) introducing the variational approximation $q(\mathbf{S}) = \prod_{n=1}^{N} q_{nt}(S_{nt})$, where $q_{nt}(S_{nt}) = M_{nt}S_{nt} + (1 - M_{nt})(1 - S_{nt})$. $M_{nt} \in [0, 1]$ describes the posterior probability of source n being active at time sample t, corresponding to the probability of S_{nt} being 1. The marginal log-likelihood is by these definitions

$$\log \sum_{\mathbf{S}} p(\mathbf{S}|\Gamma)p(\mathbf{D}|\mathbf{X}, \mathbf{S}, \beta) \geq -\sum_{\mathbf{S}} q(\mathbf{S}) \log \left(\frac{q(\mathbf{S})}{p(\mathbf{S}|\Gamma)p(\mathbf{D}|\mathbf{X}, \mathbf{S}, \beta)} \right) = -F(q, \mathbf{X}, \beta). \tag{4}$$

where F is an estimate of model evidence and is the so-called "variational free energy". As the free energy describes an upper bound on the negative log-likelihood it is minimized to find the optimal solution. As suggested in (Kappen, 2011; Kappen and Gómez, 2013) we apply a *dual formulation* to reduce the computational complexity by defining $Z_{kt} = \sum_n A_{kn}M_{nt}X_{nt}$ and Lagrange multipliers, λ_{kt}. The free energy with the Markovian prior can be derived from eq. (4) and is given by

$$F = -\frac{KT}{2} \log \frac{\beta}{2\pi} + \frac{\beta}{2} \sum_{t,k} (Y_{kt} - Z_{kt})^2 + \frac{K\beta}{2} \sum_{t,n} M_{nt}(1 - M_{nt})X_{nt}^2 \chi_{nn} \tag{5}$$

$$- \sum_{n,t} M_{nt} \log \frac{\Gamma_{10}}{\Gamma_{00}} + M_{n,t-1} \log \frac{\Gamma_{01}}{\Gamma_{00}} + (M_{nt}M_{n,t-1}) \log \frac{\Gamma_{00}\Gamma_{11}}{\Gamma_{01}\Gamma_{10}} \tag{6}$$

$$+ NT \log \frac{1}{\Gamma_{00}} + \sum_{n,t} [M_{nt} \log(M_{nt}) + (1 - M_{nt}) \log(1 - M_{nt})] \tag{7}$$

$$+ \sum_{t,k} \lambda_{kt} \left(Z_{kt} - \sum_n A_{kn}M_{nt}X_{nt} \right). \tag{8}$$

Here we define $\boldsymbol{\chi} \in \mathbb{R}^{N \times N}$ to be the covariance of the forward model $\mathbf{A}$.

Calculating the partial derivatives of the free energy and equating them to zero, yields

6

the following equation set

$$X_{nt} = \frac{1}{K\beta} \frac{1}{(1 - M_{nt})\chi_{nn}} \sum_k \lambda_{kt} A_{kn}, \quad Z_{kt} = Y_{kt} - \frac{1}{\beta}\lambda_{kt}, \tag{9}$$

$$\beta = \frac{1}{TK} \sum_{t,k,c} \lambda_{kt} \lambda_{ct} C_{kct} \tag{10}$$

$$\text{where} \quad C_{kct} = \delta_{kc} + \frac{1}{K} \sum_n \frac{M_{nt}}{(1 - M_{nt})\chi_{nn}} A_{kn} A_{cn}, \tag{11}$$

$$\lambda_{ct} = \beta \hat{Y}_{ct} \text{ def. } \sum_c C_{kct} \hat{Y}_{ct} = Y_{kt}, \tag{12}$$

$$M_{nt} = \sigma \left(\frac{K\beta}{2} \chi_{nn} X_{nt}^2 + \gamma_1 + \gamma_2 \left(M_{n,t-1} + M_{n,t+1} \right) \right), \tag{13}$$

where $\sigma(x) = (1 + \exp(-x))^{-1}$ and where the estimated source strength of source n in time sample t is given by $V_{nt} = M_{nt} X_{nt}$. While Kappen et al. solve the equation set by iteration, we implement gradient descent for the variational mean to ensure convergence. The complexity of the equations set is dominated by the computation of the tensor $\mathbf{C}$ and its inversion; which are of order NK^2T and K^3T, respectively.

Inspecting the modified VG equations, it is clear that the combination of the Markov parameters $\gamma_1 = \log\left(\frac{\Gamma_{10}\Gamma_{01}}{\Gamma_{00}^2}\right)$ and $\gamma_2 = \log\left(\frac{\Gamma_{00}\Gamma_{11}}{\Gamma_{01}\Gamma_{10}}\right)$ dictates how sparse and smooth the solution will be. The parameter γ_2 thus determines the degree of temporal smoothness, and γ_1 corresponds to a sparsity control parameter, where more negative values will yield more sparse solutions. We note that if $\Gamma_{01} + \Gamma_{10} = 1$ then $\gamma_2 = 0$ and the original VG formulation of the variational mean is obtained.

In Fig. 1 we show how applying different combinations of sparsity and smoothness affect the solution in a simulation. In the example we synthesized 25 time samples containing non-stationary support on the activation for two out of 500 sources (Fig. 1A and B). A random forward model of size 50×500 was used to project the signal to 50 observations. From Fig. 1C and D we see that if the sparsity is too high some of the relevant time samples are turned off and that this can only be partly remedied by a high temporal smoothness. If on the other hand the sparsity is too low we obtain activity in other than the relevant source locations. It is furthermore evident from Fig. 1 that having too little temporal smoothness will cause the solver to miss activation in the time samples of low magnitude activity. However, the temporal smoothness must not be too large or activity outside the activated period will emerge. We can therefore conclude that only the right
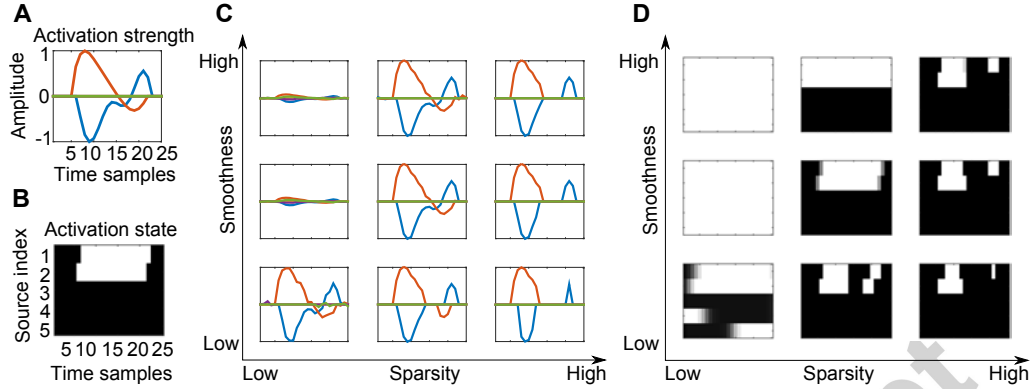
7

