Large-Scale EEG/MEG Source Localization with Spatial Flexibility

Stefan Haufe [a,b,*], Ryota Tomioka [c], Thorsten Dickhaus [a], Claudia Sannelli [a], Benjamin Blankertz [a,d,b], Guido Nolte [d] and Klaus-Robert Müller [a,b]

[a] *Machine Learning Group, Department of Computer Science, Berlin Institute of Technology, Franklinstr. 28/29, D-10587 Berlin, Germany*
[b] *Bernstein Focus Neurotechnology, Berlin, Germany*
[c] *Information-Theoretic Machine Learning and Data Mining Group, Department of Mathematical Informatics, Graduate School of Information Science and Technology, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan.*
[d] *Intelligent Data Analysis Group, Fraunhofer Institute FIRST, Kekuléstr. 7, D-12489 Berlin, Germany*

Abstract

We propose a novel approach to solving the electro- / magnetoencephalographic (EEG / MEG) inverse problem which is based upon a decomposition of the current density into a small number of spatial basis fields. It is designed to recover multiple sources of possibly different extent and depth, while being invariant with respect to phase angles and rotations of the coordinate system. We demonstrate the method's ability to reconstruct simulated sources of random shape and show that the accuracy of the recovered sources can be increased, when interrelated field patterns are co-localized. Technically, this leads to large-scale mathematical problems, which are solved using recent advances in convex optimization. We apply our method for localizing brain areas involved in different types of motor imagery using real data from Brain-Computer Interface (BCI) sessions. Our approach based on single-trial localization of complex Fourier coefficients yields class-specific focal sources in the sensorimotor cortices.

Key words: EEG, MEG, inverse problem, basis field, large-scale optimization, motor imagery, brain-computer interfaces

1. Introduction

Measuring electrical field distributions allows to localize cognitive processing and is thus of high value for neuroscience research and medical diagnosis. While invasive measurements provide a very local assessment of neuronal activations, such a procedure is only possible in humans where electrodes are already implanted for treatment/diagnosis of neurological diseases, e.g., epilepsy. Noninvasive localization techniques based on electro- and magnetoencephalography (EEG and MEG) are applicable without restriction and are therefore highly useful. They have become standard tools for analyzing fast brain signals such as somatosensory-evoked potentials (SEPs) or ongoing oscillations. For understanding the respective cognitive processes spatial patterns (scalp maps) derived from EEG/MEG, however, only give a rough estimate of the true underlying sources. Thus, for revealing a more detailed picture a full source reconstruction is required which involves a mathematical inversion of the (approximately) known mapping from sources to sensors. Unfortunately, this is an ill-defined inverse problem since any measurement can be equally well explained by infinitely many different source distributions.

Therefore, in order to "solve" the inverse problem it is necessary to impose additional constraints on the solution. Dipole fits (e.g. Scherg and von Cramon, 1986) and scanning techniques (Schmidt, 1986; Mosher and Leahy, 1999; Veen and Buckley, 1988; Van Veen et al., 1997) correspond to directly constraining the number of dipolar sources. Imaging methods, in contrast, model a large number of dipoles and thus allow us to estimate activity in the entire brain at once. Constraints are here imposed by a dedicated penalty functional reflecting assumptions on the sources. Perhaps the two most common assumptions are smoothness (Hämäläinen and Ilmoniemi, 1994; Pascual-Marqui et al., 1994; Pascual-Marqui, 2002) and focality (Matsuura and Okabe, 1995; Gorodnitsky et al., 1995; Uutela et al., 1999; Huang et al., 2006; Ou et al., 2008; Ding and He, 2008; Bolstad et al., 2009), both of which can be motivated by neurophysiological arguments. Nevertheless both approaches may deliver implausible results in practice since "smooth methods" tend to estimate sources that spread over a considerable part of the brain which is not always

* Corresponding author.
Email address: stefan.haufe@tu-berlin.de (Stefan Haufe).

physiologically meaningful. Estimates obtained by "sparse methods" tend to be unstable and scattered around the true sources. Two recent studies suggest that estimates with more plausible extent and shape can be obtained by encouraging both smoothness and focality of the sources (Haufe et al., 2008; Vega-Hernández et al., 2008) using two penalties. Such a hybrid approach has been shown to outperform purely smooth respectively focal methods when distinguishing two or three simulated as well as real sources (Haufe et al., 2008).

In this paper we propose a novel method for EEG/MEG source reconstruction that achieves a compromise between smoothness and focality, allowing to model extended sources. This is achieved by expanding the current density into a sparse combination of spatial basis fields. Compared with our previous approach Focal Vectorfield Reconstruction (FVR, Haufe et al., 2008), the presented method achieves similarly good localization, while allowing a simpler mathematical formulation. The novel cost function enables the deployment of a very efficient optimization scheme by which it becomes possible to solve reconstruction problems involving orders of magnitude more variables than previously. These additional variables can be used to localize larger datasets, or to increase the spatial resolution.

We will derive the proposed methodology in Section 2. Section 2 describes the application of our method to two datasets. The first one is simulated and used for assessing the source reconstruction quality of our method in comparison to FVR and other approaches. The second dataset consists of EEG responses acquired during sucessful brain-computer interface (BCI) sessions, where the task was to modulate local μ-rhythms by means of motor imagery of different limbs. These data are an ideal testbed for source reconstruction algorithms, since there exists a strong prior knowledge about the neurophysiological basis underlying a good BCI performance. Results on the physical origin of class-related EEG activity during BCI sessions are presented in Section 3. Section 4 contains a general discussion of strengths and potential drawbacks of current distributed inverse methods and their inherent assumptions, as well as practical issues regarding regularization. The paper finishes with concluding remarks in Section 5.

2. Methods and Materials

2.1. *Localization using Sparse Basis Field Expansions*

In EEG/MEG source reconstruction we are equipped with measurements of the scalp electrical potential (EEG) or magnetic field (MEG), from which we would like to infer the generating electrical current density (sources) in the brain. The EEG/MEG activity is comprised in a vector $\mathbf{z} = (z_1, \ldots, z_M)^\top \in \mathbb{C}^M$, where M is the number of sensors. As the data $\mathbf{z}$ could possibly contain responses to Fourier or wavelet filters it is allowed to take complex values. Let $\mathbb{B} \subset \mathbb{R}^3$ be the volume covered by the brain (i.e. white and gray matter). The current density is a vector field $\mathbf{y} : \mathbb{B} \rightarrow \mathbb{C}^3$ assigning a (complex) vectorial current source to each location in the brain. Considering a discrete sample of locations (voxels) and source currents $(\mathbf{x}_n, \mathbf{y}(\mathbf{x}_n) =: \mathbf{y}_n), n = 1, \ldots, N$, we denote by $Y = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_N^\top)^\top$ the $N \times 3$ matrix of sources and by $\mathbf{vec}\,(Y)$ a column vector containing the stacked transposed rows of Y. The forward mapping from the sources Y to the measurements $\mathbf{z}$ is linear and can be written as

$$\mathbf{z} = F\,\mathbf{vec}\,(Y) \tag{1}$$

using the lead field matrix $F \in \mathbb{R}^{M \times 3N}$, which can be computed for a known geometry of the head and known conductive properties of brain, skull and skin tissues (Baillet et al., 2001).

