

# As above, so below? Towards understanding inverse models in BCI

Jussi Lindgren

► **To cite this version:**

Jussi Lindgren. As above, so below? Towards understanding inverse models in BCI. Journal of Neural Engineering, IOP Publishing, 2017, 15 (1), 10.1088/1741-2552/aa86d0 . hal-01669325

**HAL Id: hal-01669325**

**<https://hal.inria.fr/hal-01669325>**

Submitted on 20 Dec 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# As above, so below?

## Towards understanding inverse models in BCI

Jussi T. Lindgren\*<sup>1</sup>

<sup>1</sup>Inria, France

August 9, 2017

### Abstract

In Brain-Computer Interfaces (BCI), measurements of the users brain activity are classified into commands for the computer. With EEG-based BCIs, the origins of the classified phenomena are often considered to be spatially localized in the cortical volume and mixed in the EEG. Does the reconstruction of the source activities in the volume help in building more accurate BCIs? The answer remains inconclusive despite previous work. In this paper, we study the question by contrasting the physiology-driven source reconstruction with data-driven representations obtained by statistical machine learning. Our analysis suggests that accuracy improvement from physiological source reconstruction in BCI may be expected mainly when machine learning cannot be used or where it produces suboptimal models. However, we argue that despite the use of physiology-based source reconstruction, data-driven techniques remain necessary to attain accurate BCI systems. Finally, we observe that many difficulties of the surface EEG classification remain challenges in the reconstructed volume.

**Keywords:** Brain-Computer Interface, electroencephalogram, inverse problems, machine learning, classification

## 1 Introduction

Brain-Computer Interfaces (BCI) operate by classifying measurements of users brain activity and using the predictions to control applications. The

---

\*Inria Rennes, 263 Avenue du General Leclerc, 35042 Rennes Cedex, France.  
E-mail: jussi.lindgren@inria.fr

electroencephalogram (EEG) -based BCIs are seriously challenged by their low information transfer rates [1, 2]. One possible reason for this is due to the EEG signal generation. Before the electrical potentials are measured by the surface EEG, the electrical activity of the brain has propagated through the neural tissues, the skull and the scalp. This *volume conduction* mixes, blurs and possibly cancels the contributions of the different electrical sources in the volume [3]. The volume conduction effects may then partially explain the difficulties in building usable BCI systems, as the obtained EEG signal is a mixture of the different sources and noise.

In typical BCI studies, machine learning techniques are used to estimate signal processing parameters and classifiers using EEG data recorded from the user [2, 4, 5]. These methods operate on the signal after it has been affected by the volume conduction, and need to work with the potentially introduced difficulties in the data. For example, if the discriminable part of the signal originated from a single brain region, these methods would need to implicitly discover and extract that part from the EEG mixture. To do so, they need training data, and sometimes they find only suboptimal solutions [6]. These properties are problematic, as the required data gathering (or calibration) step is an inconvenience to the user and yet the resulting accuracies may remain low. Can we require less data or obtain improvements to BCI accuracy if we use physiological knowledge and electromagnetic modeling to attempt to explicitly invert the volume conduction process? Subsequent classification would then proceed using the reconstructed signal sources in the volume (e.g. [7]). Here we call this method of classification using the reconstructed sources the *volume* (source reconstruction, inverse) approach in contrast to the *surface* approaches that also classify the EEG signal, but without explicitly attempting to recover sources in terms of a geometric volume model. Typically the volume approaches rely on a physiological head model, whereas the surface approaches tend to be based on statistical modeling of the measurements.

The source reconstruction can be expected to mitigate the volume conduction effects such as the mixing of the sources [8–10]. In addition it is appealing as the data in the volume representation is more easy to interpret in relation to the known structures of the brain [11]. Further, it allows setting physiological constraints to the classifier. For example, the solution can be restricted to a specific Region of Interest (ROI, e.g. [12]) which may be useful as it is known that popular BCI approaches such as motor imagery [13, 14] and Steady-State Visually Evoked Potentials (SSVEP, [15]) rely on detecting patterns that originate from cortical sources that are considered to be localized to specific anatomical and functional regions. Restricting the solution to the relevant regions potentially prevents unrelated sources from

interfering in the classification.

A nonnegligible amount of research work has gone into empirical EEG studies where inverse modeling is used in the BCI classification pipeline [7, 12, 16–31]. Despite this work, the evidence that volume reconstruction would lead to better classification accuracies than the surface approaches remains inconclusive. Although some empirical studies claim benefits for the volume approach [7, 21, 24, 25, 30, 31], yet other reports present accuracies that do not much differ from those obtained by surface methods [12, 18, 23, 26, 28]. In addition, some research groups have changed their opinion over time, for example the initial supportive claims of Goel et al. [25] about the volume approach were later replaced by a much more guarded statement [26]. There are no mathematical proofs of classification accuracy improvement arising from source reconstruction that the current authors are aware of.

A central problem with many empirical demonstrations is that the reported results cannot be conclusively attributed to the source reconstruction. In many studies the proposed volume approach is not compared with an equivalently structured surface approach with parameters optimized by machine learning, but to a more simple approach or not at all [7, 18, 24, 31]. This effectively inserts a structural bias to the comparison. In general, the fairness of the comparisons between the surface and volume approaches can be questioned, as well as the lack of independent confirmation studies (observed e.g. in [29]).

Instead of presenting another empirical study, here we take a different route. We examine more theoretically how the surface and volume solutions are related. In particular, we will explain that source reconstruction and using statistically estimated surface spatial filters in BCI are both approaches to obtain alternative linear representations of the data. We discuss the various related approaches and their expected properties. Finally, we consider what happens to certain difficult issues in BCI classification when the problem is considered in the volume instead of on the surface.

The rest of the paper is organized as follows. In sections 2 and 3 we briefly overview source reconstruction and BCI, respectively. In section 4 we explain how the two are related through the use of linear representations. In section 5, we examine the expected properties of different approaches to obtain linear representations. Then, we continue in section 6 to discuss how the representation in the volume affects several important challenges in BCI. In section 7 we describe considerations that should be kept in mind when comparing volume- and surface-based approaches. We suggest some future work in section 8 and finally present our conclusions in section 9.

## 2 Source reconstruction

The volume conduction propagates the electrical activity in the volume to surface potentials measured by EEG. The standard approach to reconstructing the sources is based on modeling the volume conduction (the *forward* problem) and then finding the source activities that explain the observations given the model. This equals solving an *inverse* problem. Let us briefly describe these two problems.

### 2.1 The forward problem

The EEG forward problem concerns computing the EEG measurements given the sources in the volume. A linear superposition model is generally accepted as a solution and can be derived from electromagnetics [3, 8, 32]. The model can be stated as follows [8–10],

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t), \quad (1)$$

where sample  $\mathbf{x} \in \mathfrak{R}^n$  contains the measured potentials from the  $n$  electrodes,  $\mathbf{s} \in \mathfrak{R}^m$  represents the  $m$  source amplitudes (usually  $m \gg n$ ),  $\mathbf{A} \in \mathfrak{R}^{n \times m}$  is the user-specific *leadfield* (gain) matrix and  $t$  is the sampling time<sup>1</sup>. The leadfield matrix encodes the volume conduction model that maps from the volume sources to the surface potentials. The matrix depends e.g. on anatomy and how the anatomy is modeled, the number of sources and electrodes and their positions, the degrees of freedom per source, and the measurement units. The matrix  $\mathbf{A}$  itself is nonlinear with respect to the source positions and orientations [8]. However, if the source positions and orientations have been fixed, the forward problem of eq. 1 can be considered linear. Here we use the common assumption that  $\mathbf{A}$  is a numeric, fixed matrix derived from physiology for a chosen source and electrode layout. The indexes of  $\mathbf{s}$  correspond to prefixed source positions in the 3D volume of the used head model. For details on leadfield construction see e.g. [8, 32].