Figure 1: Example of sparsity and smoothness dependency. A) The simulated signal. Two out of 500 sources are active. These were projected through a random forward model with 50 observations and added with noise to give a signal-to-noise ratio (SNR) of 10 dB. B) True activation states for five of the sources. White indicate active state while black illustrates an inactive state. Source 1 and 2 were active in some time samples, and the remaining 498 were completely turned off. C) and D) The MarkoVG estimated signal and their activation state for different combinations of sparsity and smoothness degree. The parameter setting for the solution in the mid insets of C) and D) was found through four-fold cross-validation on the 50 observations.

amount of temporal smoothness improves the solution. Importantly we also demonstrate that we are able to match the true signal's properties (Fig. 1A and B) using four fold cross-validation to find the optimal level of sparsity and smoothness (mid inset in Fig. 125 1C and D).

2.3. Spatial coherence

EEG activity arises when regional active neurons are active in synchrony (Baillet et al., 2001) and therefore many EEG inverse solvers incorporate an assumption of spatial smoothness (Phillips et al., 2002; Pascual-Marqui et al., 2002; Friston et al., 2008; Haufe 130 et al., 2008). To obtain a spatially smooth source distribution we introduce spatial basis functions following the implementation described in the multiple sparse priors model (MSP) (Friston et al., 2008). The basis functions are based on the adjacency matrix, which describes how the source space is connected. The connectivity contained in the adjacency matrix is propagated to neighbors' neighbors in eight steps and finally a thresholding is 135 performed. This translates to basis functions extending from their center to maximally
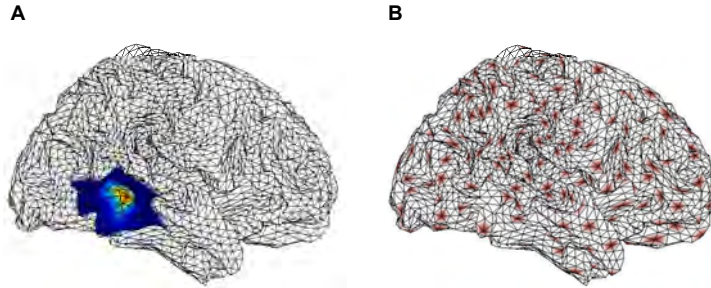
ACCEPTED MANUSCRIPT

Figure 2: The spatial basis functions. A) Example of one basis function's spatial distribution. Red/blue indicate high/low numeric activity. B) Centers of the 776 sampled basis functions.

their eighth-order neighbors (Fig. 2A). The degree of smoothness is controlled by a parameter that is set as suggested in (Friston et al., 2008). With this setting each basis function covers between 98-128 dipoles when the cortex surface has been segmented into a mesh of 8196 nodes.

¹⁴⁰   In order to reduce the complexity of the inverse solver not all dipoles (or nodes) in the source space will serve as centers of basis functions. Friston et al. placed the centers by first sampling 256 evenly spaced source indices (Friston et al., 2008), and then also included their symmetrically located sources on the other hemisphere. Finally the hemispherical symmetric centers were paired to create 256 additional basis functions. In ¹⁴⁵ total 768 basis function were created with 512 representing unilateral activity and 256 bilateral activity (Friston et al., 2008).

To ensure an even distribution in space, rather than in the source index, we propose to sample the basis function centers based on the connectivity of the mesh and thereby obtain better coverage of the cortex. We thus let the adjacency matrix determine whether ¹⁵⁰ a randomly sampled center should be included. The precise requirement is that there must be at least three vertices between all centers. By seeding the random generator the locations of the basis function centers are controlled. With our applied seeding, 776 centers are obtained, shown in Fig. 2B. In both sampling techniques all locations are part of more than 1 basis function. In the original method each source location is included in ¹⁵⁵ 3 to 17 basis functions while in our method each source location is a part of 8 to 16 basis functions.

9

2.4. MarkoVG

The inference scheme for VG with a Markovian prior is explained in eqs. (1)-(13) and seen implemented in https://github.com/STherese/VG_inverse_solvers. The
160 spatial basis functions are included simply by projecting the forward model $\mathbf{A}$ onto these. This produces the reduced forward model $\mathbf{A}_{\text{basis}} = \mathbf{AB}$, where $\mathbf{B}$ contains the basis functions in the columns. The optimum level of sparsity and smoothness is determined through four-fold cross-validation on the electrodes. The free energy is used to estimate the optimum setting for each fold, and the median across these defines the parameter
165 setting. Since we apply 70 electrodes 17-18 electrodes are in each fold, and we therefore believe that it is likely that the brain activity will be seen to some degree by all four folds. We compare our proposed method with the below three inverse solvers, that all produce temporally stationary source distributions.

2.5. Multiple sparse priors

170 MSP (Friston et al., 2008) specifies another Bayesian approach of finding sparse source distributions. As previously described, inference is based on a number of spatial basis functions with compact support, these are pruned or incorporated in the solution through a restricted maximum likelihood procedure. Smooth temporal source distributions are obtained by creating temporal projectors from the EEG signal. In effect this creates
175 rather stationary temporal activation patterns. The implementation used in the following experiments is from the SPM12 software (Ashburner et al., 2014).

2.6. T-MSBL

T-MSBL (Zhang and Rao, 2011) is an extension of the multiple measurement vectors (MMV) sparse Bayesian learning (SBL) (Wipf and Rao, 2007) method that ex-
180 ploits temporal correlation to obtain smooth temporal dynamics. T-MSBL assumes a block-structure where temporal correlations are modeled in the blocks. Automatic relevance determination (ARD) (Hansen and Rasmussen, 1994; MacKay, 1995) is applied to identify the active sources (blocks) and prune the irrelevant. In the following experiments we employ the implementation provided in the toolbox by Zhilin Zhang
185 http://dsp.ucsd.edu/~zhilin/TMSBL.html. We apply two versions of T-MSBL. In the first version we follow the recommendations of the toolbox and set the noise level to

10

"large" if SNR< 6 dB and to "mild" if SNR≥6. These two intervals translate into two numerical noise levels. We refer to the first version as "T-MSBL" which is partly favored in the simulations as the true noise level is provided. In the second version we perform four-fold cross-validation to estimate the regularization parameter, we call this version "T-MSBL cross".

