A Sparse Dictionary Learning Framework to Discover Discriminative Source Activations in EEG Brain Mapping

Feng Liu,[1] Shouyi Wang,[1] Jay Rosenberger,[1] Jianzhong Su,[2] Hanli Liu[3]
Department of Industrial, Manufacturing & Systems Engineering[1],
Department of Mathematics[2], Department of Bioengineering[3]
The University of Texas at Arlington, Arlington, TX 76019, USA
feng.liu@mavs.uta.edu,{shouyiw,jrosenb,su,hanli}@uta.edu

Abstract

Electroencephalography (EEG) source analysis is one of the most important noninvasive human brain imaging tools that provides millisecond temporal accuracy. However, discovering essential activated brain sources associated with different brain status is still a challenging problem. In this study, we propose for the first time that the ill-posed EEG inverse problem can be formulated and solved as a sparse over-complete dictionary learning problem. In particular, a novel supervised sparse dictionary learning framework was developed for EEG source reconstruction. A revised version of discriminative K-SVD (DK-SVD) algorithm is exploited to solve the formulated supervised dictionary learning problem. As the proposed learning framework incorporated the EEG label information of different brain status, it is capable of learning a sparse representation that reveal the most discriminative brain activity sources among different brain states. Compared to the state-of-the-art EEG source analysis methods, proposed sparse dictionary learning framework achieved significant superior performance in both computing speed and accuracy for the challenging EEG source reconstruction problem through extensive numerical experiments. More importantly, the experimental results also validated that the proposed sparse learning framework is effective to discover the discriminative task-related brain activation sources, which shows the potential to advance the high resolution EEG source analysis for real-time non-invasive brain imaging research.

Introduction

In the past few decades, numerous noninvasive measurements of brain activity have been proposed, implemented and applied in clinical treatment and scientific research communities. One of the most popular techniques is electroencephalography (EEG) for its advantages of low cost, portable and high temporal resolution. EEG signals measure the electrical voltage on a variety of locations on the scalp. The electrical potentials on the scalp are superimposition of electric activity of neurons inside the brain and several types of physiological and non-physiological artifacts (Haufe 2011; Haufe et al. 2013). To be specific, the EEG signal arises as a result of synchronous intracellular current flows from the depolarized membranes of apical dendrite and non-excited cell soma and basal dendrites (Baillet, Mosher, and Leahy 2001). Those above currents (sources) propagate through the conductive medium, which is approximated with the conductivity properties of different tissues overlaying each other, to the scalp which is measured by EEG electrodes.

To infer the brain sources from the scalp recorded EEG signals belongs to the class of *inverse problem*. Precise localization of neuronal activity inside the brain can offer an insightful understanding of how brain is functioning given certain cognitive and motion tasks. Recent years have witnessed a shift in neuroscience community from traditional "segregation" perspective to "integration" perspective in which the functional and effective connectivity between different regions of brains (Haufe 2011; Hipp et al. 2012; Liu et al. 2016) are investigated using complex network characteristics measurement (Watts and Strogatz 1998; Barabási and Albert 1999; Guan et al. 2012; Newman 2003). Connectivity between different parcellations of brain is established by measuring the similarity of reconstructed sources. As the source reconstruction or solving the inverse problem is tenably the first and primary step for connectivity analysis of the brain (Sockeel et al. 2016; Ma et al. 2016), precise localization of sources is required in order to gain solids result using complex network measurement in the latter steps.

In this paper, we aim to calculate the discriminative sources to facilitate the understanding of brain mechanism under different cognitive tasks or different neurological disorders by incorporating a simple linear classifier which can be interpreted as discriminative filters for different brain patterns. The label information is leveraged to get a consistent and robust solution to the inverse problem.

Related Work

On contrary to forward problem, which consists of modeling the contribution of each voxel to the EEG sensors by solving Maxwell's equation, the inverse problem is ill-posed since the number of interior brain voxels taken into account is far greater than the number of sensors outside the scalp. To precisely estimate the responsible sources of EEG activity from at least several thousands of potential contributing locations which is evenly distributed across the brain requires prior knowledge. Given different neurophysiological assumptions or prior beliefs on the structure of possible source configurations, the goal is to find a unique and stable solution that best explain the signal we observed in the EEG channels. The most commonly used priors in EEG source reconstruction are based on the ℓ_2 norm, leading to what is known as the minimum norm (MN) inverse solver (Hämäläinen and Ilmoniemi 1994). This MN inverse solver leads to a minimum norm estimates (MNE) of the sources. However, ℓ_2-based solvers suffer from several limitations, e.g. the solution inside the brain will be diffuse.