2.1.1. *Model*

Instead of estimating the currents $\mathbf{y}_n$ directly, we propose to model the current density as a linear combination of (potentially many) spatial *basis fields*, the coefficients of which are to be estimated. A basis field is defined here as a vector field, in which all output vectors point in the same direction, while the magnitudes are proportional to a scalar (basis) function $b : \mathbb{B} \rightarrow \mathbb{R}$. Given a set of functions $b_l, l = 1, \ldots, L$ (called a "dictionary"), the basis field expansion reads

$$\mathbf{y}(\mathbf{x}) = \sum_{l=1}^{L} \mathbf{c}_l b_l(\mathbf{x}) \;, \tag{2}$$

with coefficient vectors $\mathbf{c}_l \in \mathbb{C}^3, l = 1, \ldots, L$. By including one complex coefficient for each dimension, we learn orientations and amplitudes as well as phases of the complex current vectors in this model. Let $C = (\mathbf{c}_1, \ldots, \mathbf{c}_L)^\top \in \mathbb{C}^{L \times 3}$ contain the coefficients and

$$B = \begin{pmatrix} b_1(\mathbf{x}_1) & \ldots & b_L(\mathbf{x}_1) \\ \vdots & \ddots & \vdots \\ b_1(\mathbf{x}_N) & \ldots & b_L(\mathbf{x}_N) \end{pmatrix} \in \mathbb{R}^{N \times L} \tag{3}$$

be the basis functions evaluated at all locations $\mathbf{x}_n$. The forward model then reads

$$\mathbf{z} = F\,\mathbf{vec}\,(BC) \;. \tag{4}$$

2.1.2. *Sparsity, Rotational Invariance and Phase Invariance*

Solving Eq. (4) for C does not yield a unique solution if the number of coefficients is larger than the number of electrodes M, which is the common situation. The ambiguity can be overcome by *regularization*, i.e., by imposing additional constraints on the variables. Here, we assume that, for an appropriately chosen dictionary, the current density can be well approximated by a *small* number of basis fields. This can be achieved by estimating a *sparse* coefficient matrix C, i.e., a matrix that has mostly zero entries. Besides the regularizing effect, sparse decompositions also provide a way of interpreting current densities by looking at the selected basis functions (those having corresponding nonzero

coefficients in C). The premise for such interpretability is that the basis functions itself are simple enough, which should be ensured when designing the dictionary.

An important property of EEG/MEG source reconstruction algorithms is rotational invariance. That is, the estimated current density should not change when the coordinate system is rotated. This holds in general for ℓ_2-norm a.k.a. Tikhonov regularized methods, which deliver non-sparse sources/coefficients. However, if sparsity is desired, additional effort is needed. For example, penalizing the ℓ_1-norm (the sum of absolute values of the entries) of C leads to a sparse expansion, but not to rotational invariance. The ℓ_1-norm penalty does not couple the three dimensions of the current density, making it very probable that different coefficients are set to zero for each of them. This amounts to selecting different basis functions in each dimension. As a result, the tendency of ℓ_1-norm regularized methods to favor zero coefficients also creates a bias towards current orientations that are perpendicular to one or more of the axes of the coordinate system and are physiologically meaningless.

It has recently been pointed out, that rotational invariance of vectorial quantities can be maintained by choosing a so-called $\ell_{1,2}$-norm penalty, which minimizes the (sparsity inducing) ℓ_1-norm of vector *amplitudes* (Ding and He, 2008; Haufe et al., 2008; Ou et al., 2008; Bolstad et al., 2009). The difference between "standard" ℓ_1-norm and the $\ell_{1,2}$-norm is that the former leads to entry-wise sparsity, while the latter sets whole rows of C jointly to zero. Importantly the chosen coordinate system does not influence whether or not a row is set to zero by the $\ell_{1,2}$-norm, while it does affect the pruning of entries by the ℓ_1-norm. For a geometrical explanation of why ℓ_1- and $\ell_{1,2}$-norm penalties lead to sparsity at all we refer to Tibshirani (1996) and Yuan and Lin (2006).

The $\ell_{1,2}$-norm regularizer is defined by

$$\mathcal{R}(C) = \|C\|_{1,2} = \sum_{l=1}^{L} \|\mathbf{c}_l\|_2 \ . \tag{5}$$

Technically, $\mathcal{R}(C)$ is rotationally invariant due to the use of the ℓ_2-norm

$$\|\mathbf{c}\|_2 = \sqrt{\sum_{d=1}^{3} c_d^2} \tag{6}$$

in output space, which does not change under rotation. Let $Q \in \mathbb{C}^{3\times 3}, Q^\dagger Q = I$ be a unitary matrix, where $Q^\dagger$ is the adjoint of Q. Now

$$\sum_{l=1}^{L} \|Q\mathbf{c}_l\|_2 = \sum_{l=1}^{L} \sqrt{\mathrm{tr}(\mathbf{c}_l^\dagger Q^\dagger Q \mathbf{c}_l)} = \sum_{l=1}^{L} \|\mathbf{c}_l\|_2 \ . \tag{7}$$

Note that the class of unitary matrices covers both rotations $Q \in \mathbb{R}^{3\times 3}, Q^\top Q = I$ as well as phase shifts $Q = \exp(i\phi) I_{(3\times 3)}$ as special cases.

2.1.3. *Dictionary*

The idea of enforcing smoothness *and* focality in the inverse solution is to avoid the scattering of activity found for many purely focal approaches, while at the same time to maintain their high spatial resolution and the associated ability to distinguish multiple sources. In other words, we are looking for source estimates with spatially constricted but smooth active regions. In Haufe et al. (2008) a combination of two penalties was used to achieve that effect. Here, it is addressed by designing an appropriate basis function dictionary. We consider an expansion into Gaussians. These are smooth, but also well localized due to exponentially decaying tails. Thanks to the latter, sparse combinations of Gaussian bases give rise to good spatial separation of sources. Using a redundant dictionary containing Gaussians of different scales, we further expect that sources with arbitrary shape can be reconstructed with few basis elements.

Formally, we consider spherical Gaussians

$$b_{n,s}(\mathbf{x}) = \left(\sqrt{2\pi}\sigma_s\right)^{-3} \exp\left(-\frac{1}{2}\|\mathbf{x} - \mathbf{x}_n\|_2^2 \, \sigma_s^{-2}\right) \tag{8}$$

being centered at nodes $\mathbf{x}_n, n = 1, \ldots, N$ and having S different spatial standard deviations $\sigma_s, s = 1, \ldots, S$ (see Fig. 1 for examples).

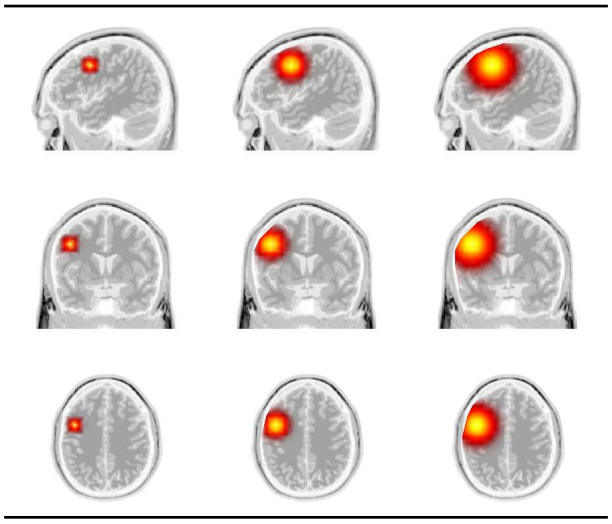

Fig. 1. Examples of Gaussian basis functions $b_{n,s}(\mathbf{x})$ with spatial standard deviations $\sigma_1 = 0.5$ cm, $\sigma_3 = 1$ cm and $\sigma_5 = 1.5$ cm.

2.1.4. *Normalization*

The proposed $\ell_{1,2}$-norm based regularization aims at selecting the smallest possible number of basis fields necessary to explain the measurement. This approach, however, is heuristic, since not the number of nonzero coefficient vectors, but their magnitudes enter the cost function. It is therefore important to normalize the basis functions in order not to prefer some of them a-priori. Let B_s be the $N \times N$ matrix containing all basis function evaluations with standard deviation σ_s. The large matrix

$$B = \left(\frac{B_1}{\| \mathbf{vec}\,(B_1) \|_1}, \ldots, \frac{B_S}{\| \mathbf{vec}\,(B_S) \|_1} \right) \in \mathbb{R}^{N \times SN} \quad (9)$$

is constructed using normalized B_s. By this means, no length scale is preferred a-priori.