The forward equation 1 can be naturally extended to multiple measurements (e.g. [8, 10, 33]), and we have

$$\mathbf{X} = \mathbf{A}\mathbf{S}, \quad (2)$$

where matrix  $\mathbf{X} \in \mathfrak{R}^{n \times l}$  contains  $l$  consequent multidimensional surface samples  $\mathbf{x}(t)$ , and  $\mathbf{S} \in \mathfrak{R}^{m \times l}$  denotes the corresponding matrix of sources  $\mathbf{s}(t)$ .

---

<sup>1</sup>It is common to include an additive noise term, i.e.  $\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t)$ . Here we omit the noise term to allow a more readable presentation.

If we assume  $\mathbf{A}$  to be a good model for the volume conduction, this implies several difficulties for EEG processing that tries to make inferences about the underlying volume activity. First, the matrix  $\mathbf{A}$  is typically not square as the number of potential sources  $m$  greatly exceeds the number of usable electrodes  $n$ , see e.g. [10]. In inverse-based BCI studies, the ratio  $m/n$  is usually between 20 and 100, see e.g. [24, 28, 30]. As a consequence, many different vectors  $\mathbf{s}$  can result in the same measurement  $\mathbf{x}$  (plus noise). The related inversion problem is called *ill-posed*. The matrix  $\mathbf{A}$  is also typically *ill-conditioned* [9, 10], i.e. the singular values of the matrix decay quickly. This can be due to nearby electrodes having almost collinear projections in the matrix, suggesting their volume conduction to be similar. An ill-conditioned matrix  $\mathbf{A}$  scales different source space directions very differently. If  $\sigma_i > 0$  is the  $i$ :th singular value of the matrix  $\mathbf{A}$ , we know from linear algebra that  $\mathbf{A}\mathbf{v}_i = \sigma_i\mathbf{u}_i$  for the corresponding left and right singular vectors  $\mathbf{u}_i$  and  $\mathbf{v}_i$  of the matrix  $\mathbf{A}$ . Thus, given the orthonormality of the singular vectors, constructing  $\mathbf{s} = \sum_i \alpha_i \mathbf{v}_i$  with scalars  $\alpha_i$ , we obtain  $\mathbf{x} = \mathbf{A}\mathbf{s} = \mathbf{A}(\sum_i \alpha_i \mathbf{v}_i) = \sum_i \alpha_i \sigma_i \mathbf{u}_i$ . In other words, the leadfield scales each component by its singular value  $\sigma_i$ . It may be difficult to differentiate the activity along strongly dampened source directions (small  $\sigma_i$ ) from measurement noise. Conversely, small measurement noise or even rounding errors in  $\mathbf{x}$  may require large changes to the source estimates if eq. 1 is satisfied exactly for each subsequent sample, even if the real source activities changed very little. In this sense an ill-conditioned matrix  $\mathbf{A}$  can be considered unstable.

## 2.2 The inverse problem

In the inverse problem the task is to recover the sources given the measurements. Non-uniqueness of the solutions and the ill-conditioning of the forward model prevent simply inverting the equation 1. A standard inverse cannot be used as  $\mathbf{A}$  is not square, and quick singular value decay can cause the solutions obtained with pseudoinverse to be dominated by noise. These difficulties can be mitigated by regularizing the solution [10, 34]. The regularized source reconstruction can be posed as an optimization problem (e.g. [35]),

$$\hat{\mathbf{S}} = \arg \min_{\mathbf{S}} [L_{data}(\mathbf{A}\mathbf{S}, \mathbf{X}) + \lambda L_{regul}(\mathbf{S})], \quad (3)$$

where  $L_{data} \geq 0$  and  $L_{regul} \geq 0$  are data and regularization loss functions, respectively, with the scalar  $\lambda \geq 0$  controlling their relative importance. The loss functions define the cost of each candidate  $\mathbf{S}$  and the solution to the reconstruction problem is  $\hat{\mathbf{S}}$ . Intuitively, the data loss is used to penalize

solutions that deviate from eq. 2, whereas the regularization loss attempts to penalize unlikely solutions that are expected to be erroneous, e.g. due to noise. For example, representing measurement noise may require large coefficients in  $\mathbf{S}$  due to the singular value decay of the leadfield. Such solutions can be avoided by assigning them a high cost with the regularization loss, with the possible tradeoff that the data loss increases.

In the above definition, the regularization loss can be considered as a *prior*. With suitable choices of the functional forms, the loss functions can correspond to an observation likelihood term and a prior term for specific probability distributions. For example, if  $l = 1$  (i.e. single sample inverse), choosing quadratic loss functions and solving the equation 3 corresponds to assuming Gaussian distributions [36] and equals to the well-known Minimum Norm algorithm. It can also be seen as Tikhonov regularization [10]. If a probabilistic interpretation exists for the chosen loss functions, the optimization can be seen as a maximum likelihood parameter estimation in probabilistic inference. For an extensive treatment of regularization in inverse problems, see e.g. [37].

Solutions to the regularized source reconstruction problem can be estimated by numerous algorithms that correspond to different assumptions and loss functions. These include the Minimum Norm (MN, [38]), Weighted Minimum Norm (WMN, e.g. [39]), and Laplacian Weighted Minimum Norm (LWMN, [40]) algorithms and yet others [10]. The optimal choice of the algorithm may depend on the data and what the reconstruction is used for [9, 35].

For many classical algorithms the solution can be obtained by a linear transform of the measurements, i.e.

$$\hat{\mathbf{S}} = \mathbf{W}\mathbf{X}, \quad (4)$$

for some matrix  $\mathbf{W}$  derived from  $\mathbf{A}$ . This is the case for example for the algorithms MN, WMN, and LWMN (see e.g. [10]). Interestingly, the mentioned algorithms do not require data  $\mathbf{X}$  to estimate the coefficients of the transform  $\mathbf{W}$ . EEG data can nevertheless be used to choose a suitable value for the regularization parameter  $\lambda$ .

There are also loss functions leading to solutions that are not obtained by linear transforms of the measurements. For example, using  $l_1$  norm - like regularization in eq. 3 states a preference for sparse solutions. Solving the corresponding optimization problem may require iterative methods [41]. Some other well-known nonlinear inverse algorithms such as FOCUSS [39] and sLoreta [42] use linear transforms like eq. 4 to provide initial solutions that are further refined.

### 3 Brain Computer Interfaces

Instead of aiming to recover the volume sources, EEG-based Brain-Computer Interfaces try to predict which category of mental activity resulted in the observed trial  $\mathbf{X}$ . Formally, we can specify a decision function

$$f_{\Theta}(\mathbf{X}) \mapsto \{1, \dots, k\}, \quad (5)$$

for  $k$ -class classification, where  $\Theta$  are the parameters of the function. If a decision function is found that accurately classifies short trial segments of different classes and generalizes that performance to previously unseen trials, a high bit-rate of the BCI is obtained. The user controls the end-application by alternating between the recognized mental activities.

In principle all the parameters  $\Theta$  of the decision function could be jointly estimated by machine learning using training data. To obtain the data, the user is given instructions to repeatedly perform mental tasks of different classes (e.g. imagine moving left hand, imagine moving right hand, ...). The EEG recording of each trial is then associated with the corresponding class label to form a training set.

Parameter estimation can be difficult for complicated, nonlinear functions [6]. In BCI, the overall decision function  $f_{\Theta}$  is often constructed stagewise as  $f_{\Theta}(\mathbf{X}) = g_{\theta_1}(h_{\theta_2}(\mathbf{X}))$ , where  $g$  is a classifier and  $h$  is a feature extractor, both with specific parameters to either specify or estimate. The feature extraction function  $h$  is usually customized for the type of BCI in question, and it can perform one or more transforms including linear temporal filtering, linear spatial filtering, averaging over time, and nonlinear transforms, among others. Then, with the features available, off-the-shelf supervised learning methods such as Linear Discriminant Analysis or Support Vector Machines are used to estimate the parameters of the classifier  $g$  [2, 4, 5].