2.7. M-FOCUSS

The FOCal Underdetermined System Solver (FOCUSS) employs a reweighted norm minimization and finds sparse solutions by defining the regularization norm to be equal to or less than 1 (Gorodnitsky and Rao, 1997). M-FOCUSS is an MMV extension developed in (Cotter et al., 2005) and also here extended to be applicable to noisy data. The latter version is the so-called regularized M-FOCUSS which performs iterative weighting using the diagonal matrix $\mathbf{W} \in \mathbf{R}^{N \times N}$ to find the dipole estimates $\mathbf{X}$, i.e. in iteration it

$$\mathbf{W}_{n,n}^{(it)} = ||X_{n,:}^{(it-1)}||_2^{1-p/2}, \text{ with } p \in [0, 2] \tag{14}$$

$$\mathbf{X}^{(it)} = \mathbf{W}^{(it)}\mathbf{W}^{(it)^\top}\mathbf{A}^\top(\mathbf{A}\mathbf{W}^{(it)}\mathbf{W}^{(it)^\top}\mathbf{A}^\top + \lambda\mathbf{I})^{-1}\mathbf{Y}. \tag{15}$$

We use the implementation of the regularized M-FOCUSS provided in the same toolbox as the T-MSBL algorithm. The regularization parameter, λ, can be approximated by the noise variance of the data (Zhang and Rao, 2011) and in the simulations we therefore use the exact noise variance for this parameter. The M-FOCUSS is thus favored in the simulations. As suggested by Cotter et al. we set the norm to be $p = 0.8$, which should, according to the authors, provide a reasonable balance between being sparse and not having to many local minima.

2.8. Simulations

We first evaluate MarkoVG in a simulation study. In line with previous EEG simulation studies (Friston et al., 2008; Stahlhut et al., 2011; Montoya-Martinez et al., 2012; Gramfort et al., 2013) we generated synthetic EEG signals by randomly planting one to four sources and projecting their temporal dynamics through a forward model. The sources were modeled as having a spatial distribution given by the earlier described basis functions where the centers could be placed in any of the dipoles of the mesh (and not only in the reduced set used for reconstruction). The source signal was projected to scalp
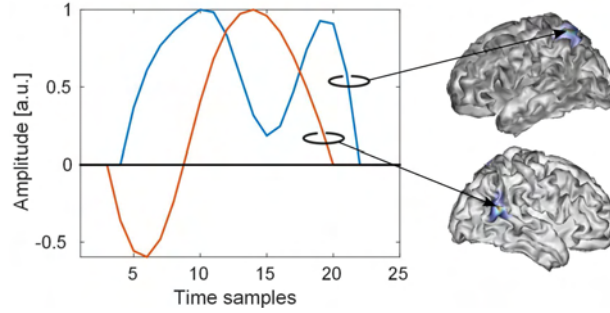
11

Figure 3: Example of two simulated sources. In each repetition one to four sources were planted.

EEG electrodes through a forward model generated for subject "A" in the real data experiment described below. The forward model contains the projection of 8196 dipoles to 70 EEG electrodes.

The temporal dynamics of the sources were each generated from random white noise which was low pass filtered to yield frequency content up to 20 Hz. We created a signal of 25 time samples. To obtain varying degrees of non-stationarity we only kept the activity in the mid section of these time samples. An example of the temporal dynamics of a set of sources is shown in Fig. 3. Noise was added to yield SNRs of 0 to 14 dB.

Across the applied SNRs 100 data sets were used to compare MarkoVG to MSP, T-MSBL and M-FOCUSS with all performing reconstruction using the 776 earlier described basis functions. The performance was judged based on a source retrieval score called the F_1-measure (Rijsbergen, 1979; Makhoul et al., 1999), as well as the source localization error. The F_1-measure is defined as

$$F_1\text{-}measure = \frac{2 \cdot \text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} = \frac{2 \cdot \text{TP}}{TP + FP + P}, \tag{16}$$

where TP, FP and P are the true, false and actual positives, respectively. It is noted that this is a rather strict measure as only a perfect correspondence between planted and estimated activity will yield perfect source retrieval, i.e. $F_1\text{-}measure = 1$. Since we only used a subset of the possible basis function centers (776 out of 8196) to reconstruct from, perfect reconstruction was only obtainable when the planted sources were basis functions from the subset. However, since the basis functions describe locally coherent activation some of the actual activity can be recovered even if the planted source component is not directly contained in the set used for reconstruction.

12

We defined the localization error as the Euclidean distance between each estimated
₂₂₅ source and the nearest planted source. The reported error is the average over all estimated
sources and all time samples containing simulated activity. We proceeded in this way as
considering only the maximum magnitude source from the estimation would disregard
any spurious activity located far from the true sources.

2.9. Benchmark EEG data

₂₃₀ To further investigate MarkoVG we applied it to EEG recorded during a well studied
paradigm, namely the multi-subject multimodal dataset studying face recognition (Wake-
man and Henson, 2015). Images of famous faces, unfamiliar faces and scrambled faces
were presented to 19 subjects in six runs of 7.5 minutes. We investigated face perception
from the 70-channel EEG data recorded in run 1 for three subjects, here termed "A",
₂₃₅ "B" and "C". In this run the subjects were presented with approximately 50 famous, 50
unfamiliar and 50 scrambled faces. As we are interested in finding the response to faces
we averaged over the two face conditions and subtracted the average of the scrambled face
condition (see EEG sensor data in Supplementary Fig. 1 and 2A). For further informa-
tion on the experimental setup used in the data collection we refer to the documentation
₂₄₀ provided by Wakeman et al. (Wakeman and Henson, 2015). We built forward models in
SPM8 using a three layered boundary element method head model (Phillips, 2000). The
head model was the result of segmenting T1-weighted MRI scans of the three investigated
subjects.

It has been shown that face perception exhibits partially bilateral activation (Eimer
₂₄₅ and McCarthy, 1999; Henson et al., 2009). In the source reconstruction we therefore
employed the basis functions set described by Friston et al. comprising both unilateral
and bilateral basis functions. For comparison we also show the solutions obtained using
MSP, T-MSBL and M-FOCUSS; also with the basis function set described by Friston
et al. As the EEG signal was averaged over many repetitions we judged the noise level
₂₅₀ needed for T-MSBL to be "mild". For M-FOCUSS an estimate of the noise variance was
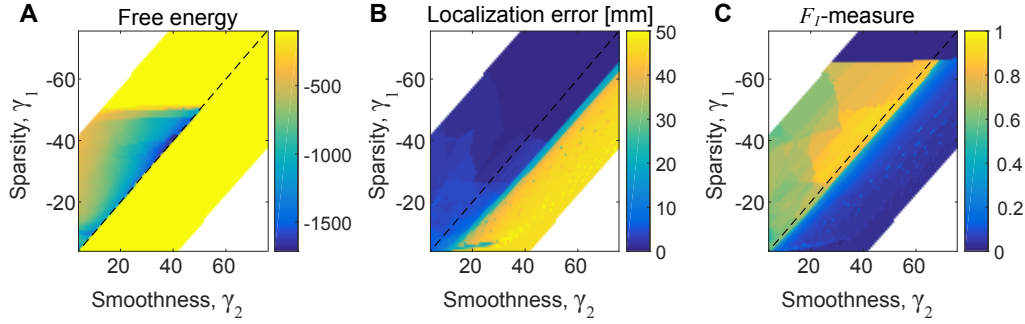calculated based on a 100 ms pre-stimulus window.