Other assumption priors are also presented with different methodologies, such as, Multiple Signal classifier (MUSIC) and Recursively applied and projected MUSIC (RAP MUSIC) (Mosher and Leahy 1998)(Mosher and Leahy 1999) which adopted spatio-temporal independent topographies (IT) model with recursive subspace projection; low resolution brain electromagnetic tomography (LORETA) (Pascual-Marqui et al. 2002) and standardized LORETA (Pascual-Marqui and others 2002) which enforces spatial smoothness of the source located on neighboring voxels, focal under determined system solution (FOCUSS) (Gorodnitsky, George, and Rao 1995) which combines the advantages of distributed dipole modeling method and linear estimation method by allowing current sources to take arbitrary shape with high resolution; weighted minimum norm-LORETA (WMN-LORETA) (Song, Zhuang, and Wu 2006) which makes compensation to the deeper sources; invariance with respect to the orientation of the coordinate system, and b) a preference for sparsity of the solutions and their spatial derivatives; Focal Vector Field Reconstruction (FVR), which combined sparsity and rotational invariance source reconstruction (Haufe et al. 2008); Mixed Norm Estimates (MxNE), which imposes sparsity over space and smoothness over time using $\ell_{1,2}$-norm regularization (Gramfort, Kowalski, and Hämäläinen 2012), etc.

As summarized above, based on different assumptions, different algorithms solving the inverse problem were proposed, implemented and validated. However, to the best of our knowledge, there is no literature addressing simultaneously estimation of brain sources and distinguishing different sources given different status of the brain. We propose a new supervised formulation of the inverse problem and with efficient algorithms to solve it. The new formulation is composed of two ingredients, source reconstruction and supervised source classification. The contributions of this paper is fourfold, including:

1. First proposed a model with discriminative power to solve EEG inverse problem.
2. First described the EEG inverse problem as an overcomplete dictionary learning problem and show the opportunities of using algorithms from compressive sensing and computer vision community.
3. Proposed revised version of K-SVD algorithm to solve the optimization model good accuracy.
4. Employed the most recently developed high accurate head model rather than approximated head model compared to previous studies.

The structure of the rest paper is as follows: In Section 2, the problem formulation is given. In Section 3, the optimization method is proposed. In Section 4, the numerical experiments and the effectiveness of our proposed framework, conclusions and future work are given in Section 5.

Discriminative Source Reconstruction

In this section, we first briefly review the inverse problem, and then the proposed model in form of discriminative dictionary learning is described, which comprises the source reconstruction term and label guided discriminative term. The motivation of such a discriminative inverse model will be discussed in details.

The Inverse Problem

The electromagnetic field measured by EEG can be described as the following linear model:

$$X = LS + \epsilon \tag{1}$$

where $X \in \mathbb{R}^{N_c \times N_t}$ is the EEG data measured at a set of N_c electrodes for N_t time points, $L \in \mathbb{R}^{N_c \times N_d}$ is the lead field matrix which maps the source signal to sensors on the scalp, each column of L represents the activation pattern of a particular source to the EEG electrodes, $S \in \mathbb{R}^{N_d \times N_t}$ represents the corresponding driving potential in N_d sources locations for all the N_t time points. ϵ is the noise. Generally, an estimate of S can be found by minimizing the following cost function, which is composed of a quadratic error and a regularization term:

$$\arg\min_S \|X - LS\|_F^2 + \lambda\Theta(S) \tag{2}$$

The penalty function $\Theta(S)$ is to discourage unnecessary complicated source configurations and enforces neurophysiologically plausible solutions, and $\|\cdot\|_F$ is the Frobenius Norm. The regularization term take the form of ℓ_2, ℓ_1 or mixed norm, spatially smooth formulation as in LORETA estimation or spatially sparse formulation with least absolute shrinkage and selection operator estimate. For example, to restrict the total number of activated sources less than T, the following ℓ_0-Norm formulation can be used:

$$\arg\min_S \|X - LS\|_F^2 \;\; s.t. \; \|s_i\|_0 \leqslant T, \tag{3}$$

As is well know that ℓ_0-norm is the best intuitive formulation to restrict number of activated sources, almost of neuroscience researchers, if not all, for solving EEG inverse problem use approximated norm such as ℓ_1 to avoid the solution being NP-hard. For the ith time point, the ℓ_1 regularized formulation is given below:

$$s_i = s^*(x_i, L) = \arg\min_x \|x_i - Ls_i\|_2^2 + \gamma\|s_i\|_1 \tag{4}$$

The ill-posed problem of Eqn.1 originates from the fact that L is a matrix with column number far much greater than the row number. We view the L matrix as a dictionary, and each column in L is an atom of the dictionary. Given the EEG recordings at a time point, which is denoted as ith column x_i of X matrix, we want to represent the signal with minimum error by trying to find the best linear representation from activation patterns (atoms) in the over-complete dictionary L. The solution s_i is the sparse coding for the x_i in the dictionary L, the non-zero entries in s_i corresponding to a column in the dictionary matrix L represent the activated regions inside the brain.