In the scope of the current paper, we consider features  $\hat{\mathbf{S}}$  obtained using eqs. 3 or 4. If a leadfield with a volume interpretation is used to solve either equation, we have a volume BCI approach. Otherwise, we consider it a surface BCI. The latter equation is commonly used in surface BCI by directly specifying spatial filters as the rows of  $\mathbf{W}$ . Such filters have been found to help in reaching higher classification accuracies in BCI (e.g. [15, 43–45]). For example, in motor imagery, a spatial Laplacian filter has a feature-enhancing effect when positioned on the electrodes over the motor cortex [43]. Similarly in the P300 [46] and SSVEP paradigms [15], spatial filters can be estimated that extract the discriminable aspects of the signal and discard unrelated variation from it [15, 45].

Finally, before classification, the representation in  $\hat{\mathbf{S}}$  may undergo other



processing and nonlinearities, depending on the type of BCI. For example, a phase-locked P300 pattern can be detected without a nonlinear feature transform [46], whereas motor imagery uses square transforms to compute spectral powers (e.g. [13]).

## 4 Linear representations

Representation of the data with a linear superposition model of eq. 1 is common to both volume and surface methods in BCI. Due to the linear generative model, we speak of *linear representations*. The equation may hold only approximately in practice due to noise, the matrix  $\mathbf{A}$ , and the choice of the coefficients  $\hat{\mathbf{S}}$ . Solving equation eq. 2 for a given  $\mathbf{X}$  is known as dictionary learning or representation learning [47], and eq. 3 is one way to specify this objective.

For source reconstruction, the connection to dictionary methods has been made before (e.g. in [33]). The forward model  $\mathbf{A}$  of volume conduction can be seen as a basis or a *dictionary* to represent each trial  $\mathbf{X}$  using the corresponding *representation*  $\hat{\mathbf{S}}$  that is obtained by solving eq. 3. The columns of  $\mathbf{A}$  form a dictionary of spatial surface templates. The matrix  $\hat{\mathbf{S}}$  consists of the represented samples  $\hat{\mathbf{s}}$ . Each coefficient of  $\hat{\mathbf{s}}$  weights a linear template to participate in representing the corresponding  $\mathbf{x}$  with eq. 1. This can be easily seen by reordering the sums in the equation so that a sample  $\mathbf{x}$  is generated by a linear sum of weighted columns of  $\mathbf{A}$ , each multiplied by a single source-specific scalar coefficient from  $\hat{\mathbf{s}}$ . If the number of dictionary templates exceeds the dimension of the observed data ( $m > n$ ) as is typically the case for EEG forward models, the dictionary is called *overcomplete* [33, 47], although the corresponding equation system in eq. 1 is underdetermined. This is because from the viewpoint of representing the measurements, an overcomplete dictionary allows many different ways to represent each  $\mathbf{x}$ . The preferred representation can be selected by priors.

To see that the linear superposition model also exists for the spatial transforms used in BCI, we can observe starting from eq. 4 that  $\mathbf{W}^\dagger \hat{\mathbf{S}} = \mathbf{W}^\dagger (\mathbf{W}\mathbf{X}) = (\mathbf{W}^\dagger \mathbf{W})\mathbf{X} \approx \mathbf{X}$ , where  $\mathbf{W}^\dagger$  is the pseudo-inverse of  $\mathbf{W}$ . If we additionally assume that  $\mathbf{W}$  has rank  $n$ , the equation can be made exact. Denoting  $\mathbf{A} = \mathbf{W}^\dagger$ , we have obtained eq. 1, the forward model. Hence, we can interpret  $\hat{\mathbf{S}}$  obtained by spatial filters as a trial of abstract source activities that are mixed by  $\mathbf{W}^\dagger$  to obtain the measurements  $\mathbf{X}$ . Although here we illustrated that some forward model is implicitly encoded in the matrix  $\mathbf{W}$ , surface approaches where the model is explicit have also been proposed in the scope of BCI [48–50].

## 5 Finding representations for BCI

Using features based on linear representations in BCI requires specifying how  $\hat{\mathbf{S}}$  is obtained. There are various approaches for this purpose, and relying on source reconstruction is only one of them. Some other techniques estimate also the dictionary  $\mathbf{A}$ . If the result is achievable by eq. 4, it suffices to optimize  $\mathbf{W}$  as  $\hat{\mathbf{S}}$  follows by a simple multiplication. In the following we will discuss the various approaches to find representations and what kind of properties we can expect from them. We have chosen to categorize the different approaches based on the origin of the dictionary  $\mathbf{A}$  (or the corresponding transform  $\mathbf{W}$ , if available).

Note that the technical details of estimating the coefficients of  $\mathbf{A}$  are out of scope of the current work. Suffice to say that this task is mostly solved with nonlinear techniques, regardless if  $\mathbf{A}$  originates from anatomy or is estimated from the data. The reader is advised to consult the referred publications for details of the specific approaches.

### 5.1 Case 1 - Dictionaries from filter design

One way to find the representation is to make assumptions about the signal generation, and then design filters that work optimally when the assumptions are met. For example, it might be assumed that the superficial, radial electrical sources of the brain are of interest, and subsequently spatial Surface Laplacian filters can be obtained to enhance such sources. This technique is sometimes called Scalp Current Density (SCD) [3, 51, 52]. These spatial filters can be encoded in a matrix  $\mathbf{W}$ , each row of the matrix containing one filter centered on a different scalp location. Such filters are not fitted to the characteristics of any specific user or BCI task, and they do not explicitly localize cortical sources. Yet, they may mitigate volume conduction effects (e.g. [52]) and have been classically used to improve motor imagery BCI [43]. The representation is obtained by eq. 4, and a corresponding dictionary can be seen by taking a pseudoinverse of the filter matrix. The approach has a clear parallel in image coding where a set of pre-defined basis functions such as wavelets can be chosen as a dictionary to represent the data [47].

### 5.2 Case 2 - Leadfield as a dictionary

If we have more knowledge about the user's anatomy and the physics of volume conduction, these can be encoded in a leadfield  $\mathbf{A}$ . The studies proposing BCI based on source reconstruction generally vary in their choices for the leadfield parameterization and the used inverse algorithm (priors).

Specifying the leadfield as a fixed parameter can also be considered a very strong prior: it is a hard constraint specifying the templates with which the EEG signal should be represented. Solving eq. 3 provides a representation in terms of the reconstructed sources. For some algorithms, the solution may have the form of eq. 4 (see section 2.2).

If the source reconstruction objective can be solved by a linear transformation  $\mathbf{W}$  derived from a physiological model  $\mathbf{A}$ , the rows of matrix  $\mathbf{W}$  are spatial filters, each corresponding to recovering a specific source that is geometrically localized in the head model. To recover the sources, these projections perform spatial integration across the surface electrode measurements, possibly again resembling differently localized spatial Laplacian transforms [3]. The inverse transform uses information from the volume conduction model  $\mathbf{A}$ , but the result is not optimized for any specific BCI task. As the sources usually densely populate the volume or its subset (e.g. the cortex), if the information usable for discrimination is spread over a large region of different sources, this region must be then singled out and integrated in some manner after the source reconstruction. The source reconstruction is not doing task-specific volume integration or feature selection. To address these issues, many studies ultimately use recorded, labeled EEG data and statistical techniques to find appropriate integrative transforms and classifiers even when source reconstruction is used (e.g. [7, 22–24, 26, 28, 31]).

One important consideration when using source reconstruction for BCI is that not all inverse methods preserve the temporal structure of the signal [53]. As the major BCI paradigms (Motor Imagery, P300, and SSVEP) are based on detecting spatiotemporal patterns in the EEG, it may be preferable to use such inverse methods that retain the temporal structure.