An estimation bias is also introduced by the location of the sources. Due to volume conduction, the signal captured by the sensors is much stronger for superficial sources compared to deep sources. In Pascual-Marqui (2002) the variance estimate $\hat{S} = \bar{F}^\top \left(\bar{F}\bar{F}^\top\right)^{-1} \bar{F} \in \mathbb{R}^{3N \times 3N}$ is derived for the (least-squares) estimated sources, where $\bar{F} = HF$ and $H = I_{(M \times M)} - \mathbf{1}_{(M)}\mathbf{1}_{(M)}^\top / M$ is the common-average reference transform. We found that $\hat{S}$ can be used for alleviating the location bias (Haufe et al., 2008). This can be done by penalizing activity at locations with high variance. Let $W_n \in \mathbb{R}^{3 \times 3}$ denote the inverse of the matrix square root of the n-th 3×3 blockdiagonal part of $\hat{S}$, we define the depth-compensation matrix

$$W = \begin{pmatrix} W_1 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & W_N \end{pmatrix} \in \mathbb{R}^{3N \times 3N} . \quad (10)$$

2.1.5. *Estimation*

Using the definitions from above the coefficients are sought which provide a defined compromise between sparsity and model error, i.e.

$$\hat{C} = \arg\min_C \; \mathcal{R}(C) + \lambda \mathcal{L}(C) \quad (11)$$

where $\mathcal{L}(C) = \|\mathbf{z} - \Gamma\,\mathbf{vec}\,(C)\|_2^2$ is the quadratic loss function, $\Gamma \equiv FW(B \otimes I_{(3\times 3)}) \in \mathbb{R}^{M \times 3SN}$ and λ is a positive constant controlling the tradeoff between loss function and regularization. Minimizing the weighted sum of two objectives is a measure to achieve a compromise between the two (cf. Zou and Hastie, 2005; Haufe et al., 2008; Vega-Hernández et al., 2008).

Given the coefficients the estimated current density at node $\mathbf{x}_n$ is defined by

$$\hat{\mathbf{y}}_n = W_n \sum_{l=1}^{SN} \hat{\mathbf{c}}_l b_l(\mathbf{x}_n) . \quad (12)$$

This solution has been termed sparse basis field expansion (S-FLEX) solution in a precursory conference paper (Haufe et al., 2009).

2.1.6. *Comparison to Focal Vectorfield Reconstruction*

Note that Eq. (11) has a structural similarity to our previous approach FVR (Haufe et al., 2008). The FVR solution is obtained by setting $B = I_{(N \times N)}$ (i.e., the coefficients $\mathbf{c}_i$ are equal to the sources $\mathbf{s}_i$) and adding the additional regularizer $\alpha \sum_{n=1}^N \|\mathbf{t}_n\|_2$, where $T = (\mathbf{t}_1, \ldots, \mathbf{t}_N)^\top = DW^{-1}C$ and D is a discrete spatial second derivative (Laplacian) operator. The additional term in FVR effectively enforces spatial smoothness or continuity of the current density by rewarding sparse second derivatives. Hence, FVR and our current approach achieve a very similar effect using contrary strategies (namely, sparsity before and after linear transformation). However, it is not possible to transform one problem into the form of the other, since B and D are generally not invertible. As we will see later (Section 2.1.8), this prevents our hereby proposed optimization algorithm to be applied to FVR, making generalization of FVR to large-scale scenarios harder.

2.1.7. *Extension to Multiple Measurements*

While Eq. (12) considers only single field patterns, we would now like to extend S-FLEX to the localization of multiple measurements. The goal is to estimate T current densities $\mathbf{y}_n(t)$ based on T patterns $\mathbf{z}(t)$. Let $Z = (\mathbf{z}(1), \ldots, \mathbf{z}(T)) \in \mathbb{C}^{M \times T}$ and $\mathbf{c}_l(t) \in \mathbb{C}^3$ be the coefficient vector describing the contribution of the l-th basis field to the t-th pattern. Defining $\tilde{\mathbf{c}}_l = (\mathbf{c}_l(1)^\top, \ldots, \mathbf{c}_l(T)^\top)^\top \in \mathbb{R}^{3T}$ and

$$\tilde{C} = \begin{pmatrix} \mathbf{c}_1(1) & \ldots & \mathbf{c}_1(T) \\ \vdots & \ddots & \vdots \\ \mathbf{c}_L(1) & \ldots & \mathbf{c}_L(T) \end{pmatrix} \in \mathbb{R}^{3L \times T} , \quad (13)$$

we propose to estimate

$$\hat{\tilde{C}} = \arg\min_{\tilde{C}} \; \tilde{\mathcal{R}}(\tilde{C}) + \lambda \tilde{\mathcal{L}}(\tilde{C}) \quad (14)$$

with $\tilde{\mathcal{R}}(\tilde{C}) = \sum_{l=1}^L \|\tilde{\mathbf{c}}_l\|_2$ and $\tilde{\mathcal{L}}(\tilde{C}) = \left\| \mathbf{vec}\left(Z - \Gamma\tilde{C}\right) \right\|_2^2$, which is equivalent to Eq. (11) for $T = 1$. However, for $T > 1$ it is not equivalent to solving T problems of type Eq. (11) separately, as in our case the $3T$ coefficients belonging to a certain basis function are tied under a common ℓ_2-norm penalty and can only be pruned to zero at the same time. Thus, the selection of basis functions which contribute coherently to several patterns is facilitated, while at the same time orientations, amplitudes and phases of the corresponding fields are allowed to differ per pattern. Such joint (or co-) localization was already suggested in previous work. The idea originates from Polonsky and Zibulevsky (2004) and appears also in Malioutov et al. (2005), Wipf and Rao (2007), Ou et al. (2008) and Bolstad et al. (2009). Malioutov et al., Ou et al. and Bolstad et al. (2009) use the technique for spatio-temporal source localization, where the ℓ_2-norm penalty in temporal domain prevents from artificial jumps in the time course of the estimated sources. Both studies suggest that joint localization leads to a better noise suppression compared to the single-timepoint estimator. A similar effect has been reported in a pure regression setting, where joint regularization of Fourier coefficients lead to improved BCI classification rates (van Gerven et al., 2009).

2.1.8. *Optimization*

Eqs. (11) and (14) form convex problems, composed of a quadratic loss function and a convex nondifferentiable regularizer. These problems share similarities with the problems discussed in Polonsky and Zibulevsky (2004); Haufe et al. (2008); Malioutov et al. (2005); Ou et al. (2008); Ding and He (2008); Wipf and Nagarajan (2009) and Bolstad et al. (2009). In the majority of these papers, the cost function is

reformulated as an instance of second-order cone programming (SOCP) (Lobo et al., 1998). The proposed interior-point-based SOCP solvers are, however, only applicable to small- and medium-sized problems not exceeding several ten thousands of variables. For this reason, some authors perform a dimensionality reduction step in order to reduce the number of variables and/or observations (Malioutov et al., 2005; Ou et al., 2008).

Here, we make use of a more recent advance in numerical optimization that enables us to solve S-FLEX instances involving millions of model parameters and thousands of observations. The proposed algorithm is based on deriving the Fenchel dual of the optimization problem and applying the augmented Lagrangian technique. It has thus been termed Dual Augmented Lagrangian (DAL, see Tomioka and Sugiyama (2009)). Usage of augmented Lagrangians was also proposed by Polonsky and Zibulevsky (2004) who apply the technique to the primal (original) problem. However, it is shown in Tomioka and Sugiyama (2009) that a dual formulation is more efficient when the number of unknown variables is much larger than the number of observations, which is the typical scenario in distributed source modeling.