Dictionary Learning Fused with Label Information

As the brain has different emotion/task related states, classification of different status is important in Brain Computer Interface(BCI) application, also it helps us understand the mechanism how brain is functioning. Mapping the EEG to the source give us a direct sense of how the sources are evoked and evolved in different states. The motivation of the supervised inverse problem formulation can be explained using a simple demonstrative example as it's illustrates in Fig.1. The electrical potential mentioned at x_1 can be formulated as $x_1 = a_1s_1 + a_2s_2 + a_3s_3 + \epsilon$ and the same case for x_2 channels, where $a_i(i = 1, 2, 3)$ describe the conductivity from for electricity traveling from site s_i to channel x_1. According previous studies (Raichle 2006), only a small portion of electrical energy are task related and it's reasonable to assume that s_1 represents to the non-task resting state source and contribute most of the potential measured in sensors. Assume s_2 is activated when performing task A and s_3 is related to task B. Under the condition of low signal noise ratio (SNR), the reconstructed source tends to be only s_1 without explicitly using the supervising label. Here we leverage the label information explicitly in the hope of successful reconstruction of the discriminative source s_2 and s_3. Here

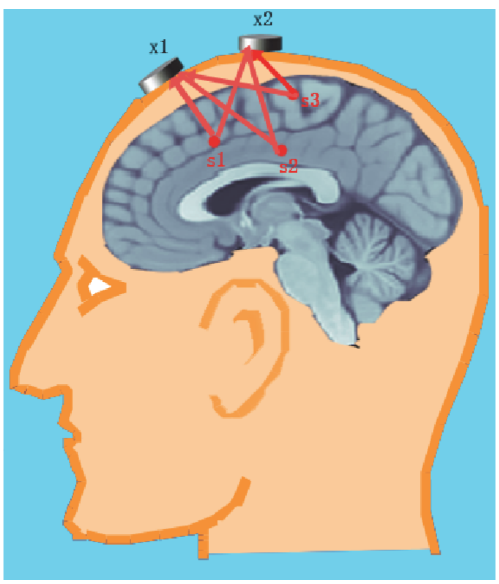

Figure 1: source to electrode

we present a new framework that can infer the source signal guided by the label information. A classification of different brain status based on the sparse coder s_i is obtained by determining its model parameters W, where

$$W = \arg\min_{W} \sum_i \ell\{h_i, f(s_i, W)\} + \lambda \|W\|_F^2 \tag{5}$$

where $\ell\{\cdot\}$ is the loss function for classification accuracy based on the ground truth and classification model $f(\cdot)$, and h_i is the label vector where non-zero entry denotes the corresponding class. Traditional procedure is first solve the pure inverse problem ignoring the supervising label and then train the sparse coding s_i with classification model. Separating the inverse problem and classification problem can be misleading, we argue that since we have the brain status information, it's better to use it as a label to make the inverse solution exhibiting discriminative capability. With this thought and inspired by literatures in computer vision community (Jiang, Lin, and Davis 2013; Yang et al. 2014; Zhang and Li 2010; Pham and Venkatesh 2008), the following sparse discriminant inverse model is given:

$$\begin{aligned}\langle W, S\rangle &= \arg\min_{W,S} \|X - LS\|_F^2 + \beta \sum_i \ell\{h_i, f(s_i, W)\}\\ &+ \lambda \|W\|_F^2 \; s.t. \; \forall i, \;\; \|s_i\|_0 \leqslant T\end{aligned} \tag{6}$$

The first term is the reconstruction error, the second term represents the classification loss, the third term is the regularization of W to avoid over-fitting. This formulation aims to simultaneously learn the sparse coding and the classification model. Using the multi-class classifier $f(\cdot)$ instead of one-against-all classifiers is efficient for classification, by suppressing features sharing among classes and trying to explicitly extract different sparse representation among different classes. In this paper, We focus on an inverse solution with more balanced reconstructive and discriminative power by adding the classification regularization term λ. A summary of our proposed framework is illustrated in Fig.2.

Source Reconstruction Based on Linear Classifier

From Eqn.6, we reduce to the following optimization problem by using a simple linear classifier.

$$\begin{aligned}\langle W, S\rangle &= \arg\min_{W,S} \|X - LS\|_F^2 + \beta \|H - WS\|_F^2\\ &+ \lambda \|W\|_F^2 \; s.t. \; \forall i, \;\; \|s_i\|_0 \leqslant T\end{aligned} \tag{7}$$

Here $H = [h_1, h_2, \ldots, h_N] \in \mathbb{R}^{m \times N_t}$, with each row h_i, $i = 1, \ldots, N_t$ being the label vector corresponding to an EEG signal x_i. In order to solve the optimization problem (7), the K-SVD algorithm and its derivatives can be used. However, our proposed method is different from these previous methods in several aspects, owing to it being tailored to solve the EEG inverse problem.