### 5.3 Case 3 - Dictionary is learned

If we want to decompose the signal but do not have access to physiological models, we have a *blind source separation* problem (e.g. [54]). In blind source separation it is typically assumed that we know the function class of the mixing transform (e.g. a linear matrix), but not its parameter values. These values are then estimated from the data. To do so for linear representations, objective functions resembling eq. 3 are used, with the difference that in this case the dictionary  $\mathbf{A}$  is estimated as well. A loss term or hard constraints can be specified for the matrix  $\mathbf{A}$ , for example to prefer orthogonality or a small norm of the matrix. Again, priors on  $\hat{\mathbf{S}}$  are used to guide the optimization towards solutions that have desirable properties over the training data, such as sparsity or non-gaussianity [54]. Examples of approaches in this category include such well-known projection pursuit and unsupervised learn-

ing algorithms as Principal Components Analysis (PCA) [54], Independent Component Analysis (ICA) [54] and various dictionary learning techniques [47, 48, 50].

If the used technique provides a solution in the form of eq. 4, often two matrices are returned,  $\mathbf{A}$  with the columns corresponding to a dictionary of spatial templates and  $\mathbf{W}$ , with the rows containing spatial filters that extract the corresponding feature weights  $\hat{\mathbf{S}}$ . It should be noted that in this case the obtained representations do not have explicit connection to cortical anatomy or localized sources. However, the statistical methods may implicitly discover representations that correspond to physiologically localized sources [55–57] suggesting that to extract physiologically meaningful aspects from the data or mitigating the volume conduction effects does not necessarily require manually inserting explicit physiological knowledge to the processing pipeline.

Unlike representations found by source reconstruction that attempt to reconstruct each individual source, representations obtained from unsupervised methods can do integration both in the surface and implicitly in the volume, if dictated by the data and the optimized objective function. For example, if the method relies on the computation of eq. 4, the estimated transform  $\mathbf{W}$  would do volume integration implicitly by linearly mixing together different volume source estimates if that improves the objective function value. To give a more concrete example, if one brain region’s electrical activity were independent from the rest, ICA should attempt to combine the sources of this region and segregate them as one or more independent components separate from the rest of the activity (e.g. [58]).

Unsupervised methods can be used to obtain undercomplete, bijective and overcomplete dictionaries and the corresponding representations. If the used assumptions are valid and compatible with the data, then in principle these methods may find meaningful representations. The unsupervised methods fit parameters to the EEG of the user, but they do not explicitly attempt to tune the representation to differentiate between mental tasks in BCI.

## 5.4 Case 4 - Dictionary is learned using labels

The dictionary can also be optimized for a specific BCI task following well-established practices in BCI [4]. In addition to fitting classifiers with the labeled data, the used representation itself is optimized to make the data easier to classify. Although the representation and the classifier are still mostly optimized separately or even sequentially, this can be seen as a step towards estimating all the decision function parameters  $\Theta$  jointly from the data. As a difference to the unsupervised techniques we described before, here the trial labels are used together with the EEG data to define the priors

Problem	Description
Insufficient data	The amount of training data is insufficient to find a good statistical model, overfitting
Wrong assumptions	Objective function, model family, constraints or assumptions about the data distribution can be unsuitable
Bad regularization	The solution may be under- or overregularized
Local minima	The optimization was not able to find the globally optimal solution

Table 1: Some potential difficulties in fitting statistical models.

in the optimization objective.

Supervised techniques for representation learning in BCI are for example the Common Spatial Patterns (CSP) [44] for Motor Imagery, and xDAWN [45] for P300. Again, as a result, these methods typically return a linear transform  $\mathbf{W}$  and a corresponding dictionary  $\mathbf{A}$ . Methods also exist which use iterative techniques to find the representation, such as Discriminative Dictionary Learning (DDL) [49]. The dictionaries found with supervised learning are typically strongly undercomplete: a few basis functions can lead to better classification accuracies than using more of them [44, 45, 49]. This may be because the corresponding projections perform integration explicitly on the surface and implicitly in the volume, attempting to extract the relevant part of the data only. Recall that the BCI classification ultimately maps each trial  $\mathbf{X}$  to  $\log_2(k)$  bits for  $k$ -class classification. For BCI, this dimension reduction can already start in the representation stage, as there is no need to represent such parts of  $\mathbf{X}$  that are not useful for the classifier  $g$ . Having unrelated dimensions in the representation may increase the risk of overfitting the classifier [6].

A cost of using the statistical, data-driven methods is that a significant amount of training data might be needed to learn a representation and a classifier that are able to generalize well to new, unseen data [59]. Various other issues could interfere with the effort as well. We list some of the possible difficulties in table 1. Further discussions can be found in standard textbooks on statistical learning (e.g. [6]).

## 5.5 Case 5 - Combination approaches

The preceding cases 1 and 2 use prior knowledge in the form of fixed linear transforms or dictionaries. The cases 3 are 4 more data-driven, as then the dictionaries are estimated from the data. It is also possible to consider different combinations. In particular, approaches that combine physiological

	Physiological	Unsupervised	Supervised
Algorithms	MN, WMN, ...	PCA, ICA, ...	CSP, xDAWN, ...
From physiology	$\mathbf{A}$	-	-
From EEG	$\hat{\mathbf{S}}, \lambda$	$\mathbf{A}, \hat{\mathbf{S}}$	$\mathbf{A}, \hat{\mathbf{S}}$
Fitted to BCI task	-	-	$\mathbf{A}, \hat{\mathbf{S}}$
Size of $\mathbf{A}$	As provided, usually $m > n$	Selectable, depends on method	Selectable, usually $m < n$ .
Can use priors	Yes	Yes	Yes
Combines electrodes	Yes	Yes	Yes
Combines sources	Not intended	Implicitly	Implicitly
Source reconstruction	Yes	No	No

Table 2: Examples of properties that different approaches to representation learning can have (cases 2-4 in the text). The table is intended as illustrative and not conclusive. It is possible to design methods that change one or more properties of a specific approach. For details, see text.

leadfields and BCI trial label information have been proposed using a technique called beamformers (e.g. [18, 21, 23, 28]). These methods typically estimate projections  $\mathbf{W}$  that emphasize information originating from specific volume sources while suppressing information coming from elsewhere, allowing the methods to do volume integration in addition to spatial integration. Beamformers can be optimized to discriminate between different volume sources or regions, but they can also be combined with CSP-like objectives to find projections that extract the part of the EEG signal related to different classes [18, 21].

## 5.6 Summary

We summarize some typical properties of the cases 2-4 in Table 2. The main difference between the approaches is the origin of the dictionary and the extent that EEG data and trial labels are used to influence the parameters. In other respects the approaches are more similar as for example regularization can be considered whenever there are parameters to fit. However, physiological priors may be difficult to specify without a physiological model.

## 6 As above, so below

We now turn to illustrate that source reconstruction may not make the BCI classification problem trivial. We consider how the classification problem is affected by the volume reconstruction when the representation has been

obtained by solving eq. 3 using a physiology-based leadfield but without optimizing the representation for classification purposes. This corresponds to the section 5.2 before. In particular, we will discuss information recovery by the source reconstruction, the feature selection and integration question related to the regions of interest in the volume, as well as nonstationarity of the EEG signal generation.

## 6.1 What is not there can be difficult to restore

Classification of EEG trials in BCI is subject to the effects of the volume conduction that can cause blurring and canceling of the electrical activity during the propagation [3]. This is reflected in the properties of the leadfield models  $\mathbf{A}$ . The forward transform can both lose information completely as well as dampen it, as we described in section 2.1. Mathematically, different configurations of source activities  $\mathbf{s}$  map to the same measurements  $\mathbf{x}$  (plus random noise), and some vectors in the volume space have very weak projections that can be difficult to separate from the noise in  $\mathbf{x}$ .

Although reconstructing the sources may mitigate the volume conduction effects, it should be stressed that source reconstruction (or any other deterministic transform of the measurements) cannot restore such information that was never in the sources in the first place, and information lost by the volume conduction will remain lost. Nevertheless, it may be compensated with prior belief. Let us examine these two claims.