We use the reference implementation of DAL, which is provided as open source software (Tomioka, 2009). Note that DAL is not only suitable for computing S-FLEX solutions, but could also be applied to solve large instances of the problem arising in Polonsky and Zibulevsky (2004); Ou et al. (2008); Ding and He (2008); Bolstad et al. (2009). Unfortunately, sparsity of *linearly transformed* variables is not efficiently handled by DAL, preventing our previous approach FVR to benefit from DAL. For this reason, we believe, that our current formulation is more suitable for achieving spatial flexibility in large-scale source localization tasks.

2.2. *Simulations*

2.2.1. *Assessing Single-Measurement Localization Performance*

Validation of methods for inverse reconstruction is generally difficult due to the lack of a "ground truth". The measurements $\mathbf{z}$ do not provide such a truth, as the main goal here is not to find a functional representation for the EEG, but for the underlying current density $\mathbf{y}(\mathbf{x})$, which is unknown. Therefore, a standard way of evaluating inverse methods is to assess their ability to reconstruct known functions. This is done here by reconstructing simulated current sources, which are generated as follows. A realistic head model is obtained from high-resolution MRI (magnetic resonance imaging) slices of a human head (Holmes et al., 1998). Inside the brain, $N = 2142$ dipole locations $\mathbf{x}_n, n = 1, \ldots, N$ are defined according to a cubic grid of 10mm inter-dipole distance. Corresponding current vectors $\mathbf{y}_n$ are sampled from a multivariate standard normal distribution. The resulting function $(\mathbf{x}_n, \mathbf{y}_n)$ is spatially smoothed using a Gaussian lowpass filter with standard deviation 2.5 cm. Finally, denoting by p_k the k-th percentile of the current lengths $\|\mathbf{y}_n\|_2, n = 1, \ldots, N$, each $\mathbf{y}_n$ is scaled to have length $\max(\|\mathbf{y}_n\|_2 - p_{90}, 0)$, i.e., only the 10% largest currents are retained. Source distributions obtained by this procedure usually feature two-three activity patches (sources) with small to medium extent and smoothly varying magnitude and orientation (see Fig. 2 for an example). The lead field $F \in \mathbb{R}^{M \times 2142 \cdot 3}$ is constructed according to Nolte and Dassios (2005) taking into account the realistic head geometry.

The localization is carried out using the proposed sparse basis field expansion (S-FLEX) approach, the commonly used approaches of LORETA (Pascual-Marqui et al., 1994), minimum ℓ_1-norm estimate (denoted as L1 in the following) (Matsuura and Okabe, 1995), and our recently proposed Focal Vectorfield Reconstruction (FVR) technique (Haufe et al., 2008). Note that these methods cover the full spectrum from smooth spread-out solutions (LORETA) to sparse solutions (L1). We use a variant of L1, in which the original depth compensation approach is replaced by the approach outlined in Section 2.1.4. As the data was simulated without noise, perfect reconstruction is required for all methods. For S-FLEX, basis functions with three different standard deviations $\sigma_1 = 0.5$ cm, $\sigma_2 = 1$ cm, $\sigma_3 = 1.5$ cm are used. The tradeoff parameter α for FVR is chosen as suggested in Haufe et al. (2008).

Five current densities are simulated and respective pseudo EEG measurements for 118 channels are computed. For each measurement and method a 5×5-fold cross-validation is conducted. That is, the EEG electrodes are randomly partitioned into five groups of approximately equal size. Each union of four electrode groups gives rise to a "training set", while the remaining channel groups are called "test sets". The procedure is carried out five times with different randomizations, yielding 25 training sets with corresponding test sets. Inverse reconstructions are carried out based on the "training sets". In each of the 25 cross-validation runs, two criteria are evaluated. Most importantly the *reconstruction error*, defined as

$$\mathrm{REC} = \left\| \frac{\mathbf{vec}\,(Y)}{\|\,\mathbf{vec}\,(Y)\,\|_2} - \frac{\mathbf{vec}\left(\hat{Y}^{\mathrm{tr}}\right)}{\|\,\mathbf{vec}\left(\hat{Y}^{\mathrm{tr}}\right)\,\|_2} \right\|_2 , \qquad (15)$$

is considered, where $\hat{Y}^{\mathrm{tr}}$ are the vector field outputs at nodes $\mathbf{x}_n, n = 1, \ldots, N$ estimated using only the training set. Apart from the pointwise reconstruction, we also consider the earth-mover's distance (EMD) between true and estimated current density, which measures the effort needed to transform one density into the other. The EMD is described in Rubner et al. (2000) and has been introduced in the context of EEG/MEG inverse solution evaluation in Haufe et al. (2008).

A third quantity of interest is the *generalization error*, i.e., the error in predicting the activity at those channels in the test set from the sources that are estimated from the

training set. This is defined as

$$\mathrm{GEN} = \left\| \mathbf{z}^{\mathrm{te}} - F^{\mathrm{te}}\, \mathbf{vec}\left(\hat{Y}^{\mathrm{tr}}\right) \right\|_2^2 , \qquad (16)$$

where $\mathbf{z}^{\mathrm{te}}$ and F^{te} are the parts of $\mathbf{z}$ and F belonging to the test set.

2.2.2. *Effect of Joint Localization*

To illustrate the effect of co-localization, we performe the following experiment. A single dipolar source is placed in a cortical region of the brain and the resulting field pattern is computed. Ten different phase-shifted versions of the pattern are constructed by multiplication with a random unit-length complex number $\exp(i\phi)$. Each resulting pattern is superimposed by equal amounts of measurement and biological noise. Measurement noise is drawn from a Gaussian distribution with identity covariance matrix. Biological noise is correlated noise stemming from noisy brain sources. This brain noise is generated for each location in the brain using a Gaussian distribution with identity covariance matrix. The leadfield is used to project the noisy sources to the EEG. The signal-to-noise ratio is set to 1. Note that in this scenario, the SNR cannot be increased by averaging, since both signal and noise are zero-mean complex quantities. Source localization is carried out using both the single- and multiple-measurement variants of S-FLEX, where the regularization constant is set to match the exact SNR. The source estimates of all patterns are averaged to yield the estimated dipole amplitude per voxel. The obtained source amplitude map is compared with the true map, in which only one dipole is active, using the earth mover's distance. The experiment is repeated 100 times.

2.3. *Localization of Sensori-Motor Rhythms*

We will now consider real data from experiments recently conducted within the Berlin Brain-computer Interface (B)BCI project (Müller et al., 2008; Blankertz et al., 2007, 2008a,b; Tomioka and Müller, 2010; Blankertz et al., 2010a). These experiments originally had the purpose of screening subjects with respect to BCI aptitude. Note that the presented BCI data mainly serves as a *testbed* to explain the S-FLEX technique on non-simulated EEG data.

Brain-computer interfacing aims at providing paralyzed patients a communication device that "reads thoughts" and thereby obviates the need of using the usual motor pathway. A particularly successful approach to BCI is motor imagery, i.e., a system is controlled by a user deliberately switching between movement imaginations of certain limbs. It is known that for most people the associated contralateral sensorimotor cortex becomes active already during imagination of a movement. This leads to attenuation of local sensorimotor rhythms (predominantly in the μ-range) detectable by EEG/MEG; a phenomenon which is also known as Event-Related Desynchronization (ERD) (Pfurtscheller and Lopes da Silva, 1999).