Optimization with K-SVD Algorithm

For Equation 7, it can be rewritten as

$$\begin{aligned}\langle W, S\rangle &= \arg\min_{W,S} \left\| \begin{pmatrix} X \\ \sqrt{\beta} H \end{pmatrix} - \begin{pmatrix} L \\ \sqrt{\beta} W \end{pmatrix} S \right\|_F^2\\ &+\lambda \|W\|_F^2 \qquad s.t. \; \forall i, \;\; \|s_i\|_0 \leqslant T\end{aligned} \tag{8}$$

Let $X_{new} = (X^t, \sqrt{\beta}W^t)^t$, $L_{new} = (L^t, \sqrt{\beta}W^t)^t$, the optimization of Equation 8 is equivalent to solving the following problem:

$$\begin{aligned}\langle L_{new}, S\rangle &= \arg\min_{L_{new}, S} \|X_{new} - L_{new}S\|_F^2\\ &+ \lambda \|W\|_F^2 \qquad s.t. \; \forall i, \;\; \|s_i\|_0 \leqslant T\end{aligned} \tag{9}$$

In neuroscience community, the lead field matrix is rarely normalized (Grech et al. 2008). We use normalized lead field matrix L to meet the requirement of K-SVD algorithm. It's more important to find an explanatory activation pattern compared to magnitude of the signal as a common practice

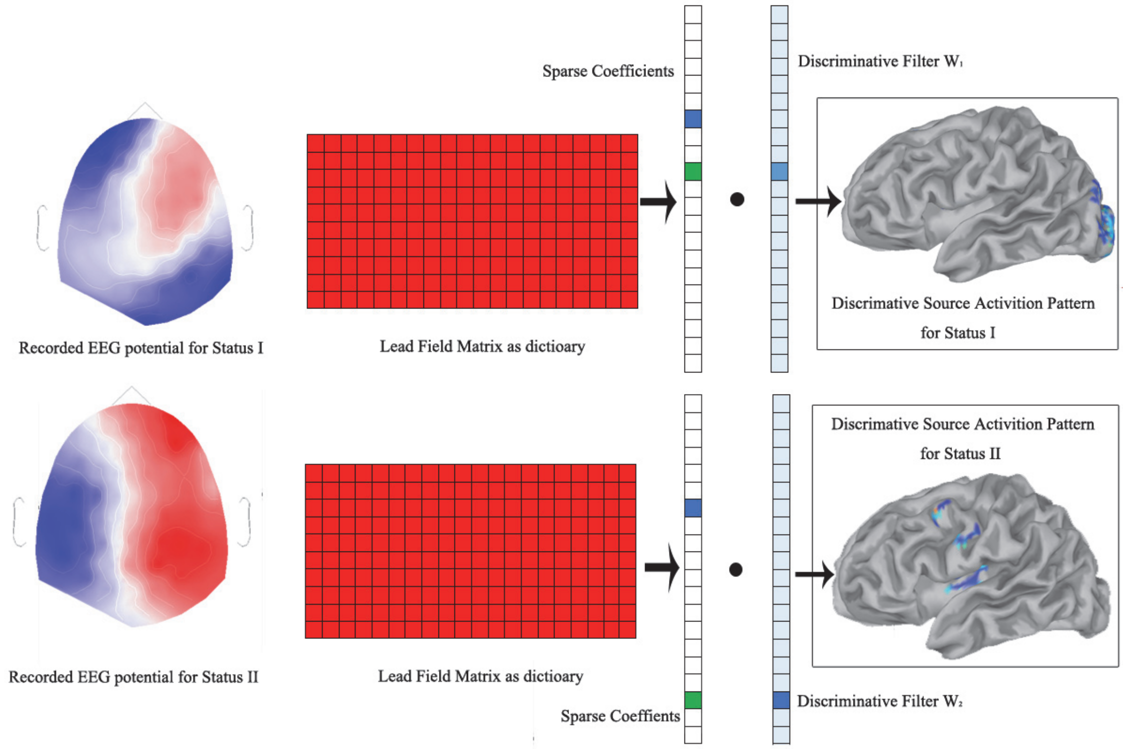

Figure 2: Discriminative Source Reconstruction Framework: The left two topoplots represent the recorded EEG potentials on the scalp for two stimulus status (e.g. finger tapping and comedy video stimulus), and the lead field matrix are represented as overcomplete dictionary, the sparse coefficients are the codes for the source activation location and activation potentials; The sparse coefficients and W matrix are estimated simultaneously. Each row of W matrix is termed as discriminative filter because the Hadamard product of the source code coefficient and the discriminative filter can highlight the corresponding stimulus activated source signal by masking the common background or resting activation signals which is share by other different brain stimulus inputs. The rightmost pictures are exemplary reconstructed discriminative source activation patterns on the cortex