First, it is possible that in some cases, the original source activities themselves are difficult to classify. Motor imagery studies report that classification performance above random guessing accuracy could not be attained with some users, and further that in general the accuracy can improve by training the users [1]. It is common that the users in motor imagery studies have practiced the imagery for many hours over separate sessions (e.g. [13, 14, 43]). Unless the volume conduction and subsequently the forward model  $\mathbf{A}$  is assumed to change with imagination training, we can say that the training effect rather occurs in the nature of the original source activities  $\mathbf{S}$ , making the resulting EEG more amenable to classification. If the user does not generate discriminable activity in the real sources, reconstructing the sources should not be expected to help BCI.

Second, if we again assume  $\mathbf{A}$  to be an accurate model of volume conduction, a simple argument shows that it may cause irrevocable information loss. We know from linear algebra that vectors  $\mathbf{s}_0 \neq \mathbf{0}$  exist such that  $\mathbf{A}\mathbf{s}_0 = \mathbf{0}$ , as  $\mathbf{A}$  has a null-space with dimension  $m - n$ . We can now construct each column of a matrix  $\mathbf{S}_0$  from arbitrary linear combinations of such vectors, and encode information in the summation weights with the result that  $\mathbf{A}\mathbf{S}_0 = \mathbf{0}$ .

Consider now  $\mathbf{S} = \mathbf{S}' + \mathbf{S}_0$ . It follows that  $\mathbf{A}\mathbf{S} = \mathbf{A}\mathbf{S}' = \mathbf{X}$ , and hence if  $\mathbf{S}_0$  encoded information that was independent of  $\mathbf{S}'$ , that information cannot be recovered from  $\mathbf{X}$ . Nevertheless, priors such as the regularization loss in eq. 3 can help making assumptions about the missing information in the sense that the priors can translate the solutions  $\hat{\mathbf{S}}$  along the null-space of  $\mathbf{A}$ , whereas this translation is not visible for the data loss term. However, a solution influenced by the priors is not necessarily the original  $\mathbf{S}$ : the inserted information equals inserted assumptions. For example, choosing a solution  $\hat{\mathbf{S}}$  with the minimum norm by no means proves that the minimum norm configuration caused the measurements together with noise.

If lost information cannot be recovered unambiguously, it can be asked if the potential information loss can nevertheless be made smaller. One possible approach is to increase the number of electrodes used. It is known that dense electrode coverage may improve the source localization [60, 61] and thus provide more information about the source activity in general. However, it is not yet clear how useful this is with respect to the phenomena that are classified in BCI. Empirical evidence suggests that surface BCIs typically do not significantly improve after a certain point when more electrodes are added. On the contrary, it has been often observed that a small amount of well-placed electrodes can do better than a larger set (e.g. in motor imagery [62]), possibly due to increased possibilities to overfit the used statistical models. More generally, Graimann & al. state in their BCI introduction that "most groups try to minimize the number of electrodes to reduce setup time and hassle" [63]. It is not well-understood whether the limitations of the used signal processing techniques prevent benefiting from a high number of electrodes, or if after a certain point, additional electrodes recover only little additional information relevant for BCI. The latter could be the case if the volume conduction is very similar for nearby electrodes.

## 6.2 Where is my region of interest?

Instead of increasing the amount of discriminative information, a more plausible source of benefit from source reconstruction is that it allows a more intuitive interpretation of the used signal processing (e.g. allow less 'black-box' BCI [11]). This is because the signal components in the volume have a localized interpretation that can be visualized and contrasted to results from other imaging techniques operating in the volume and to knowledge from neurophysiology (e.g. [7, 12]). Subsequently, this knowledge can be used to reduce the complexity of the data.

For example, previous research results could be used to select a region of interest that is suspected to correspond to the mental activity that is being



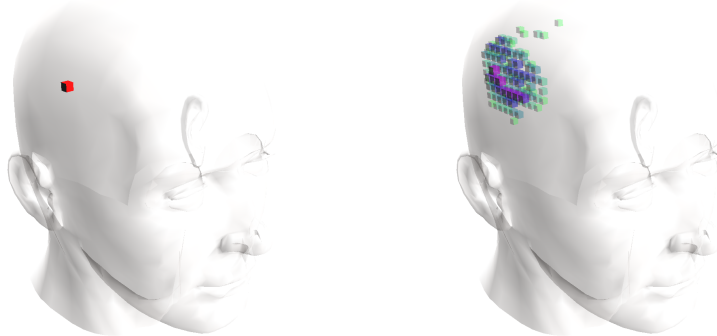


Figure 1: Illustration of source leakage. Left, a single active dipolar source. Right, reconstruction of the source activity visualized by the Euclidean norms of the sources. The strongest sources with norms in the highest 95% percentile are shown. For details, see text.

classified and to remove the activities of other sources from consideration (e.g. [7]). This is one way of setting priors in Bayesian modeling (see for example [6]), i.e. to bias further modeling and classification to prefer some explanations over others. With a ROI prior, some sources are ruled out altogether from further consideration<sup>2</sup>.

The setting of ROIs in the volume is nontrivial. First, for established paradigms there can be a reasonable prior guess about the shape, the extent and the location of the ROI due to previous research. For example, in motor imagery the kinesthetic imagination of right hand movement is known to affect the electrical activity on the contralateral side of the motor cortex [64]. But for novel paradigms the ROI is largely unknown and may be formed of distant cortical regions [24]. Second, although the total number of sources  $m$  is often large, the effective resolution of the reconstruction is typically low [8]. The reconstructed source activity blends with that of other nearby sources. This unintentional mixing of source activities in the volume can be seen as cross-talk or point spread between reconstructed sources [65] and is sometimes called *source leakage* [5]. The phenomenon is illustrated in Figure 1. The figure was made as follows. A single dipolar source was set to unit value in all 3 directions (visualized as a voxel). Other sources were set to zero. This source vector was projected with a leadfield  $\mathbf{A}$  to the surface and then reconstructed with the linear Minimum Norm solution. The leadfield was obtained from LORETA-KEY software (as in [22]). The

<sup>2</sup>Other priors have been studied e.g. in [12].

model was based on three concentric spheres, and had 2394 sources with 3 degrees of freedom each, their positions constrained to the cortical gray matter, and 60 electrodes. In these conditions, the originally active source does not have the highest strength in the MN reconstruction. The aperture of non-zeroes ('leakage') in the recovered volume covers a much wider area than a single dipole. This blurring effect reflects the information lost by the forward transform. The shape and extent of the blur and the possible mislocalization depends on the position and activation direction of the active dipole, the properties of the particular  $\mathbf{A}$  and the constraints used to obtain the reconstruction. In our case, the source leakage remains present when the dipole position and the direction of its activation are varied, and becomes worse for deep sources. Naturally, in normal brain operation, many different cortical regions are active simultaneously with the resulting electrical activity potentially canceling and interfering during the volume conduction, making accurate reconstruction even more challenging (not shown). The consequence of the source leakage is that ROIs in reconstructed volumes may need adjustment to account for the leakage effects.

Hence, the optimal ROI may not be generally known, it may be large or disconnected, and the source reconstruction may spread the activity to a larger area in the reconstructed volume than what was anticipated. If the ROI is larger than a single source, the BCI pipeline must specify how these reconstructed sources are combined to make predictions.

In general, forcing a strict ROI without looking at the data risks some of the information being reconstructed outside the ROI and thus not usable for classification. Perhaps due to this, data-driven statistical techniques are often used to answer the ROI selection problem in practice. A classical machine learning approach would be to do feature selection in the reconstructed volume [22]. It is also possible to narrow a manually selected ROI by statistical feature selection [30]. Interestingly, sometimes the ROI selection is effectively done with surface EEG data before applying the inverse transformation [18, 31] hinting that the feature selection problem might be perceived as more difficult in the high-dimensional volume than on the surface. Besserve & al. use the whole volume instead of selecting a ROI, letting a support vector machine to do implicit feature selection [24].