13

Figure 4: Example of the dependence of sparsity and smoothness on the MarkoVG solution illustrated on simulated data. The simulated source distribution for this example can be seen in Fig. 3 and was created using an EEG forward model. The black dashed line indicates $\gamma_1 = -\gamma_2$. White areas in the plots illustrate where combinations of γ_1 and γ_2 are not meaningful, see text. A) The free energy calculated on the validation sets in a four-fold cross-validation scheme. Shown is the mean across these folds. B) The localization error averaged across time and estimated sources, and C) F_1-*measure* of the MarkoVG solution; 0 indicates no correct sources are retrieved/many false sources are retrieved and 1 indicates all correct and no false sources are recovered.

3. Results

3.1. Simulations

In Fig. 4 we investigated the effect different combinations of sparsity and smoothness
255 levels have on the MarkoVG solution. Note that these combinations have to respect the specification of the prior probabilities, i.e. the columns of the matrix in eq. (2) must sum to 1 and have elements with values between 0 and 1. The relevant combinations of sparsity and temporal smoothness were therefore investigated in the band shown in Fig. 4.

260 Fig. 4A demonstrates that low free energy calculated on the validation folds in a cross-validation scheme coincides with low localization error (Fig. 4B) and high F_1-*measure* (Fig. 4C). This is evidence that the free energy can be used to optimize parameters for performance. The optimal solution was located in the vicinity of the dashed line where $\gamma_2 = -\gamma_1$ and more precisely just above the dashed line, particularly when considering
265 the F_1-*measure*. In the following we assumed the relation $\gamma_2 = -0.9\gamma_1$, and thus reduced the search space of the optimal parameter setting from two to one parameter. Some intuition on the implications of the defined relation can be gained by propagating the
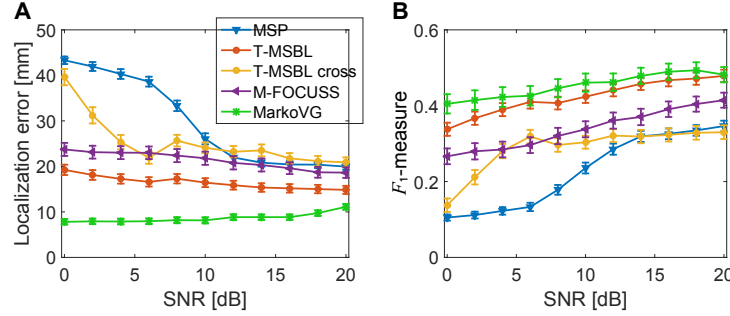
14

Figure 5: Performance on simulated data; created with a real EEG forward model. 100 repetitions were run with simulated activity consisting of one to four sources (basis functions) randomly placed on the cortex, each having non-stationary temporal dynamics, see example in Fig. 3. Errorbars indicate standard error of the mean. A) Localization error averaged across time and sources. B) The source retrieval score, F_1-measure; 1 indicates optimal retrieval. Note, perfect performance is not expected as the locations of the planted sources were drawn from the entire mesh and sought reconstructed based on a subset.

relation to the transition probabilities, i.e., setting $\gamma_2 = -\gamma_1$ implies that the probability of staying in an inactive and active state are equal ($\Gamma_{00} = \Gamma_{11}$). Hence, implies that 270 transitioning from an inactive to an active state is as likely as the reverse. In this case there is no sparsity bias, only temporal smoothness is enforced if $\Gamma_{00} > 0.5$. To promote sparse solutions we heuristically applied the factor 0.9, based on complete scans of the parameter space as seen in Fig. 4.

In Fig. 5 we applied the above mentioned sparsity-smoothness relation and per-275 formed cross-validation on one parameter for MarkoVG. We compared MarkoVG to MSP, T-MSBL, T-MSBL with cross-validation, and M-FOCUSS, in 100 repetitions and with different levels of noise. It can be observed that MarkoVG achieved the best localization error and that MarkoVG and T-MSBL outperformed the other methods with respect to the F_1-measure. We further observe that cross-validation was not effective for T-MSBL.