(Haufe 2011). Later we show that the normalization doesn't effect the solution in case of ℓ_0 norm. The normalization is defined as:

$$\begin{aligned} L' &= \{l'_1, l'_2, ..., l'_{N_d}\} \\ &= \{\frac{l_1}{\|l_1\|_2}, \frac{l_2}{\|l_2\|_2}, ..., \frac{l_{N_d}}{\|l_{N_d}\|_2}\} \\ W' &= \{w'_1, w'_2, ..., w'_{N_d}\} \\ &= \{\frac{w_1}{\|l_1\|_2}, \frac{w_2}{\|l_2\|_2}, ..., \frac{w_{N_d}}{\|l_{N_d}\|_2}\} \end{aligned} \tag{10}$$

Suppose x_i is the EEG signal vector and we want to find corresponding source location. x_i is a sparse linear combination of the atoms in L, which can be expressed as:

$$\begin{aligned} x_i &= \sum\nolimits_{m=1}^{N_d} l_m s_i(m) = \sum\nolimits_{m=1}^{N_d} \left(\frac{l_m}{\|l_m\|_2}\right) \left(s_i(m)\|l_m\|_2\right) \\ &= \sum\nolimits_{m=1}^{N_d} l'_m s'_i(m) \end{aligned}$$

Also, the Label matrix h_i can be expressed as

$$\begin{aligned} h_i &= \sum\nolimits_{m=1}^{N_d} w'_m s'_i(m) = \sum\nolimits_{m=1}^{N_d} \left(\frac{w_m}{\|l_m\|_2}\right) \left(s_i(m)\|l_m\|_2\right) \\ &= \sum\nolimits_{m=1}^{N_d} w_m s_i(m) \end{aligned}$$

The sparse coding with or without normalization of L is equivalent in terms of ℓ_0-norm, which is $\|s_i\|_0 = \|s'_i\|_0$, thus the normalization of lead field matrix doesn't effect the reconstruction solution under condition of ℓ_0-norm. As L_{new} is always normalized column-wise, we can drop the regularization penelty term $\|W\|_F$.

$$\begin{aligned} \langle L'_{new}, S\rangle &= \arg\min_{L'_{new}, S} \|X_{new} - L'_{new} S\|_F^2 \\ & s.t.\ \forall i,\quad \|s_i\|_0 \leqslant T \end{aligned} \tag{11}$$

For similarity, we omit the apostrophe ($'$) notation when there is no confusion. When fixing S, solving L matrix can be regarded as solving a simple regression problem:

$$\hat{L} = \arg\min_L \|X - LS\|_F^2, \tag{12}$$

where $\hat{L} = XS^T(SS^T)^{-1}$. The computational complexity of $XS^T(SS^T)^{-1}$ is $O(n^3)$, it is advisable to solve it using K-SVD by updating the dictionary atom-by-atom. This optimization problem of Eqn.11 is exactly what K-SVD algorithm (Aharon, Elad, and Bruckstein 2006) solves and the only difference is that the upper L part of dictionary L_{new}

will not be updated. We adopt the procedure in the original K-SVD algorithm.

Following K-SVD, denote l'_k as the kth column in the L'_{new}, and s_k is the corresponding kth row in S. The second term $L'_{new}S$ can be decomposed into the following formulation:

$$L_{new}S = \sum_{k=1}^{N_d} l_k * s_k$$

Let $E_k = (X - \sum_{j \neq k}(l_j * s_j))$, representing the error without using the atom l_k, the main idea of K-SVD is to update each atom in the dictionary sequentially to the projected direction that most reduces the error. Let $\tilde{s}_R^k$ and $\tilde{E}_k$ denote the result of discarding the zero entries in x_R^k and E_k, respectively. As a result, l_k and $\tilde{s}_R^k$ can be computed using

$$\langle l_k, \tilde{s}_R^k \rangle = \arg\min_{l_k, \tilde{s}_R^k} \left\| \tilde{E}_k - l_k \tilde{s}_R^k \right\|_F^2 \qquad (13)$$

The above optimization problem can be easily solved by employing an SVD composition of $\tilde{E}_k$, namely, $U\Sigma V^t = SVD(\tilde{E}_k)$, and using the SVD result and update the l_k and $\tilde{s}_R^k$ with $l_k = U(:,1)$, $\tilde{s}_R^k = \Sigma(1,1)V(1,:)$. $U(:,1)$ denotes the first column of matrix U, and $V(1,:)$ is the first row of V, $\Sigma(1,1)$ is the first diagonal value of Σ. The upper part of the L_{new} matrix will not be updated, and only the lower part composed of W matrix is updated. The detailed algorithm is given in the following algorithm 1 with matlab indexing notation.