Selection of a spatial region is not the only sense in which a ROI may need to be considered. The ROI can also be temporal or spectral. For example, in P300, a short temporal pattern needs to be learned and detected over repeated trials (flashes) [46], suggesting that a relevant volume ROI is a temporal segment with some apriori unknown spatial aperture. For motor imagery, the optimal spectral power bands for classification may depend on the user [13, 21, 24]. Even using volume-based BCI, the spectral feature

selection problem is finally addressed by statistical means (e.g. [7, 31]).

### 6.3 Nonstationarity of EEG

In the previous sections we have considered difficulties that can already arise in situations where the generative process of the signal is stable but possibly different from user to user. Unfortunately, the generative properties change in a single user over both short and long periods of time [11, 66, 67]. These changes may be due to fatigue, alertness, attention, adaptation, plasticity, disease, and so on. In practice even the changes during short periods of time can lead to degradation of the classification accuracy in BCI and the interface must adapt [66, 67]. As a consequence, it is customary to estimate new decision function parameters for each BCI session with freshly obtained training data from the user in question.

The phenomenon of temporal change is called *concept drift* in machine learning and *nonstationarity* in statistics. As its consequence, either the decision function must be invariant to this kind of change, or the optimal parameterization of the decision function may need to change over time to account for the changes in the volume signal generation. If these changes concern the brain regions that produce the discriminative patterns, it seems unlikely that source reconstruction would avoid the requirement for adaptive models, only this time in the volume. On the other hand, if the signal changes are in non-related regions or subspace, source reconstruction with a ROI may allow removing the nonstationary phenomena from consideration. However, in that case properly fitted statistical models should also ignore that linear subspace.

The nonstationarity observed in the surface measurements can be attributed to nonstationarity in the volume, unless the measurement device adds a nonstationary component to the signal. This is because if  $\mathbf{A}$  is assumed fixed, it cannot cause nonstationarity, so the nonstationarity must originate from the process generating the source activities  $\mathbf{s}$ . With source reconstruction, this nonstationarity then needs to be addressed in the volume if it is interfering with the classification. Surface methods naturally face the same challenge.

### 6.4 Summary

Missing information, the differences in the generated activity from user to user, as well as the changes over time in the users' brain activity are among the challenges that can be expected to be present in the reconstructed sources if they were present on the surface. Table 3 summarizes these challenges.

Problem	Description
Missing information	Even the best (unknown) decision function would classify some $\mathbf{X}$ wrong
Classifier not known	The decision function needs to be chosen and parameterized for the user, includes ROI and feature selection
Nonstationarity	Optimal decision function may change over time if the generating process changes

Table 3: Potential challenges in BCI classification that can be expected to remain after source reconstruction.

## 7 Comparing volume and surface BCIs

It is difficult to conclusively attribute empiric results to be due to a specific approach. Based on our discussion, a fair comparison of volume and surface approaches in BCI should keep the signal processing pipelines otherwise identical except for the parameterization of the linear representation. This practice is sometimes followed [21, 30] but unfortunately not always [7, 18, 31]. Sometimes it is possible to go halfway. For example, Besserve & al. [24] use a transform  $\mathbf{W}(\mathbf{x} - \mathbb{E}[\mathbf{x}])$  both for volume and surface methods after some artifact clean-up. Here  $\mathbb{E}[\mathbf{x}]$  denotes the spatial mean, sometimes called Common Average Reference (CAR). However, for the surface method, they force  $\mathbf{W}$  to be an identity transform [24]. This way, the surface method is prevented from – even implicitly – performing localizing integration either spatially or in the volume before the nonlinear and noninvertible transforms of the spectral band power computations are done. In that case, the volume and surface decision functions are no longer members of the same function class. The surface approach is forced to work in a more limited space of functions with respect to  $\mathbf{X}$ . This kind of structural bias may potential explain the advantage the study reports for the volume method.

It is possible that the accuracies obtained with the volume methods are affected by the quality of the forward model, as the model is used to reconstruct the sources. Many volume BCI solutions use spherical, off-the-shelf leadfield models [12, 18, 21, 28]. Other studies have used average brain models [30, 31], with a few studies relying on user-specific head models [7, 23, 24]. In source localization in general, the effects of improving the forward model  $\mathbf{A}$  are well known (e.g. [9, 60]), but in BCI context there has been little work comparing forward models, with the exception of the study of Besserve & al. [24] who found increasing the volume resolution to improve the information transfer rate. Some papers take such benefits for granted [18]. The effect of the quality of the forward model  $\mathbf{A}$  is confounded by the practice

of using both physiological and statistical models in the same pipeline (e.g. [18, 24, 30, 31]). The effect of having an unoptimal forward transform can be compensated by statistical methods in the volume if the discriminable information from the surface is nevertheless retained in  $\hat{\mathbf{S}}$ . We can illustrate this for linear transforms. If  $\mathbf{A}\hat{\mathbf{S}} \approx \mathbf{X}$  (approximate due to noise), then  $\mathbf{W}\hat{\mathbf{S}} = (\mathbf{W}\mathbf{A})\hat{\mathbf{S}} = \mathbf{V}\hat{\mathbf{S}} \approx \mathbf{W}\mathbf{X}$ . To put it another way, any surface transform  $\mathbf{W}$  can be approximated with the reconstructed sources by a volume transform  $\mathbf{V} = \mathbf{W}\mathbf{A}$ . This way, further statistical modeling in the volume may discover a  $\mathbf{V}$  that compensates for inaccurate forward models or inverse transforms by implicitly reconstructing the measurements. The prerequisite for this possibility is that the reconstructed sources have not been processed in a manner that the function class optimizable by the statistical method is prevented from (implicitly) reverting the previous steps. For example, the typical squaring or power transform operation that is used to obtain frequency band power features used in motor imagery [43] is one such non-reversible and information-losing transform (for signed data). Selecting only a subset of the sources with a ROI is another example.

Finally, the quality of different BCI solutions must be weighted against their complexity and cost. For inverse based BCIs these costs can be high if dense-electrode EEG hardware and user-specific physiological models are desired. The largest user-specific cost comes from obtaining a custom leadfield. In this case, an anatomical Magnetic Resonance Imaging (MRI) scan must be taken of the user. Then, the obtained volume model must be segmented, meshed and used as a basis for a computationally costly estimation of the leadfield matrix [8]. This leadfield should ideally contain projections corresponding to the actual electrode placements on the user. Alternatively, an additional multiplicative matrix or interpolation can be constructed to map between the model and the actual electrodes. In this latter case, the electrode placements should be measured to compute the mapping. It is clear that the complete process of leadfield acquisition is not a casual procedure, and yet it is not complete. The estimation of the leadfield requires a tissue conductivity parameter which cannot be easily measured from live users [8]. A wrong guess can result in large source localization errors [60]. Perhaps due to these associated costs many volume-based BCI studies do not use user-specific models despite their better localization accuracy.

## 8 Future work

Future empirical studies demonstrating differences between volume and surface approaches should more clearly segregate the causes for the reported

results. In particular, identical training data should be used to construct two structurally identical pipelines, where the volume approach uses a linear representation obtained with a physiological leadfield as the dictionary, and the surface approach learns the dictionary using an appropriate statistical method such as CSP. Also, it would be of interest to understand which statistical fitting problems the source reconstruction has addressed (see table 1), if the volume approach provides better results than the surface approach. Conversely, surface approach performing better may reveal shortcomings in the inverse modeling and the used assumptions.

The current paper focuses on single-user BCI and builds on the common assumption that the volume conduction is modeled well by a linear and unchanging transform. The effects of allowing nonlinearities and adaptivity should be further studied, while keeping in mind that surface methods may be similarly extendable.