280 *3.2. Face perception EEG data*

Fig. 6 presents the source distributions of the face perception data as estimated by MSP, T-MSBL, M-FOCUSS and MarkoVG for three subjects. The temporal dynamics of the basis function having highest acitivity in the time interval 130 to 200 ms after stimuli onset is shown in blue in the top panel with the locations marked in blue in the lower
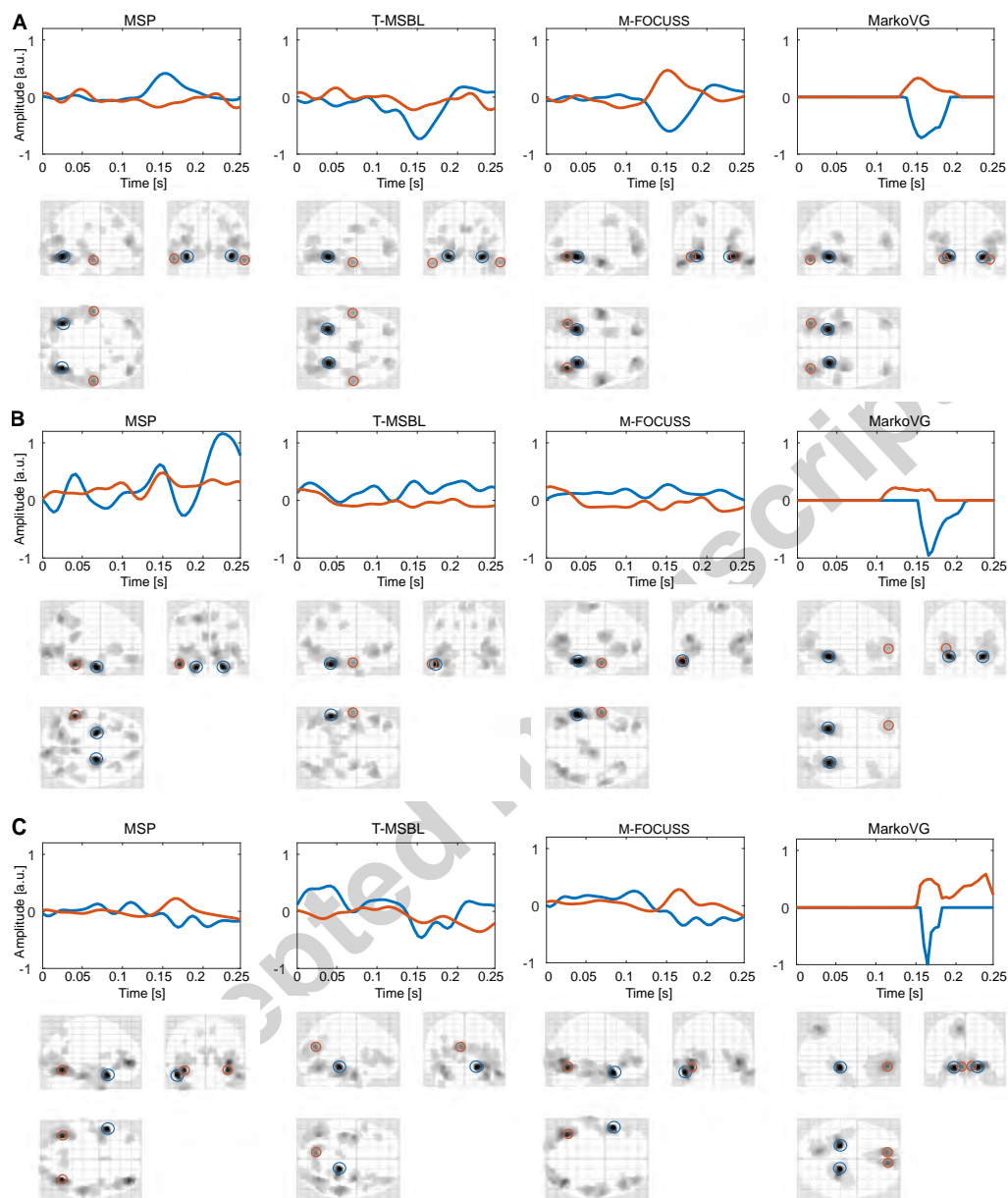
15

Figure 6: Reconstruction of EEG face perception data for subject "A", "B" and "C". The two strongest basis functions' temporal dynamics (top) and their locations (bottom) for MSP, T-MSBL, M-FOCUSS and MarkoVG. The highest magnitude source is shown in blue and second highest in red. The highest magnitude basis functions were for several of the examples bilateral. The glass brains show the activity for the 512 maximum magnitude dipoles at the time sample with highest magnitude source, here at 151-170 ms after stimuli onset. Source strengths are directly comparable between sources, subjects and solvers but are in arbitrary unit due to lack of units of the forward model Litvak (2016)
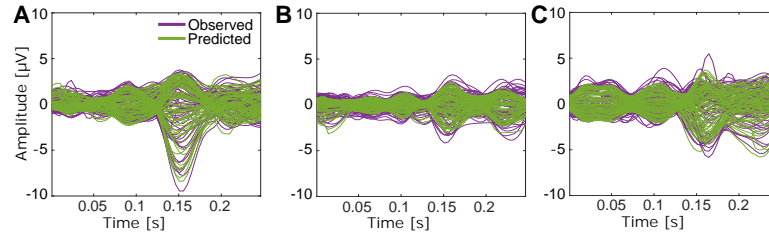
16

Figure 7: Averaged ERP as observed and as predicted by MarkoVG. Shown are all electrodes for the time interval 0 to 250 ms after stimuli onset for subjects "A", "B" and "C".

₂₈₅ panel. The second largest valued basis function is similarly shown in red. These basis functions were bilateral in several of the shown examples.

Most solvers recovered contrast activity in or close to the expected areas, i.e. the left and right occipital face area (OFA) and fusiform face area (FFA). More specifically MarkoVG placed the strongest activation (blue circles) in or near the FFA for all sub-₂₉₀ jects. Focusing on subject "A" the strongest basis function for T-MSBL, M-FOCUSS and MarkoVG were located in the FFAs, marked with blue circles in Fig. 6A. All four methods found activation in the OFAs; MSP had its strongest activated sources close to the OFA, the same for the second strongest activation for M-FOCUSS and finally MarkoVG had its second strongest activation in the OFA.

₂₉₅ The presented inverse solvers generally showed the well-known temporal response to viewing faces, i.e. the N170 ERP component. The N170 component normally appears 130-200 ms after presentation of a face (Itier and Taylor, 2004) The solvers peaked between 150 ms and 170 ms after stimuli, however, the N170 peak was less defined for MSP, TMSBL and M-FOCUSS for subject "B". It is further noted that MarkoVG differed ₃₀₀ from the other methods by being temporally as well as spatially more sparse.

In Fig. 7 we show the averaged observed ERP as well as the ERP predicted by MarkoVG. These are quite similar, however with a slight bias towards zero of the MarkoVG predicted ERPs. This is similarly demonstrated for MSP, T-MSBL and M-FOCUSS in Supplementary Fig. 1. To avoid scaling issues in a comparison we show in supplementary ₃₀₅ Fig. 2B the temporal correlations across channels between the observed and predicted EEG signals for subject "A". These were again similar for all methods.

4. Discussion

Solving the ill-posed inverse problem of EEG and obtaining detailed spatio-temporal knowledge of cognitive processes require us to make relevant prior assumptions on the
₃₁₀ solution. Such assumptions should be based on prior knowledge of the brain, for example, from brain anatomy and physiology. Common assumptions include on the spatial side sparsity and smoothness, meaning that the source distribution of interest is believed to consist of relatively few source patches, each having temporally coherent source strength. Sparsity is a common assumption when solving ill-posed inverse problems in general as it
₃₁₅ mitigates the non-uniqueness of the problem (Donoho et al., 2006). In EEG imaging it is, as mentioned earlier, further motivated by the existence of more short ranging connections than long ranging. Sparsity has previously been obtained through regularization of the inverse problem, e.g., by imposing the L_p-norm, where $p \leq 1$ (Gorodnitsky and Rao, 1997; Matsuura and Okabe, 1995). When $p = 1$ the problem is still convex however
₃₂₀ the correct solution is only guaranteed under certain conditions that are usually not met because of the highly correlated columns of the EEG forward model (Donoho et al., 2006). Furthermore, studies have shown that the L_1-norm produces spurious sources (Liu et al., 2004; Hansen et al., 2013c). Employing L_p-norms where $p < 1$ will produce more sparse solutions, it however also implies non-convexity.