2.3.1. *Experimental Setup*

Forty healthy BCI-naive subjects (33.6 ± 13.1 years old, 22 female) participated in the part of the study considered here [1] . During the experiment, they sat in a comfortable chair with arms and legs resting conveniently. EEG was acquired from 119 Ag/AgCl electrodes (reference at nasion). In a calibration session, arrows pointing left, right or down were presented on a screen and had to be responded by five seconds of left hand, right hand or foot motor imagery, respectively. Each arrow was presented 75 times. Using heuristics, well-discriminating contiguous post-stimulus time-intervals and frequency bands were identified (Blankertz et al., 2008b). Furthermore, a fixed number of spatial filters was computed for each subject. The log-bandpower of the temporally and spatially filtered EEG data was used to train regularized linear discriminant analysis (RLDA) classifiers discriminating between each pair of classes of motor imagery. The two classes showing best cross-validated separability on the calibration set were selected for use in a subsequent online session.

The locations of μ-ERD generators in hand- and foot areas of the sensorimotor cortex are well known from the literature (e.g., Golaszewski et al., 2002; Leonardo et al., 1995; Porro et al., 1996; Ciccarelli et al., 2005). These results obtained by functional Magnetic Resonance Imaging (fMRI) serve as a hypothesis here to be confirmed by EEG source localization. Since reliable source localization can only be expected from clean EEG signals, we restrict ourselves to cases in which EEG traces are clearly separable. That is, for each subject we are only interested in the two motor imagery classes showing best classifiability, and consider only those subjects who achieved very good BCI performance (error rates below 10%) in both the calibration and the online session. Ten subjects fell into that category. For nine of them, discriminability was found in the μ-band, as expected. A tenth subject, that achieved best BCI control using a broad band covering β- and γ-range, was excluded from the present investigation.

2.3.2. *Single-trial Localization of Fourier Coefficients*

In Yuan et al. (2008), Fourier-transformed single-trial EEG acquired during left/right hand motor imagery was subjected to a cortically-constrained version of the weighted minimum ℓ_2-norm source estimator (WMNE). A subsequent voxel-wise statistical test revealed maximal source discrimination between classes in the sensorimotor cortex. Here we pursue a similar approach. To this end, S-FLEX is applied to complex Fourier coefficients, which are calculated for each trial in the preselected band and pre-stimulus time interval by means of an FFT. Within the selected frequency band, five equidistantly-sampled Fourier coefficients are taken for each subject, regardless of the length of the selected time interval. Localization is

[1] Another forty experiments using the same paradigm were conducted at University of Tübingen. These are considered in Blankertz et al. (2010b), but not here.

conducted in the standard head model using a source grid of $N = 6\,249$ dipoles (7mm inter-dipole distance). We consider basis functions with widths $\sigma_1 = 0.75$ cm, $\sigma_2 = 1$ cm, $\sigma_3 = 1.25$ cm and $\sigma_4 = 1.5$ cm, which is slightly different from the dictionary used in the simulations. The regularization parameter to cope with noise influence is selected by means of five-fold cross-validation. That is, the parameter is chosen that minimizes the generalization error GEN on average. A multiple-measurement variant of S-FLEX is employed, where all Fourier-coefficient patterns of a trial are co-localized.

Note that the optimization problems to be solved in this joint localization setting are truly large-scale. Due to the comprehensiveness of the basis function dictionary used here ($L = 24\,996$ Gaussians with $N = 6\,249$ varying center positions and 4 different widths), the localization of a single EEG pattern ($M \approx 80$ electrodes) already amounts to fitting $3 \cdot 24\,996 = 74\,988$ model parameters (recall that three coefficients exist per basis function, one for each of the three spatial dimensions of the current density). When five Fourier are co-localized, the number of observations and variables effectively increases by a factor of ten, due to real and imaginary parts. Using DAL, however, the computational demands for such models are still moderate. Solving the optimization problem for a single choice of the regularization constant typically takes around three minutes on a modern laptop computer (1.58 GHz dual-core CPU, 2 GB RAM). Note that this is roughly the time required by our previously suggested method FVR to localize a single EEG pattern in a coarse source grid with 1 cm inter-dipole distance.

3. Results

3.1. *Simulations*

Fig. 2 shows a simulated current density along with reconstructions according to LORETA, L1, FVR and S-FLEX. From the Figure it is apparent, that LORETA and L1 do not approximate the true current density well. While the LORETA solution is rather blurry, not reflecting the shape of the true source, the L1 solution exhibits a number of spikes, which could be misinterpreted as different sources. The estimates of FVR and S-FLEX approximately recover the shape of the sources.

As quantified in Tab. 1, S-FLEX generalizes slightly better than its peers. More importantly, S-FLEX outperforms the other methods in terms of reconstruction accuracy.

Joint- as compared to single-pattern localization leads to a significantly better reconstruction of a single dipole, as indicated by a lower earth mover's distance (3.9 ± 0.1 as compared to 4.6 ± 0.1). An example is shown in Figure 3. While both estimated source distributions here peak similarly close to the true source location (indicated by a red cross-hair), the multiple-measurement approach has the advantage of being less spread-out. This illustrates that joint localization effectively removes the noise-induced spatial instability seen in single-trial estimates.

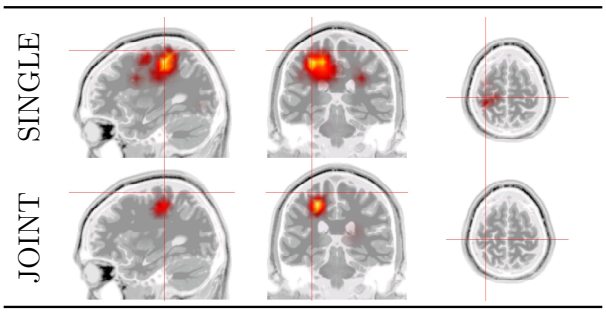

Fig. 3. Comparison of individual (SINGLE) and joint localization of ten simulated noisy measurements. The location of the true simulated source is indicated by a red cross-hair.

	REC	GEN
LORETA	1.00 ± 0.01	2.87 ± 0.78
L1	1.21 ± 0.01	1.86 ± 0.57
FVR	0.95 ± 0.02	1.21 ± 1.00
S-FLEX	0.71 ± 0.04	0.92 ± 0.28

Table 1
Ability of LORETA, L1, FVR and S-FLEX to reconstruct (REC) simulated currents and generalization performance (GEN) with respect to the EEG measurements.

3.2. *Localization of Sensori-Motor Rhythms*

The preprocessing steps performed prior to source localization were subject-specific, i.e., including filtering in individual frequency bands, the selection of individual time intervals and class combinations, and so on. For this reason, calculation of grand-averages is not possible, and results are presented here for single subjects. Table 2 lists the heuristically-chosen frequency-bands, time-intervals and the optimal class combinations for all subjects.

SUBJECT	BAND	INTERVAL	CLASSES
js	10.0 – 13.0 Hz	1000 – 4250 ms	LEFT/RIGHT
kp	8.0 – 12.5 Hz	1000 – 4500 ms	LEFT/RIGHT
ks	8.5 – 12.5 Hz	770 – 3950 ms	LEFT/RIGHT
kg	8.5 – 13.5 Hz	830 – 3440 ms	LEFT/FOOT
jj	8.5 – 13.5 Hz	1170 – 3970 ms	LEFT/FOOT
jl	9.5 – 13.5 Hz	1150 – 4160 ms	LEFT/FOOT
jy	7.5 – 11.5 Hz	1240 – 4070 ms	LEFT/FOOT
kc	9.5 – 13.5 Hz	1270 – 4000 ms	LEFT/FOOT
kd	9.5 – 13.5 Hz	1340 – 4110 ms	LEFT/FOOT

Table 2
Maximally-discriminating frequency-bands, time-intervals and optimal class combinations for nine subjects.