Regarding the properties of the BCI classification problem in general, it should be better understood how the spatiotemporal signal characteristics used for BCI (e.g. ERP patterns in P300, the flicker frequency -specific power effects in SSVEP, and the beta-band depression in Motor Imagery) are affected by the volume conduction. If these characteristics are dampened differently in terms of signal to noise ratio (SNR) in different user models due to anatomical differences, this might help to explain the large variability of BCI accuracies obtainable with different users. Understanding the BCI SNR and source activity distributions better in the volume could also be illustrative regarding the prospects of using source reconstruction in BCI.

## 9 Conclusion

Although source reconstruction has been often studied and proposed for BCI [7, 12, 16–31], it remains an exotic approach which is covered only cursorily by BCI reviews and textbooks [2, 5]. We have argued that physiological source reconstruction and the more common spatial transforms used in BCI are both methods that seek alternative representations of the EEG in terms of linear superposition models. The difference is that source reconstruction methods rely on leadfields derived from physiology, whereas statistical surface methods estimate the corresponding parameters using the EEG data and labels. We have discussed various ways to obtain these representations and their expected properties. Finally, we have considered three major difficulties in EEG classification to illustrate that source reconstruction does not make the BCI problem trivial. Instead, source reconstruction transforms surface classification problem into a volume classification problem, and data-driven

techniques remain necessary to address challenges that are still present in the reconstructed volume.

Regarding surface and volume classification, it appears that there is above, but no below: all EEG classification is classification of the same surface measurements. The difference is in the data representation and in the eye of the beholder. Obtaining a volume representation using a user-specific leadfield entails a potentially large cost and increased complexity in addition to the classification problem that remains to be solved, now with the transformed data. Whether this increased modeling complexity is worth its cost, or simply something Occam warned against, remains to be seen.

## Acknowledgement

The author wishes to thank Fabien Lotte, Anatole Lécuyer, Theodore Papadopoulo and the anonymous reviewers for their helpful comments and suggestions on earlier versions of this manuscript. This work was funded by the Labex CominLabs project SABRE.

## References

- [1] Allison B, Neuper C. Could Anyone Use a BCI? In: Tan DS, Nijholt A, editors. *Brain-Computer Interfaces. Human-Computer Interaction Series*. Springer; 2010. p. 35–54.
- [2] Nicolas-Alonso LF, Fernando L, Gomez-Gil J. Brain Computer Interfaces, a Review. *Sensors (Basel)*. 2012;12(2):12111279.
- [3] Nunez PL, Srinivasan R. *Electric Fields of the Brain*, 2nd edition. Oxford University Press; 2006.
- [4] Lotte F, Congedo M, Lécuyer A, Lamarche F, Arnaldi B. A review of classification algorithms for EEG-based brain-computer interfaces. *Journal of Neural Engineering*. 2007;4(2).
- [5] Clerc M, Bougrain L, Lotte(eds) F. *Brain-Computer Interfaces*. vol. 1-2. ISTE-Wiley; 2016.
- [6] Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning*, 2nd edition. Springer; 2008.

- [7] Cincotti F, Mattia D, Aloise F, Bufalari S, Astolfi L, Fallani FDV, et al. High-Resolution EEG Techniques for Brain-Computer Interface Applications. *Journal of Neuroscience Methods*. 2008;167(1).
- [8] Baillet S, Mosher JC, Leahy RM. Electromagnetic brain mapping. *IEEE Signal Processing Magazine*. 2001;18:14–30.
- [9] He B, Lian J. Electrophysiological Neuroimaging. In: He B, editor. *Neural Engineering*. Springer US; 2005. p. 221–261.
- [10] Grech R, Cassar T, Muscat J, Camilleri KP, Fabri SG, Zervakis M, et al. Review on solving the inverse problem in EEG source analysis. *Journal of NeuroEngineering and Rehabilitation*. 2008;5(25).
- [11] Makeig S, Kothe C, Mullen TR, Shamlo NB, Zhang Z, Kreutz-Delgado K. Evolving Signal Processing for Brain-Computer Interfaces. *Proceedings of the IEEE*. 2012;100(Centennial-Issue):1567–1584.
- [12] Noirhomme Q, Kitney RI, Macq B. Single-Trial EEG Source Reconstruction for Brain-Computer Interface. *IEEE Transactions on Biomedical Engineering*. 2008;55(5).
- [13] Pfurtscheller G, Neuper C, Flotzinger D, Pregenzer M. EEG-based discrimination between imagination of right and left hand movement. *Electroencephalography and Clinical Neurophysiology*. 1997;103(6):642 – 651.
- [14] Wolpaw JR, McFarland DJ. Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans. *PNAS*. 2004;101(51):17849–17854.
- [15] Vialatte FB, Maurice M, Dauwels J, Cichocki A. Steady-state visually evoked potentials: Focus on essential paradigms and future perspectives. *Progress in Neurobiology*. 2010;90(4):418 – 438.
- [16] Qin L, Ding L, He B. Motor Imagery Classification by Means of Source Analysis for Brain Computer Interface Applications. *Journal of Neural Engineering*. 2004;1(3):135–141.
- [17] Kamousi B, Liu Z, He B. Classification of Motor Imagery Tasks for Brain-Computer Interface Applications by means of Two Equivalent Dipoles Analysis. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*. 2005;13(2):166–171.



- [18] Congedo M, Lotte F, Lécuyer A. Classification of movement intention by spatially filtered electromagnetic inverse solutions. *Physics in Medicine and Biology*. 2006;51(8).
- [19] Lehembre R, Noirhomme Q, Macq B. Inverse Problem Applied to BCIs : Keeping Track of the EEG's Brain Dynamics Using Kalman Filtering. In: *Proc. of the 3rd International Brain Computer Interface Workshop*; 2006. p. 32–33.
- [20] Babiloni F, Cincotti F, Marciani M, Salinari S, Astolfi L, Tocci A, et al. The Estimation of Cortical Activity for Brain-Computer Interface: Applications in a Domotic Context. *Computational Intelligence and Neuroscience*. 2007;91651.
- [21] Grosse-Wentrup M, Liefhold C, Gramann K, Buss M. Beamforming in noninvasive brain-computer interfaces. *IEEE Transactions on Biomedical Engineering*. 2009;56(4):1209–1219.
- [22] Lotte F, Lécuyer A, Arnaldi B. FuRIA: An inverse Solution based Feature Extraction Algorithm using Fuzzy Set Theory for Brain-Computer Interfaces. *IEEE transactions on Signal Processing*. 2009;57(8):3253–3263.
- [23] Fruitet J, Clerc M. Reconstruction of cortical sources activities for on-line classification of electroencephalographic signals. In: *Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology*; 2010. p. 6317–6320.
- [24] Besserve M, Martinerie J, Garnero L. Improving quantification of functional networks with EEG inverse problem: Evidence from a decoding point of view. *NeuroImage*. 2011;55:1536–1547.
- [25] Goel M, Chavarriaga R, Millán J. Cortical Current Density vs. Surface EEG for Event-Related Potential-based Brain-Computer Interface. In: *5th International IEEE EMBS Conference on Neural Engineering*; 2011. p. 430–433.
- [26] Goel M, Chavarriaga R, Millán J. Comparing BCI performance using scalp EEG- and inverse solution-based features. In: *6th International Brain-Computer Interface Conference*; 2014. .
- [27] Ahn M, Cho H, Jun SC. Calibration time reduction through source imaging in brain computer interface (BCI). In: *HCI International 2011–Posters Extended Abstracts*. Springer; 2011. p. 269–273.