₃₂₅ Bayesian approximations such as SBL (Tipping, 2001) should produce fewer local minima (Zhang and Rao, 2011) and are therefore also very promising. ARD is in SBL used to prune away variables by assigning a hyperparameter to each variable dictating whether to keep or discard the variable. Extending SBL to the MMV framework one hyperparameter controls all the time samples for each variable (Wipf and Rao, 2007).
₃₃₀ As more samples are available in determining whether a variable is relevant in the MMV model an improved solution is obtained (Wipf and Rao, 2007; Zhang and Rao, 2011). However, the assumption of common sparsity profile across time does not always hold physiologically.

More flexible ways of handling and exploiting temporal coherency have been proposed
₃₃₅ (Montoya-Martinez et al., 2012; Gramfort et al., 2013), wherein signals are modeled as being non-stationary. Structured sparsity profiles are achieved by Gramfort et al. by time-frequency analysis and modeling each active source as a summation of Gabor atoms

18

(Gramfort et al., 2013) . Montoya-Martines et al. avoid synthesizing a dictionary containing temporal patterns by using the sparse group LASSO regularizer (Montoya-Martinez

₃₄₀ et al., 2012). While these two methods both base their solution on regularization through the L_1/L_2-norm, we propose to obtain sparser solutions through the VG (Kappen, 2011) which approximates the L_0-norm regularizer. Furthermore the VG has the favorable trait of estimating both the *state of activation* (active/non-active) and the *activation strength* of the active sources. This allows for modeling the temporal dynamics in EEG as having

₃₄₅ smooth temporal support, while allowing for more rapid changes in the dipole strength, under the assumption that the location of activation varies slower than the activation strength. By applying a Markovian prior on the support, the level of smoothness in the temporal sparsity profile is adapted to the observed EEG data.

We extended MarkoVG by incorporating spatial basis functions inspired by earlier
₃₅₀ implementations (Friston et al., 2008). Spatial smooth compact source patches are motivated by knowledge of the EEG generators' spatial extension, which is estimated to be at least 5×5 mm^2 (Baillet et al., 2001) and often extending several centimeters (Michel, 2009). Incorporating spatial basis functions has the additional benefit of reducing the computational complexity when there are fewer basis functions than original sources.
₃₅₅ There is however a risk that the center of a "true" source is a source with low activity in the basis functions. This is a potential bias that is incurred to counter the uncertainty and ill-posedness of the EEG inverse problem. These assumptions are argued to impose limited bias as in (Friston et al., 2008). Another related possible concern is the use of fixed orientations and simplified forward models in general. Several studies argue
₃₆₀ for improving source reconstruction by using as detailed and accurate forward models as possible (Akalin Acar and Makeig, 2013; Windhoff et al., 2013), and we note that it is indeed possible to combine MarkoVG with any type of forward model. Incorporation of both flexible dipole orientations and spatial coherency could, for example, be achieved by the so-called 'sparse basis fields' introduced by Haufe et al. (Haufe et al., 2011). Finally,
₃₆₅ increasing the number of spatial basis functions should also be investigated in future work.

In summary we applied the physiological meaningful assumptions that the underlying EEG generators are spatially smooth and sparse, and temporally variable smooth/sparse. The effectiveness of implemented assumptions can in general be validated through simu-

lation studies. We therefore tested the performance of MarkoVG in a controlled setting

370 where we also compared the proposed algorithm to three other solvers. On the synthetic data we found that MarkoVG was better at identifying the correct active sources and time samples. MarkoVG was thus more effective in recovering the sparsity level, both spatially and temporally. Importantly, we showed that even when favoring the T-MSBL and especially the M-FOCUSS algorithm superior performance to MarkoVG was not

375 accomplished.

In real data hypothesized assumptions can be validated using other imaging techniques such as fMRI or through lesion studies which can provide information about the location of specific information processing in the brain. The estimated temporal dynamics can be validated through single cell recordings that can indicate when specific brain areas

380 are involved in an EEG response. We tested our algorithm on the EEG response to seeing faces as compared to scrambled faces. In this paradigm we know from fMRI studies (Henson et al., 2003) and combined EEG/MEG studies (Henson et al., 2009) that the activated areas include the FFA and OFA. Studies of patients with lesions in the FFA and OFA further validate the importance of these areas in face perception (Eimer

385 and McCarthy, 1999; Dalrymple et al., 2011). The face/scrambled face contrasted fMRI recordings included in the multimodal study we extracted EEG from, have been analyzed at the group level in Fig. 3b in (Wakeman and Henson, 2015). The O/FFAs were also here dominating the face response as compared to the scrambled face condition. Furthermore, the face-sensitive response was sparse and largely symmetric across hemispheres.

390 The MSP, T-MSBL, M-FOCUSS and MarkoVG confirmed the existence of face-sensitive activity in the FFA and OFA. Frontal activation was also recovered by all algorithms, partly agreeing with the before mentioned fMRI study which also contained frontal activation. The temporal dynamics recovered by the inverse solvers in the most strongly activated sources showed focused activity around the N170 component. This

395 was especially true for MarkoVG which for the shown sources only had activity 130 to 200 ms after stimuli onset. The strongest sources as estimated by MSP, T-MSBL and M-FOCUSS had activity in the entire time window including peaks around 50 ms (the latter mostly pronounced for MSP). A study of the single cell recordings from the inferior temporal cortex of the macaque brain response to faces revealed predictive power in the

400 response after approximately 100 ms (Kiani et al., 2005), thus indicating the relevant face response starts well after 50 ms. Our study thus indicates that this evidence can be transferred to humans. We also note that an ERP study on humans has shown that significant differences between faces and noise textures begins 130 ms after stimuli onset (Rousselet et al., 2008).

405 The reasoning behind promoting zero activation is a model of focal brain activation, i.e., specific brain areas become active as a response to given stimuli and then return to their baseline level. This is for example appropriate when reconstructing contrast EEG responses. Furthermore the ill-posedness of the inverse problem and the poor signal-to-noise levels of EEG in general obstruct accurate recovery of dense source activations. By 410 sparsity promoting priors we focus on activity in time samples having sufficient evidence.

For reference we provide the computation time required by each inverse solver per iteration. Computed on a laptop with 2.1-GHz 64-bit i7 processor the time spend per iteration is 225 ms for MSP, 3 ms for T-MSBL, 0.6 ms for M-FOCUSS and 100 ms for MarkoVG. We note that MSP in general requires least iterations per inverse problem, 415 and since MarkoVG performs cross-validation to estimate the sparsity level it is slowest among the tested algorithms. However, active set based optimization as implemented in TMSBL and M-FOCUSS (in which inactive variables are pruned), could dramatically reduce the computation time. This is a current topic of research. Furthermore, we note that the goal of MarkoVG is to explore the implementation of meaningful physiologically 420 priors in order to solve the severely ill-posed inverse EEG problem rather than being fast.