Algorithm 1 Revised DK-SVD algorithm

INPUT: Lead field matrix L, preprocessed EEG signal matrix X, relative controlling scalar β, label matrix H
OUTPUT: classification matrix W, EEG source matrix S
Initialization: Using K-SVD initialization described in Ref.(Aharon, Elad, and Bruckstein 2006)

set $m = 1$
while not converged **do**
  Solve the following sparse coding problem using matching pursuit algorithm for $i = 1, 2, \ldots, N$:
  $\min_{s_i} \|x_i - Ls_i\|_2^2 \quad s.t. \quad \|s_i\|_0 \leqslant T$
  while i is not equal to N_d **do**
    (1) Compute the representation error without atom l_i, $E_i = (X - \sum_{j \neq i}(l_j * s_j))$
    (2) Extract the nonzero entries of s_i and truncate the E_i to E_i^P accordingly.
    (3) SVD decomposition for E_i^P as $E_i^P = U\Lambda V$
    (4) Update l_i and s_i^T:
      $l_i(N_c+1:end) \longleftarrow U(:,1)(N_c+1:end)$,
      $\tilde{s}_R^i \longleftarrow \Sigma(1,1)V(1,:)$.
    (5) Update index $i \longleftarrow i+1$;
  end while
  $m \longleftarrow m+1$
end while

Numerical Experiments

Numerical simulations were conducted given different SNR. We compared our proposed framework with two different baseline methods. Computation time (in s) combined with three different accuracy criteria and two solution quality measurements were used as gauges. The first baseline method is formulated as ℓ_1 sparse representation ($\ell_1 SR$) and solved with Efficient Projections onto the ℓ1-Ball (EP-ℓ1B) (Liu and Ye 2009), and the second baseline method are the recently developed MxNE (Gramfort, Kowalski, and Hämäläinen 2012).

Head and Source Model

Head model is a volume conductor model which is used to describe the flow of electric current in the head. Usually, the brain model was built in 3 steps, (1) collect the MRI images; (2) tissue segmentations (3) Mesh generation and assignment of conductivities for different tissues. We used a newly developed lead field model called ICBM-NY or "New York Head" (Huang, Parra, and Haufe 2016) which is based on highly detailed standardized finite element model (FEM) of the non-linear averaged anatomical template-ICBM152. The brain tissue segmentation is divided into 6 tissue type (scalp, skull, cerebro-spinal fluid(CSF), gray matter, white matter and air cavities) with native MRI resolution of $0.5mm^3$. We imposed biological pink noise and EEG sensor measurement noise to test accuracy of different algorithm with SNR from 1.5 to 0.5. The pink noise is generated from 100 sources located randomly inside the brain. The sensor measurement noise is directly added to the measured EEG signal. Based on the criteria given by Baillet and Garnero (Baillet and Garnero 1997), the spatial and temporal accuracy should be at least better than 5 mm and 5 ms respectively. Also, we divided the brain into 8 region of interests (ROI) called Right Anterior Inferior (RAI), Right Anterior Superior (RAS), Right Posterior inferior (RPI), Right Posterior Superior (RPS), Left Anterior Inferior (LAI), Left Anterior Superior (LAS), Left Posterior inferior (LPI), Left Posterior Superior (LPS). The simulated source dynamics is generated linear autoregessive (AR) model with order of 6.

Reconstruction Accuracy

We used three different accuracy criteria to measure the reconstructed source accuracy. The first one is perfect reconstruction accuracy (PRA), which compare the calculated source location and the exact ground true. The second measurement is to use Baillet-Garnero's reconstruction accuracy (BRA) criteria (Baillet and Garnero 1997). The third measurement is to use the criteria proposed in (Haufe and Ewald 2016) denoted as Haufe reconstruction accuracy (HRA), which is to measure whether the reconstructed source is located in the ROI. To make the solution be more informative, a sparse solution is always preferred for its interpretability. The averaged number non-zero entries (NZE) in the solution is also included to measure the sparsity. $\|X - LS\|_F$ is the reconstruction error (RE).

Detailed numerical experiments performance are listed in Table 1-3. From the tables 1-3, our proposed framework outperforms the other two baseline methods for less computation time, more accurate based on 3 different criteria, more sparse solution, better reconstruction error and also with the capability of extracting discriminative sources. For the other two baseline methods, the results are based on optimized regularization parameters with a balanced trade-off between the sparsity and reconstruction error.

We found that even when the SNR is high, the traditional algorithm can only find the resting state or background noise which is set to be much larger, missing unanimously the discriminative source, meaning that the task related source can't be estimated correctly. We are not surprised that the baseline method achieved bad performance since some of the atoms are highly coherent each other and the sparsity property is hard to achieve, the ℓ_1 or ℓ_2 norm tend to assign non-zero value to those high coherent atoms simultaneously.

We demonstrated an exemplary results in Fig.3 and Fig.4. Fig.3 demonstrates the reconstructed source and calculated discriminative filters W, the discriminative filters suppress common source shared by different status while extracting and magnifying distinguished ones. Fig.4 illustrates the EEG potentials topoplots on the scalp before and after the application of our method, distinctive source activation patterns can be clearly revealed.