Source space discriminability maps obtained from localizing single-trial Fourier patterns are presented in Figure 4. These plots depict how much class-specific the brain activity during motor imagery is at each voxel. We use the signed

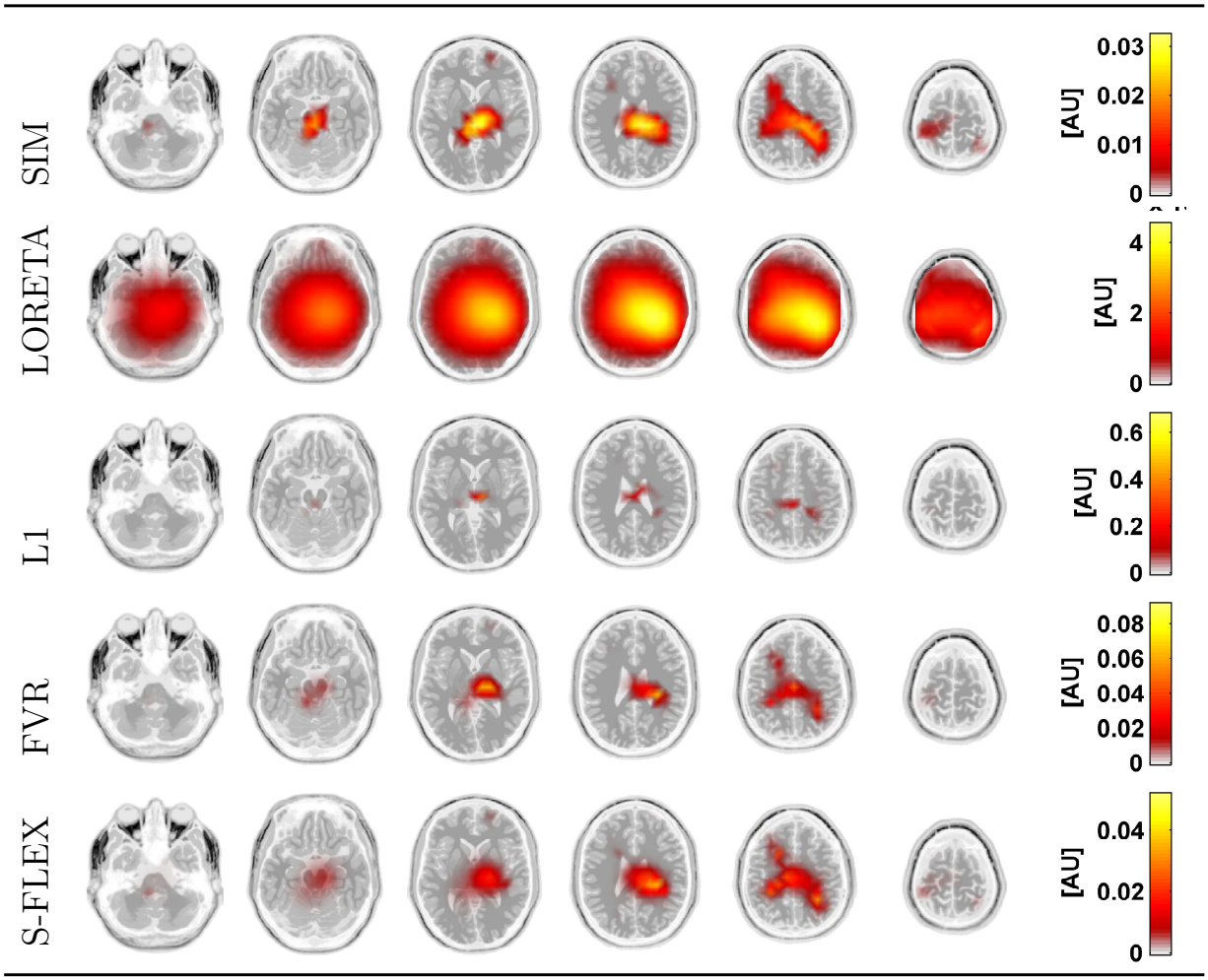

Fig. 2. Simulated current density (SIM) and reconstructions according to LORETA, L1, FVR and S-FLEX. Color encodes dipole magnitude.

r^2-value which is normalized by the average Band Amplitude (BA) as a discriminability index. Here, the Band Amplitude at a certain voxel is defined as the ℓ_2-norm of the estimated vector of Fourier-coefficients in the heuristically selected frequency band at that voxel, while the r^2-value is defined as the signed squared Pearson correlation coefficient between class label and Band Amplitude.

Subjects *js*, *kp* and *ks* achieved best BCI control by performing left vs. right hand motor imagery. This is reflected by the localization of the corresponding mental activity revealing an opposing contralateral activation of both hand areas. The other six subjects utilized foot imagery. In the left hand condition, all these subjects exhibit desynchronization of both hand areas simultaneously. This suggests that the foot condition often serves as a pseudo class, and some subject may effectively achieve one-dimensional BCI control utilizing only the presence or absence of (left *and* right) hand-related ERD.

4. Discussion

4.1. *Suitable Priors for Source Localization*

Computation of distributed EEG (or MEG) inverses heavily relies on regularization, since the physical model alone does not uniquely determine the sources. Within the last years the field has seen tremendous progress, in that the proposed regularization penalties more and more represent neurophysiologically meaningful prior knowledge. One important step has been the insight that the spatial spread of the sources should be controlled for, and ideally be tuned automatically using cross-validation or basis field selection. A second line of research was devoted to investigating how mutual information can be exploited, if several related field maps are to be localized jointly. In many cases, such as here, $\ell_{1,2}$-norm based grouping is suggested, which amounts to assuming that the sets of active brain sites have substantial overlap across EEG/MEG patterns. This assumption is certainly fulfilled for repeated measurements recorded under the same experimental condition, and will approximately hold also for narrow-band signals.