In conclusion we have introduced temporal smoothness in the support of the brain dynamics within the so-called MarkoVG framework, and demonstrated how it can adapt to the degree of temporal coherency and spatial sparsity underlying the recorded EEG signal. In simulations and in real data MarkoVG showed promise as a tool for EEG dynamic 425 imaging. Further improvements on the algorithm involves increasing the flexibility of the model by optimizing two free parameters instead of working with a fixed relation between them as here. Such more complex optimization of parameters could be accomplished by Bayesian optimization methods such as proposed by (Snoek et al., 2012).

Acknowledgements

References

Akalin Acar, Z., Acar, C.E., Makeig, S., 2016. Simultaneous head tissue conductivity and EEG source location estimation. NeuroImage 124, 168–180.

440 Akalin Acar, Z., Makeig, S., 2013. Effects of forward model errors on EEG source localization. Brain topography 26, 378–396.

Andersen, M.R., Winther, O., Hansen, L.K., 2014. Bayesian inference for structured spike and slab priors, in: Advances in Neural Information Processing Systems, pp. 1745–1753.

445 Ashburner, J., Barnes, G., Chen, C.C., Daunizeau, J., Flandin, G., Friston, K., Kiebel, S., Kilner, J., Litvak, V., Moran, R., et al., 2014. SPM12 Manual .

Baillet, S., Mosher, J.C., Leahy, R.M., 2001. Electromagnetic brain mapping. Signal Processing Magazine, IEEE 18, 14–30.

Cotter, S.F., Rao, B.D., Engan, K., Kreutz-delgado, K., Member, S., 2005. Sparse Solu-
450 tions to Linear Inverse Problems With Multiple Measurement Vectors. IEEE Transactions on Signal Processing 53, 2477–2488.

Dalrymple, K.A., Oruc, I., Duchaine, B., Pancaroglu, R., Fox, C.J., Iaria, G., Handy, T.C., Barton, J.J., 2011. The anatomic basis of the right face-selective N170 in acquired prosopagnosia: a combined ERP/fMRI study. Neuropsychologia 49, 2553–2563.

₄₅₅ De Ciantis, A., Lemieux, L., 2013. Localisation of epileptic foci using novel imaging modalities. Current opinion in neurology 26, 368.

Delorme, A., Palmer, J., Onton, J., Oostenveld, R., Makeig, S., 2012. Independent EEG sources are dipolar. PloS one 7, e30135.

Donoho, D.L., Elad, M., Temlyakov, V.N., 2006. Stable recovery of sparse overcomplete ₄₆₀ representations in the presence of noise. Transactions on Information Theory, IEEE 52, 6–18.

Eimer, M., McCarthy, R.A., 1999. Prosopagnosia and structural encoding of faces: Evidence from event-related potentials. Neuroreport 10, 255–259.

Friston, K., Harrison, L., Daunizeau, J., Kiebel, S., Phillips, C., Trujillo-Barreto, N., ₄₆₅ Henson, R., Flandin, G., Mattout, J., 2008. Multiple sparse priors for the M/EEG inverse problem. NeuroImage 39, 1104–1120.

Gorodnitsky, I., George, J., Rao, B., 1995. Neuromagnetic source imaging with FOCUSS: a recursive weighted minimum norm algorithm. Electroencephalography and clinical Neurophysiology 95, 231–251.

₄₇₀ Gorodnitsky, I.F., Rao, B.D., 1997. Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm. Signal Processing, IEEE Transactions on 45, 600–616.

Gramfort, A., Kowalski, M., Hämäläinen, M., 2012. Mixed-norm estimates for the M/EEG inverse problem using accelerated gradient methods. Physics in medicine and ₄₇₅ biology 57, 1937–1961.

Gramfort, A., Strohmeier, D., Haueisen, J., Hämäläinen, M.S., Kowalski, M., 2013. Time-frequency mixed-norm estimates: Sparse M/EEG imaging with non-stationary source activations. NeuroImage 70, 410–422.

Hämäläinen, M.S., Ilmoniemi, R., 1994. Interpreting magnetic fields of the brain: mini-₄₈₀ mum norm estimates. Medical & biological engineering & computing 32, 35–42.

Hansen, L.K., Hansen, S.T., Stahlhut, C., 2013a. Mobile real-time EEG imaging Bayesian inference with sparse, temporally smooth source priors, in: 2013 International Winter Workshop on Brain-Computer Interface (BCI), IEEE. pp. 6–7.

Hansen, L.K., Rasmussen, C.E., 1994. Pruning from adaptive regularization. Neural Computation 6, 1223–1232.

Hansen, S.T., Hansen, L.K., 2013. EEG Sequence Imaging: A Markov Prior for the Variational Garrote, in: Unpublished note for the International Workshop, Machine Learning and Interpretation in Neuroimaging, available at http://orbit.dtu.dk/fedora/objects/orbit:127330/datastreams/file_61f34d92-2e60-4871-8a41-08f1d69f5c47/content.

Hansen, S.T., Hansen, L.K., 2014. EEG source reconstruction using sparse basis function representations, in: 2014 International Workshop on Pattern Recognition in Neuroimaging, IEEE. pp. 1–4.

Hansen, S.T., Hauberg, S., Hansen, L.K., 2016. Data-driven forward model inference for eeg brain imaging. NeuroImage 139, 249–258.

Hansen, S.T., Stahlhut, C., Hansen, L.K., 2013b. Expansion of the Variational Garrote to a Multiple Measurement Vectors Model, in: Scandinavian Conference on AI 2013 (2013), SCAI.

Hansen, S.T., Stahlhut, C., Hansen, L.K., 2013c. Sparse Source EEG Imaging with the Variational Garrote, in: Pattern Recognition in Neuroimaging (PRNI), 2013 International Workshop on, IEEE. pp. 106–109.

Haufe, S., Nikulin, V.V., Ziehe, A., Müller, K.R., Nolte, G., 2008. Combining sparsity and rotational invariance in EEG/MEG source reconstruction. NeuroImage 42, 726–38.

Haufe, S., Tomioka, R., Dickhaus, T., Sannelli, C., Blankertz, B., Nolte, G., Müller, K.R., 2011. Large-scale EEG/MEG source localization with spatial flexibility. NeuroImage 54, 851–859.

Henson, R., Goshen-Gottstein, Y., Ganel, T., Otten, L., Quayle, A., Rugg, M., 2003.

24

Electrophysiological and haemodynamic correlates of face perception, recognition and priming. Cerebral cortex 13, 793–805.