- [28] Ahn M, Hong JH, Jun SC. Feasibility of approaches combining sensor and source features in brain-computer interfaces. *Journal of Neuroscience Methods*. 2012;204(1):168–178.
- [29] Fruitet J. Interfaces Cerveau-Machines basees sur l’imagination de mouvements brefs: vers des boutons controles par la pensee. PhD thesis, University of Nice - Sophia Antipolis; 2012.
- [30] Zaitcev A, Cook G, Liu W, Paley M, Milne E. Feature extraction for BCIs based on electromagnetic source localization and multiclass Filter Bank Common Spatial Patterns. In: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 2015. p. 1773–1776.
- [31] Edelman BJ, Baxter B, He B. EEG Source Imaging Enhances the Decoding of Complex Right-Hand Motor Imagery Tasks. *IEEE Transactions on Biomedical Engineering*. 2016;63:4–14.
- [32] Hallez H, Vanrumste B, Grech R, Muscat J, Clercq WD, Vergult A, et al. Review on solving the forward problem in EEG source analysis. *Journal of NeuroEngineering and Rehabilitation*. 2007;4(46).
- [33] Cotter SF, Rao BD, Egan K, Kreutz-Delgado K. Sparse solutions to linear inverse problems with multiple measurement vectors. *IEEE Transactions on Signal Processing*. 2005;53(7):2477–2488.
- [34] Michel CM, Murray MM, Lantz G, Gonzales S, Spinelli L, de Peralta RG. EEG source imaging. *Clinical Neurophysiology*. 2004;115(10):2195–2222.
- [35] Gramfort A, Kowalski M, Hämäläinen M. Mixed-norm estimates for the M/EEG inverse problem using accelerated gradient methods. *Physics in Medicine and Biology*. 2012;57:1937–1961.
- [36] Wipf D, Nagarajan S. A unified Bayesian framework for MEG/EEG source imaging. *NeuroImage*. 2009;44(3):947 – 966.
- [37] Engl HW, Hanke M, Neubauer A. *Regularization of Inverse Problems*. Springer; 2000.
- [38] Hämäläinen MS, Ilmoniemi RJ. Interpreting measured magnetic fields of the brain: estimates of current distributions. Helsinki University of Technology; 1984. Technical Report TTK-F-A559.

- [39] Gorodnitsky IF, Rao BD. Sparse Signal Reconstruction from Limited Data Using FOCUSS: A Re-weighted Minimum Norm Algorithm. *IEEE Transactions on Signal Processing*. 1997;45(3):600–616.
- [40] Pascual-Marqui RD, Michel CM, Lehmann D. Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain. *International Journal of Psychophysiology*. 1994;18(1):49 – 65.
- [41] Fuchs M, Wagner M, Köhler T, Wischmann HA. Linear and nonlinear current density reconstructions. *Journal of Clinical Neurophysiology*. 1999;16(3):267–295.
- [42] Pascual-Marqui R. Standardized low-resolution brain electromagnetic tomography (sLORETA): technical details. *Methods Find Exp Clin Pharmacol*. 2002;24:5–12.
- [43] McFarland DJ, McCane LM, David SV, Wolpaw JR. Spatial filter selection for EEG-based communication. *Electroencephalography and Clinical Neurophysiology*. 1997;103(3):386–394.
- [44] Müller-Gerking J, Pfurtscheller G, Flyvbjerg H. Designing optimal spatial filters for single-trial EEG classification in a movement task. *Clinical Neurophysiology*. 1999;110(5):787–798.
- [45] Rivet B, Souloumiac A, Attina V, Gibert G. xDAWN Algorithm to Enhance Evoked Potentials: Application to Brain-Computer Interface. *IEEE Transactions on Biomedical Engineering*. 2009;56(8).
- [46] Fazel-Rezai R, Allison BZ, Guger C, Sellers EW, Kleih SC, Kübler A. P300 brain computer interface: current challenges and emerging trends. *Frontiers in Neuroengineering*. 2012;5(14).
- [47] Rubinstein R, Bruckstein AM, Elad M. Dictionaries for Sparse Representation Modeling. *Proceedings of the IEEE*. 2010 June;98(6):1045–1057.
- [48] Hamner B, Chavarriaga R, Millán JR. Learning Dictionaries of Spatial and Temporal EEG Primitives for Brain-Computer Interfaces. In: *Proc. Workshop Structured Sparsity: Learning and Inference*; 2011. .
- [49] Zhou W, Yang Y, Yu Z. Discriminative dictionary learning for EEG signal classification in Brain-computer interface. In: *2012 12th International Conference on Control Automation Robotics Vision (ICARCV)*; 2012. p. 1582–1585.

- [50] Barthélemy Q, Gouy-Pailler C, Isaac Y, Souloumiac A, Larue A, Mars JJ. Multivariate temporal dictionary learning for EEG. *Journal of Neuroscience Methods*. 2013;215(1):19 – 28.
- [51] Perrin F, Bertrand O, Pernier J. Scalp Current Density Mapping: Value and Estimation from Potential Data. *IEEE Transactions on Biomedical Engineering*. 1987;34(4):283–288.
- [52] Burle B, Spieser L, Roger C, Casini L, Hasbroucq T, Vidal F. Spatial and temporal resolutions of EEG: Is it really black and white? A scalp current density view. *International Journal of Psychophysiology*. 2015;97(3):210 – 220.
- [53] Liu H, Schimpf PH, Dong G, Gao X, Yang F, Gao S. Standardized shrinking LORETA-FOCUSS (SSLOFO): a new algorithm for spatio-temporal EEG source reconstruction. *IEEE Transactions on Biomedical Engineering*. 2005;52(10):1681–1691.
- [54] Hyvärinen A, Karhunen J, Oja E. *Independent Component Analysis*. John Wiley & Sons; 2004.
- [55] Koles ZJ, Soong AC. EEG source localization: implementing the spatio-temporal decomposition approach. *Electroencephalography Clinical Neurophysiology*. 1998;107(5):343–353.
- [56] Kobayashi K, Akiyama T, Nakahori T, Yoshinaga H, Gotman J. Systematic source estimation of spikes by a combination of independent component analysis and RAP-MUSIC: II: Preliminary clinical application. *Clinical Neurophysiology*. 2002;113(5):725 – 734.
- [57] Delorme A, Palmer J, Onton J, Oostenveld R, Makeig S. Independent EEG Sources Are Dipolar. *PLOS ONE*. 2012;7:1–14.
- [58] Onton J, Westerfield M, Townsend J, Makeig S. Imaging human EEG dynamics using independent component analysis. *Neuroscience and Biobehavioral Reviews*. 2006;30(6):808822.
- [59] Lotte F. Signal Processing Approaches to Minimize or Suppress Calibration Time in Oscillatory Activity-Based Brain–Computer Interfaces. *Proceedings of the IEEE*. 2015;103(6).
- [60] Acar Z, Makeig S. Effects of Forward Model Errors on EEG Source Localization. *Brain Topography*. 2013;26:378–396.

- [61] Song J, Davey C, Poulsen C, Luu P, Turovets S, Anderson E, et al. EEG source localization: Sensor density and head surface coverage. *Journal of Neuroscience Methods*. 2015;256:9 – 21.
- [62] Tam WK, Tong K, Meng F, Gao S. A Minimal Set of Electrodes for Motor Imagery BCI to Control an Assistive Device in Chronic Stroke Subjects: A Multi-Session Study. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*. 2011;19(6):617–627.
- [63] Graimann B, Allison B, Pfurtscheller G. Brain–Computer Interfaces: A Gentle Introduction. In: Graimann B, Pfurtscheller G, Allison B, editors. *Brain-Computer Interfaces: Revolutionizing Human-Computer Interaction*. Springer Berlin Heidelberg; 2010. p. 1–27.
- [64] Pfurtscheller G, Neuper C. Motor imagery activates primary sensorimotor area in humans. *Neuroscience Letters*. 1997;239(2-3).
- [65] Liu AK, Dale AM, Belliveau JW. Monte Carlo simulation studies of EEG and MEG localization accuracy. *Human Brain Mapping*. 2002;16(1):47–62.
- [66] Wolpaw JR, Birbaumer N, McFarland DJ, Pfurtscheller G, Vaughan TM. Brain–computer interfaces for communication and control. *Clinical Neurophysiology*. 2002;113:767–791.
- [67] Shenoy P, Krauledat M, Blankertz B, Rao RP, Müller KR. Towards adaptive classification for BCI. *Journal of Neural Engineering*. 2006;3(1).