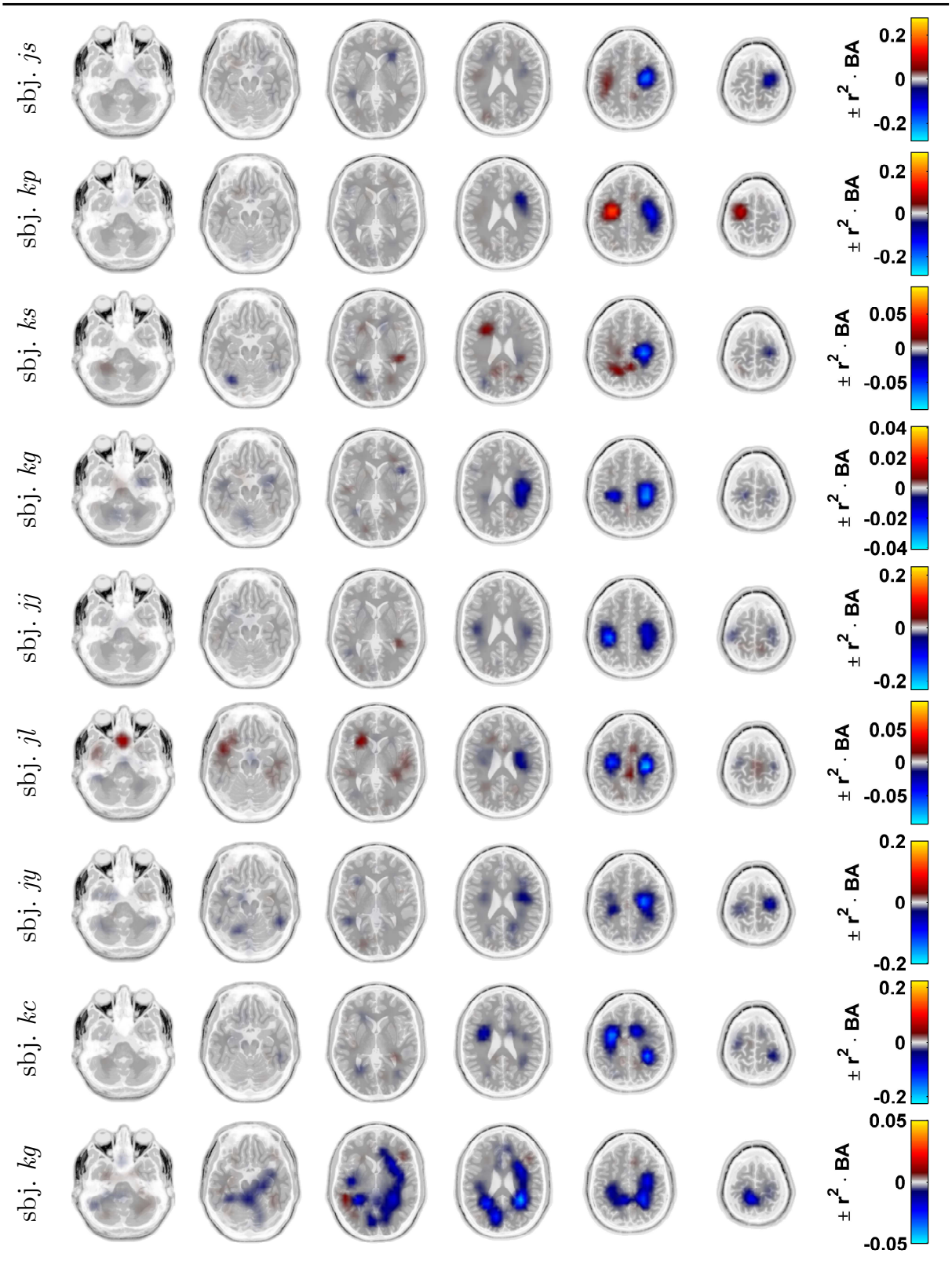

Fig. 4. S-FLEX source-space discriminability between best (discriminating) conditions for nine subjects. Source estimates are obtained from single-trial localization of Fourier-coefficients. For subjects *js*, *kp* and *ks* the optimal conditions are left and right hand imagery, while for all other subjects the optimal combination is left hand vs. foot imagery.

The validity of this model for wide-band and time-series data, however, remains unclear. In some cases, the known ordering of patterns might allow further exploitation, e.g., if smoothness (rather than concurrence) of sources is enforced in temporal or spectral directions. Since the resulting additional between-pattern coupling is currently not supported by DAL, the development of large-scale optimizers for such models will be subject to future work.

4.2. *Choice of the Regularization Parameter*

The choice of the regularization parameter, which controls how "complicated" source distributions are allowed to be, turned out to be a crucial point in our analyses. We observed that reasonable results are obtained using the cross-validated generalization error as a model selection criterion. This criterion can be efficiently computed for ℓ_2-norm regularized methods such as LORETA. Interestingly, using cross-validation for sparse methods such as L1 and S-FLEX the "optimal" model error (loss) is systematically estimated to be smaller than for LORETA, leading to seemingly too complex sources. For this and for efficiency reasons the optimal model error as estimated by LORETA was used as the regularization criterion throughout the study.

Another model selection criterion being widely used in the field of inverse problems consists in finding the corner of the L-curve, which is a log-log plot of the model error vs. the regularization term (Hansen, 1992). In our experiments, this criterion showed a tendency to regularize rather strongly, i.e. favor very simple sources. These often turn out to be deeper and more central. This can be seen in Figure 5, which is to be compared with the corresponding panels of Figure 4. In summary, both cross-validation and the L-curve method may fail under some conditions. It is a remaining challenge to derive noise estimates for oscillatory data, which could give rise to better model selection.

5. Conclusion

We contribute a novel methodology for obtaining sparse decompositions of vector fields especially tailored to EEG/MEG current density reconstruction. Furthermore, a well-suited efficient optimization scheme is suggested, that allows genuine large-scale localization. We have validated our method on simulated data and furthermore used high quality data from BCI experiments as testbed. When applied to EEG data from motor imagery BCI sessions, the signals recorded in different conditions showed highly separable source activation in the sensorimotor cortex, in agreement with the literature. This raises the question how source reconstruction methods can be used to improve BCI classification.

Acknowledgments This work was supported in part by BMBF grant No. 01GQ0850, DFG grant No. MU 987/3-1 and the FP7-ICT Programme of the European Community, under the PASCAL2 Network of Excellence, ICT-216886. We thank Vojtech Franc, Marius Kloft, Vadim Nikulin, Volker Roth and Alexander Zien for valuable discussions.

References

Baillet, S., Mosher, J. C., Leahy, R. M., 2001. Electromagnetic brain mapping. IEEE Signal Proc. Mag. 18 (6), 14–30.

Blankertz, B., Dornhege, G., Krauledat, M., Müller, K.-R., Curio, G., 2007. The non-invasive Berlin Brain-Computer Interface: Fast acquisition of effective performance in untrained subjects. NeuroImage 37 (2), 539–550.

Blankertz, B., Lemm, S., Treder, M. S., Haufe, S., Müller, K.-R., 2010a. Single-trial analysis and classification of ERP components – a tutorial. NeuroImage. In press.

Blankertz, B., Losch, F., Krauledat, M., Dornhege, G., Curio, G., Müller, K.-R., 2008a. The Berlin Brain-Computer Interface: Accurate performance from first-session in BCI-naive subjects. IEEE Trans. Biomed. Eng. 55 (10), 2452–2462.

Blankertz, B., Sannelli, C., Halder, S., Hammer, E. M., Kübler, A., Müller, K.-R., Curio, G., Dickhaus, T., 2010b. Neurophysiological predictor of SMR-based BCI performance. NeuroImage 51 (4), 1303–1309.

Blankertz, B., Tomioka, R., Lemm, S., Kawanabe, M., Müller, K.-R., Jan. 2008b. Optimizing spatial filters for robust EEG single-trial analysis. IEEE Signal Proc. Mag. 25 (1), 41–56.

Bolstad, A., Van Veen, B., Nowak, R., 2009. Space-time event sparse penalization for magneto-/electroencephalography. NeuroImage 46, 1066–1081.

Ciccarelli, O., Toosy, A. T., Marsden, J. F., Wheeler-Kingshott, C. M., Sahyoun, C., Matthews, P. M., Miller, D. H., Thompson, A. J., September 2005. Identifying brain regions for integrative sensorimotor processing with ankle movements. Exp. Brain Res. 166 (1), 31–42.

Ding, L., He, B., 2008. Sparse source imaging in EEG with accurate field modeling. Hum. Brain Mapp. 29 (9), 1053–1067.

Golaszewski, S. M., Siedentopf, C. M., Baldauf, E., Koppelstaetter, F., Eisner, W., Unterrainer, J., Guendisch, G. M., Mottaghy, F. M., Felber, S. R., Sep 2002. Functional magnetic resonance imaging of the human sensorimotor cortex using a novel vibrotactile stimulator. NeuroImage 17, 421–430.

Gorodnitsky, I. F., George, J. S., Rao, B. D., 1995. Neuromagnetic source imaging with FOCUSS: a recursive weighted minimum norm algorithm. Electroencephalogr. Clin. Neurophysiol. 95 (21), 231–251.

Hämäläinen, M. S., Ilmoniemi, R. J., 1994. Interpreting magnetic fields of the brain: minimum norm estimates. Med. Biol. Eng. Comput. 32, 35–42.