Table 1: Performance comparison at SNR=1.2

Method	Time	PRA	BRA	HRA	NZE	RE
DKSVD	**2.04**	**0.77**	**0.81**	**0.93**	**4.00**	**0.91**
$\ell_1 SR$	10.9	0.33	0.37	0.52	308	118
MxNE	10.6	0.22	0.25	0.50	449	85.6

Table 2: Performance comparison at SNR=0.8

Method	Time	PRA	BRA	HRA	NZE	RE
DKSVD	**2.26**	**0.77**	**0.79**	**0.91**	**4.00**	**1.86**
$\ell_1 SR$	10.9	0.31	0.35	0.50	316	117
MxNE	12.9	0.32	0.32	0.53	418	95.8

Table 3: Performance comparison at SNR=0.5

Method	Time	PRA	BRA	HRA	NZE	RE
DKSVD	**2.38**	**0.63**	**0.64**	**0.68**	**4.00**	**12.4**
$\ell_1 SR$	11.4	0.30	0.32	0.50	410	111
MxNE	12.0	0.25	0.28	0.50	507	89.6

Conclusion and Future Work

We aim to reconstruct discriminative sources given different brain status. A label guided dictionary learning formulation is given for the first time with ℓ_0-norm and is solved using our revised version of DK-SVD algorithm. Through numerical simulations, we showed that in terms of accuracy and speed, our method is better than the ℓ_1 or ℓ_2 related ones. The reason is high coherence of lead field matrix and sparsity constraints is easy to fail. The classification component trained a W matrix with each row corresponding certain type of brain status, which is physically meaningful, we termed as discriminative filter. Our proposed framework can achieve satisfactory result compared to traditional methods and can be extended to more specific priors such as spatially smoothness requirement or depth compensation requirement.

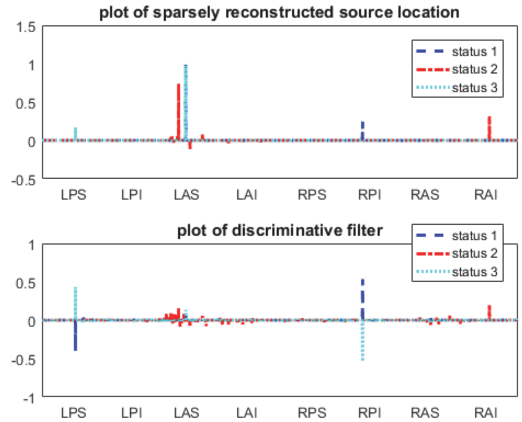

Figure 3: Sparse coding and discriminative filter for 3 different brain status: the common resting state signal is attenuated by W (as is illustrated in LAS region) and the reconstructed discriminative sources are extracted and magnified by the corresponding rows of W. The upper subfigure is the reconstructed source S and discriminative filter W is given in the lower subfigure.

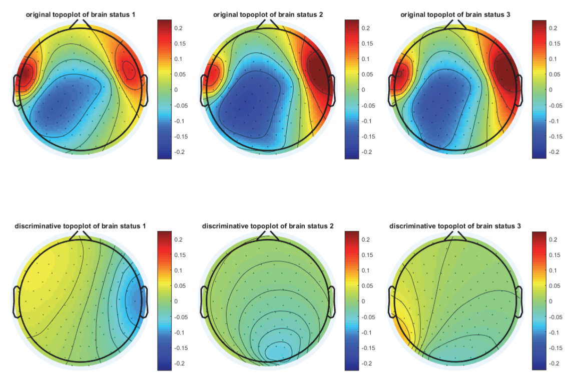

Figure 4: Discriminative filtered topoplots for 3 different brain status: the top 3 topoplots is corresponding to 3 different brain tasks with high common source from spontaneous activity or resting state potentials, it's very hard to distinguish them. Below is the topoplots after we applied our methodology to extract the discriminative expression for different brain tasks. The lower topoplots are constructed by applying the discriminative source to the forward model.

References

Aharon, M.; Elad, M.; and Bruckstein, A. 2006. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing* 54(11):4311–4322.

Baillet, S., and Garnero, L. 1997. A bayesian approach to introducing anatomo-functional priors in the eeg/meg inverse problem. *IEEE Transactions on Biomedical Engineering* 44(5):374–385.

Baillet, S.; Mosher, J. C.; and Leahy, R. M. 2001. Electromagnetic brain mapping. *IEEE Signal processing magazine* 18(6):14–30.

Barabási, A.-L., and Albert, R. 1999. Emergence of scaling in random networks. *Science* 286(5439):509–512.

Gorodnitsky, I. F.; George, J. S.; and Rao, B. D. 1995. Neuromagnetic source imaging with FOCUSS: a recursive weighted minimum norm algorithm. *Electroencephalography and clinical Neurophysiology* 95(4):231–251.