510 Henson, R.N., Flandin, G., Friston, K.J., Mattout, J., 2010. A parametric empirical Bayesian framework for fMRI-constrained MEG/EEG source reconstruction. Human brain mapping 31, 1512–31.

Henson, R.N., Mouchlianitis, E., Friston, K.J., 2009. MEG and EEG data fusion: simultaneous localisation of face-evoked responses. NeuroImage 47, 581–9.

515 Hulbert, S., Adeli, H., 2013. EEG/MEG-and imaging-based diagnosis of Alzheimers disease. Reviews in the neurosciences 24, 563–576.

Ishwaran, H., Rao, J.S., 2005. Spike and slab variable selection: frequentist and Bayesian strategies. Annals of Statistics , 730–773.

Itier, R.J., Taylor, M.J., 2004. N170 or N1? Spatiotemporal differences between object 520 and face processing using ERPs. Cerebral cortex 14, 132–142.

Kappen, H., 2011. The Variational Garrote. arXiv preprint arXiv:1109.0486 arXiv:1109.0486v2.

Kappen, H.J., Gómez, V., 2013. The Variational Garrote. Machine Learning , 1–26.

Kiani, R., Esteky, H., Tanaka, K., 2005. Differences in onset latency of macaque inferotem-525 poral neural responses to primate and non-primate faces. Journal of neurophysiology 94, 1587–1596.

Litvak, V., 2016. JISCMail - SPM Archives.

Liu, H., Gao, X., Schimpf, P.H., Yang, F., Gao, S., 2004. A recursive algorithm for the three-dimensional imaging of brain electric activity: shrinking LORETA-FOCUSS. 530 Biomedical Engineering, IEEE Transactions on 51, 1794–1802.

Lucka, F., Pursiainen, S., Burger, M., Wolters, C.H., 2012. Hierarchical Bayesian inference for the EEG inverse problem using realistic FE head models: depth localization and source separation for focal primary currents. NeuroImage 61, 1364–82.

MacKay, D.J.C., 1995. Probable networks and plausible predictions a review of practical
Bayesian methods for supervised neural networks. Network: Computation in Neural
Systems 6, 469–505.

Makhoul, J., Kubala, F., Schwartz, R., Weischedel, R., 1999. Performance measures for
information extraction, in: Proceedings of DARPA Broadcast News Workshop. Morgan
Kaufmann Pub, pp. 249–252.

Markov, N., Misery, P., Falchier, A., Lamy, C., Vezoli, J., Quilodran, R., Gariel, M.,
Giroud, P., Ercsey-Ravasz, M., Pilaz, L., et al., 2011. Weight consistency specifies
regularities of macaque cortical networks. Cerebral Cortex 21, 1254–1272.

Matsuura, K., Okabe, Y., 1995. Selective minimum-norm solution of the biomagnetic
inverse problem. IEEE transactions on bio-medical engineering 42, 608–15.

Michel, C.M., 2009. Electrical neuroimaging. Cambridge University Press.

Montoya-Martinez, J., Artes-Rodriguez, A., Hansen, L.K., Pontil, M., 2012. Structured
sparsity regularization approach to the EEG inverse problem, in: Cognitive Information
Processing (CIP), 2012 3rd International Workshop on, IEEE. pp. 1–6.

Nunez, P.L., Srinivasan, R., Westdorp, A.F., Wijesinghe, R.S., Tucker, D.M., Silberstein,
R.B., Cadusch, P.J., 1997. EEG coherency: I: statistics, reference electrode, volume
conduction, Laplacians, cortical imaging, and interpretation at multiple scales. Elec-
troencephalography and clinical neurophysiology 103, 499–515.

Ou, W., Hämäläinen, M.S., Golland, P., 2009. A distributed spatio-temporal EEG/MEG
inverse solver. NeuroImage 44, 932–46.

Pascual-Marqui, R., Esslen, M., Kochi, K., Lehmann, D., et al., 2002. Functional imaging
with low-resolution brain electromagnetic tomography (LORETA): a review. Methods
and findings in experimental and clinical pharmacology 24, 91–95.

Pascual-Marqui, R.D., Michel, C.M., Lehmann, D., 1994. Low resolution electromagnetic
tomography: a new method for localizing electrical activity in the brain. International
Journal of psychophysiology 18, 49–65.

Phillips, C., 2000. Source estimation in EEG. PhD thesis.

Phillips, C., Rugg, M.D., Friston, K.J., 2002. Anatomically Informed Basis Functions for EEG Source Localization: Combining Functional and Anatomical Constraints. NeuroImage 16, 678–695.

Rijsbergen, C.J.V., 1979. Information Retrieval. Butterworth-Heinemann. 2nd edition.

Rousselet, G.A., Husk, J.S., Bennett, P.J., Sekuler, A.B., 2008. Time course and robustness of ERP object and face differences. Journal of vision 8, 3.1–18.

Scherg, M., Von Cramon, D., 1985. Two bilateral sources of the late AEP as identified by a spatio-temporal dipole model. Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section 62, 32–44.

Schüz, A., Braitenberg, V., 2002. The human cortical white matter: Quantitative aspects of cortico-cortical long-range connectivity. Cortical Areas: Unity and Diversity , 377–384.

Snoek, J., Larochelle, H., Adams, R.P., 2012. Practical Bayesian optimization of machine learning algorithms, in: Advances in Neural Information Processing Systems, pp. 2951–2959.

Stahlhut, C., Mørup, M., Winther, O., Hansen, L.K., 2011. Simultaneous EEG source and forward model reconstruction (sofomore) using a hierarchical bayesian approach. Journal of Signal Processing Systems 65, 431–444.

Tipping, M.E., 2001. Sparse Bayesian learning and the relevance vector machine. The Journal of Machine Learning Research 1, 211–244.

Vega-Hernández, M., Martínez-Montes, E., Sanchez-Bornot, J.M., Lage-Castellanos, A., Valdés-Sosa, P.A., 2008. Penalized least squares methods for solving the EEG inverse problem. Statistica Sinica 18, 1535–1551.

Wakeman, D.G., Henson, R.N., 2015. A multi-subject, multi-modal human neuroimaging dataset. Scientific Data 2, 150001.

Windhoff, M., Opitz, A., Thielscher, A., 2013. Electric field calculations in brain stimulation based on finite elements: an optimized processing pipeline for the generation and usage of accurate individual head models. Human brain mapping 34, 923–935.

27

590   Wipf, D.P., Rao, B.D., 2007. An empirical Bayesian strategy for solving the simultaneous sparse approximation problem. Signal Processing, IEEE Transactions on, 55, 3704–3716.

Zhang, Z., Rao, B.D., 2011. Sparse signal recovery with temporally correlated source vectors using sparse Bayesian learning. Selected Topics in Signal Processing, IEEE
595   Journal of 5, 912–926.