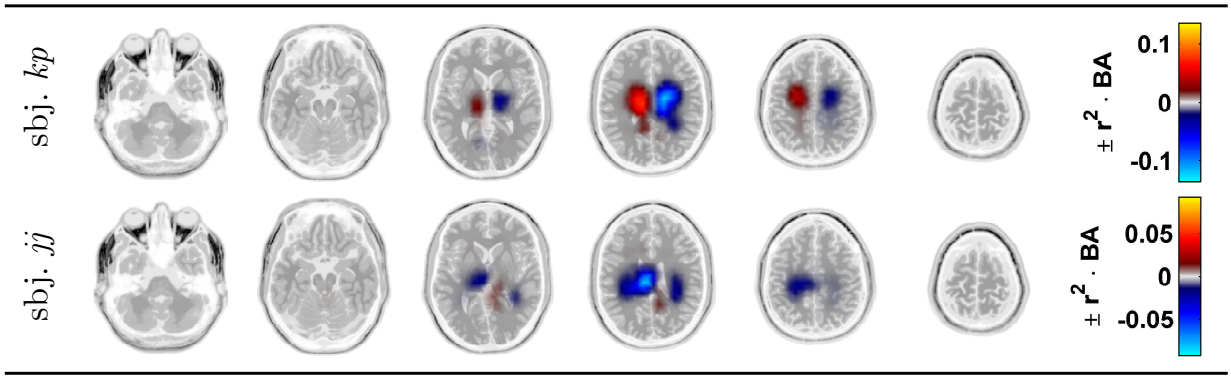

Fig. 5. S-FLEX source-space discriminability between conditions obtained from single-trial localization of Fourier-coefficients for subjects *kp* and *jj*, when the L-curve criterion is used for model selection.

Hansen, P. C., 1992. Analysis of discrete ill-posed problems by means of the L-curve. SIAM Rev. 34 (4), 561–580.

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 (2), 726–738.

Haufe, S., Nikulin, V. V., Ziehe, A., Müller, K.-R., Nolte, G., 2009. Estimating vector fields using sparse basis field expansions. In: Koller, D., Schuurmans, D., Bengio, Y., Bottou, L. (Eds.), Advances in Neural Information Processing Systems 21. MIT Press, Cambridge, MA, pp. 617–624.

Holmes, C. J., Hoge, R., Collins, L., Woods, R., Toga, A. W., Evans, A. C., 1998. Enhancement of MR images using registration for signal averaging. J. Comput. Assist. Tomogr. 22 (2), 324–333.

Huang, M.-X., Dale, A. M., Song, T., Halgren, E., Harrington, D. L., Podgorny, I., Canive, J. M., Lewise, S., Lee, R. R., 2006. Vector-based spatial-temporal minimum L1-norm solution for MEG. NeuroImage 31, 1025–1037.

Leonardo, M., Fieldman, J., Sadato, N., Campbell, G., Ibanez, V., Cohen, L., Deiber, M. P., Jezzard, P., Pons, T., Turner, R., Bihan, D. L., Hallett, M., 1995. A functional magnetic-resonance-imaging study of cortical regions associated with motor task execution and motor ideation in humans. Hum. Brain Mapp. 3, 83–92.

Lobo, M. S., Vandenberghe, L., Boyd, S., Lebret, H., 1998. Applications of second-order cone programming. Lin. Alg. Appl. 284, 193–228.

Malioutov, D., Çetin, M., Willsky, A. S., 2005. A sparse signal reconstruction perspective for source localization with sensor arrays. IEEE Trans. Signal Proces. 55 (8), 3010–3022.

Matsuura, K., Okabe, Y., 1995. Selective minimum-norm solution of the biomagnetic inverse problem. IEEE Trans. Biomed. Eng. 42, 608–615.

Mosher, J. C., Leahy, R. M., 1999. Source localization using recursively applied and projected (RAP) MUSIC. IEEE Trans. Signal Proces. 47 (2), 332–340.

Müller, K.-R., Tangermann, M., Dornhege, G., Krauledat, M., Curio, G., Blankertz, B., 2008. Machine learning for real-time single-trial EEG-analysis: From brain-computer interfacing to mental state monitoring. J. Neurosci. Meth. 176, 82–90.

Nolte, G., Dassios, G., 2005. Analytic expansion of the EEG lead field for realistic volume conductors. Phys. Med. Biol. 50, 3807–3823.

Ou, W., Hämäläinen, M. S., Golland, P., 2008. A distributed spatio-temporal EEG/MEG inverse solver. NeuroImage 44, 932–946.

Pascual-Marqui, R. D., 2002. Standardized low-resolution brain electromagnetic tomography (sLORETA): technical details. Meth. Find. Exp. Clin. Pharmacol. 24 (1), 5–12.

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

Pfurtscheller, G., Lopes da Silva, F. H., Nov 1999. Event-related EEG/MEG synchronization and desynchronization: basic principles. Clin Neurophysiol 110, 1842–1857.

Polonsky, A., Zibulevsky, M., 2004. MEG/EEG source localization using spatio-temporal sparse representations. In: Puntonet, C. G., Prieto, A. (Eds.), Independent Component Analysis and Blind Signal Separation, Fifth International Conference. Lecture Notes in Computer Science. Springer, pp. 1001–1008.

Porro, C. A., Francescato, M. P., Cettolo, V., Diamond, M. E., Baraldi, P., Zuiani, C., Bazzochi, M., di Prampero, P., 1996. Primary motor and sensory cortex activation during motor performance and motor imagery: A functional magnetic resonance imaging study. J. Neurosci. 16, 7688–7698.

Rubner, Y., Tomasi, C., Guibas, L. J., 2000. The earth mover's distance as a metric for image retrieval. Int. J. Comput. Vision 40 (2), 99–121.

Scherg, M., von Cramon, D., 1986. Evoked dipole source potentials of the human auditory cortex. Electroenceph. Clin. Neurophysiol. 65, 344–360.

Schmidt, R. O., 1986. Multiple emitter location and sig-

nal parameter estimation. IEEE Trans. Antenn. Propag. 43 (3), 276–280.

Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. J. Roy. Stat. Soc. B Meth. 58 (1), 267–288.

Tomioka, R., 2009. DAL. `http://mloss.org/software/view/183/`.

Tomioka, R., Müller, K. R., 2010. A regularized discriminative framework for EEG based communication. Neuroimage 49, 415–432.

Tomioka, R., Sugiyama, M., 2009. Dual augmented lagrangian method for efficient sparse reconstruction. IEEE Signal Proc. Let. 16 (2), 1067–1070.

Uutela, K., Hämäläinen, M., Somersalo, E., 1999. Visualization of magnetoencephalographic data using minimum current estimates. NeuroImage 10, 173–180.

van Gerven, M., Hesse, C., Jensen, O., Heskes, T., 2009. Interpreting single trial data using groupwise regularisation. NeuroImage 46, 665–676.

Van Veen, B. D., Van Drongelen, W., Yuchtman, M., Suzuki, A., 1997. Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. IEEE Trans. Biomed. Eng. 44 (9), 867–880.

Veen, B. D. V., Buckley, K. M., 1988. Beamforming: a versatile approach to spatial filtering. IEEE ASSP Magazine 5 (2), 4–24.

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

Wipf, D., Nagarajan, S., 2009. A unified Bayesian framework for MEG/EEG source imaging. Neuroimage 44, 947–966.

Wipf, D., Rao, B., 2007. An empirical bayesian strategy for solving the simultaneous sparse approximation problem. IEEE Trans. Signal Proces. 55 (7), 3704–3716.

Yuan, H., Doud, A., Gururajan, A., He, B., 2008. Cortical imaging of event-related (de)synchronization during online control of brain-computer interface using minimum-norm estimates in the frequency domain. IEEE Trans. Neural. Syst. Rehabil. Eng. 16, 425–431.

Yuan, M., Lin, Y., 2006. Model selection and estimation in regression with grouped variables. J. Roy. Stat. Soc. B Meth. 68 (1), 49–67.

Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net. J. Roy. Stat. Soc. B Meth. 67, 301–320.