Gramfort, A.; Kowalski, M.; and Hämäläinen, M. 2012. Mixed-norm estimates for the M/EEG inverse problem using accelerated gradient methods. *Physics in Medicine and Biology* 57(7):1937.

Grech, R.; Cassar, T.; Muscat, J.; Camilleri, K. P.; Fabri, S. G.; Zervakis, M.; Xanthopoulos, P.; Sakkalis, V.; and Vanrumste, B. 2008. Review on solving the inverse problem in eeg source analysis. *Journal of NeuroEngineering and Rehabilitation* 5(1):25.

Guan, Z.-H.; Liu, F.; Li, J.; and Wang, Y.-W. 2012. Chaotification of complex networks with impulsive control. *Chaos: An Interdisciplinary Journal of Nonlinear Science* 22(2):023137.

Hämäläinen, M. S., and Ilmoniemi, R. J. 1994. Interpreting magnetic fields of the brain: minimum norm estimates. *Medical & Biological Engineering Computing* 32(1):35–42.

Haufe, S., and Ewald, A. 2016. A simulation framework for benchmarking EEG-based brain connectivity estimation methodologies. *Brain topography* 1–18.

Haufe, S.; Nikulin, V. V.; Ziehe, A.; Mller, K.-R.; and Nolte, G. 2008. Combining sparsity and rotational invariance in EEG/MEG source reconstruction. *NeuroImage* 42(2):726 – 738.

Haufe, S.; Nikulin, V. V.; Mller, K.-R.; and Nolte, G. 2013. A critical assessment of connectivity measures for EEG data: A simulation study. *NeuroImage* 64:120 – 133.

Haufe, S. 2011. *Towards EEG source connectivity analysis*. Ph.D. Dissertation, Technical University of Berlin, 10623 Berlin, Germany.

Hipp, J. F.; Hawellek, D. J.; Corbetta, M.; Siegel, M.; and Engel, A. K. 2012. Large-scale cortical correlation structure of spontaneous oscillatory activity. *Nature Neuroscience* 15(6):884–890.

Huang, Y.; Parra, L. C.; and Haufe, S. 2016. The new york heada precise standardized volume conductor model for EEG source localization and tES targeting. *NeuroImage* 140:150 – 162.

Jiang, Z.; Lin, Z.; and Davis, L. S. 2013. Label consistent K-SVD: Learning a discriminative dictionary for recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35(11):2651–2664.

Liu, J., and Ye, J. 2009. Efficient euclidean projections in linear time. In *Proceedings of the 26th Annual International Conference on Machine Learning*, 657–664.

Liu, F.; Xiang, W.; Wang, S.; and Lega, B. 2016. Prediction of seizure spread network via sparse representations of overcomplete dictionaries. In *International Conference on Brain and Health Informatics*, 262–273.

Ma, G.; He, L.; Cao, B.; Zhang, J.; Philip, S. Y.; and Ragin, A. B. 2016. Multi-graph clustering based on interior-node topology with applications to brain networks. In *Joint European Conference on MLKDD*, 476–492.

Mosher, J. C., and Leahy, R. M. 1998. Recursive MUSIC: a framework for EEG and MEG source localization. *IEEE Transactions on Biomedical Engineering* 45(11):1342–1354.

Mosher, J. C., and Leahy, R. M. 1999. Source localization using recursively applied and projected (RAP) MUSIC. *IEEE Transactions on Signal Processing* 47(2):332–340.

Newman, M. E. 2003. The structure and function of complex networks. *SIAM review* 45(2):167–256.

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

Pascual-Marqui, R. D.; Esslen, M.; Kochi, K.; Lehmann, D.; et al. 2002. Functional imaging with low-resolution brain electromagnetic tomography (loreta): a review. *Methods Find Exp Clin Pharmacol* 24(Suppl C):91–95.

Pham, D. S., and Venkatesh, S. 2008. Joint learning and dictionary construction for pattern recognition. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, 1–8.

Raichle, M. E. 2006. The brain's dark energy. *Science* 314(5803):1249–1250.

Sockeel, S.; Schwartz, D.; Pélégrini-Issac, M.; and Benali, H. 2016. Large-scale functional networks identified from resting-state EEG using spatial ICA. *PloS one* 11(1):e0146845.

Song, C.; Zhuang, T.; and Wu, Q. 2006. Hybrid weighted minimum norm method a new method based LORETA to solve EEG inverse problem. In *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, 1079–1082.

Watts, D. J., and Strogatz, S. H. 1998. Collective dynamics of small-worldnetworks. *Nature* 393(6684):440–442.

Yang, M.; Zhang, L.; Feng, X.; and Zhang, D. 2014. Sparse representation based fisher discrimination dictionary learning for image classification. *International Journal of Computer Vision* 109(3):209–232.

Zhang, Q., and Li, B. 2010. Discriminative K-SVD for dictionary learning in face recognition. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2691–2698.