

Supervised Discriminative EEG Brain Source Imaging with Graph Regularization

Feng Liu¹(✉), Rahilsadat Hosseini¹, Jay Rosenberger¹, Shouyi Wang¹,
and Jianzhong Su²

¹ Department of Industrial, Manufacturing and Systems Engineering,
University of Texas at Arlington, Arlington, TX, USA
feng.liu@mavs.uta.edu

² Department of Mathematics, University of Texas at Arlington, Arlington, TX, USA

Abstract. As Electroencephalography (EEG) is a non-invasive brain imaging technique that records the electric field on the scalp instead of direct measuring activities of brain voxels on the cortex, many approaches were proposed to estimate the activated sources due to its significance in neuroscience research and clinical applications. However, since most part of the brain activity is composed of the spontaneous neural activities or non-task related activations, true task relevant activation sources can be very challenging to be discovered given strong background signals. For decades, the EEG source imaging problem was solved in an unsupervised way without taking into consideration the label information that representing different brain states (e.g. happiness, sadness, and surprise). A novel model for solving EEG inverse problem called Graph Regularized Discriminative Source Imaging (GRDSI) was proposed, which aims to explicitly extract the discriminative sources by implicitly coding the label information into the graph regularization term. The proposed model is capable of estimating the discriminative brain sources under different brain states and encouraging intra-class consistency. Simulation results show the effectiveness of our proposed framework in retrieving the discriminative sources.

Keywords: Inverse problem · Graph regularization · EEG source imaging · Sparse representation

1 Introduction

To infer the activated brain sources from the recorded EEG data is called *inverse problem*. Precise localization of neuronal firing pattern inside the brain can offer an insightful understanding of how the brain is functioning under certain cognitive and motion tasks. We also argue that source reconstruction or solving the inverse problem is the first and primary step for connectivity analysis of the brain, and precise inference of time course of brain sources is required in order to build the brain connectivity network. The latter step is to analyze the brain

network using complex networks [4, 14, 17] characteristics measurement, as we saw a shift in neuroscience community from traditional “segregation” perspective to “integration” perspective where the functional and effective connectivity between different regions of brains are intensively studied [8, 13] in recent years.

In order to solve the ill-posed inverse problem, different priors or assumptions have to be imposed to obtain a unique solution. The most traditionally used priors are based on minimum power, leading to what is known as the minimum norm estimate (MNE) inverse solver [6], or minimum magnitude, termed as minimum current estimate (MCE) [16], leading to a least absolute shrinkage and selection operator (LASSO) formulation. Other assumptions or priors are presented with different inverse algorithms, such as standardized low-resolution brain electromagnetic tomography sLORETA [15], which enforces spatial smoothness of the source located on neighboring voxels; Bernoulli-Laplace priors, which introduced $\ell_0 + \ell_1$ norm in a Bayesian framework [2]; Mixed Norm Estimates (MxNE), which imposes sparsity over space and smoothness over time using $\ell_{1,2}$ -norm regularization [3]; graph Fractional-Order Total Variation (gFOTV) [11], which impose sparsity of the spatial fractional derivatives so that it locates source peaks by providing the freedom of choosing smoothness order.

As summarized above, numerous algorithms that were based on different source configuration assumptions or prior knowledge were presented to solve the inverse problem. Traditional algorithms solve the EEG inverse problem independently for different brain states without leveraging the label information, that will make it hard to compare the reconstructed sources under different brain states due to its low SNR (Signal-to-Noise Ratio) of the EEG signal. To the best of our knowledge, few researchers come up with a model that can integrate EEG inverse problem with label information (e.g. happiness, sadness, and surprise) to find task related discriminative sources except our previous work [12]. To explicitly extract factual sources and eliminate the spurious ones, we proposed the graph regularized version of discriminative source reconstruction that has the capability of promoting intra-class consistency, and we tested it on synthetic data and illustrated its effectiveness in discovering task related sources.

2 The Inverse Problem

Under the quasi-static approximation of Maxwell’s equations, the EEG signal measurements X can be described as the following linear function of current sources S :

$$X = LS + E, \quad (1)$$

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 a wide matrix called lead field matrix that maps the brain source signal to sensors on the scalp, each column of L represents the activation pattern of a particular brain source to the EEG electrodes, $S \in \mathbb{R}^{N_d \times N_t}$ represents the corresponding neural electric field in N_d source locations for N_t time points. $E \in \mathbb{R}^{N_c \times N_t}$ is additive noise. An estimate of S can be found

by minimizing the following cost function, which is composed of a data fidelity term and a regularization term:

$$\arg \min_S \|X - LS\|_F^2 + \lambda\Theta(S), \quad (2)$$

where $\|\cdot\|_F$ is the Frobenius Norm. The regularization term $\Theta(S)$ can be used to guarantee smooth source configurations temporally or spatially and enforces neurophysiologically plausible solutions or to guarantee sparsity in source solution. For example, to restrict the total number of activated voxels to be less than or equal to k , the constraint $\|s_i\|_0 \leq k$ can be used. Even though ℓ_0 -norm is the best intuitive formulation to restrict number of activated sources, it's a common practice to use approximated norm such as ℓ_1 to avoid the problem being NP-hard when solving EEG inverse problem. For the i th time point, the ℓ_1 regularized formulation is given below:

$$\langle s_i \rangle = \arg \min_{s_i} \|x_i - Ls_i\|_2^2 + \gamma\|s_i\|_1. \quad (3)$$

Given the EEG recordings at a time point, which is denoted as i th 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 [12]. 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 [12].

3 Proposed Framework

3.1 Graph Regularized Discriminative Source Imaging

Due to the fact that EEG signal is non-stationary and typically the SNR is very low, it's important to get consistent inverse solution under the same brain status and eliminate the spurious sources that are usually not consistent within the same class. Inspired by the successful applications of graph regularization in computer vision community [1, 5], the proposed model of retrieving task related discriminative source is presented, which is termed as Graph Regularized Discriminative Source Imaging (GRDSI), and comprises the data fidelity term and label guided graph regularization term:

$$\langle S \rangle = \arg \min_S \|X - LS\|_F^2 + \gamma\|S\|_{1,1} + \frac{\beta}{2} \sum_{i,j=1}^N \|s_i - s_j\|_2^2 M_{ij}, \quad (4)$$

where the first term is the fidelity term, the second term is the cost of sparse coding, $\|\cdot\|_{1,1}$ is the ℓ_1 norm notation for a matrix, equal to the sum of the absolute value of all elements in a matrix. The third term is the graph regularization term that requires all the sparse coder within the same category remains similar pattern while making the sparse representation for different class distinct from each other. The definition of M matrix can be written as:

$$M_{ij} = \begin{cases} 1, & \text{if } (s_i, s_j) \text{ belong to the same class} \\ 0, & \text{if } (s_i, s_j) \text{ belong to different classes} \end{cases}$$

The goal of this formulation is to find discriminative sources while maintaining the robustness of in-class reconstructed sources.

Remarks on design of M matrix

When (s_i, s_j) belong to the same class, design the value of M_{ij} to be positive will add penalty when the intrinsic geometry (s_i, s_j) is different, thus promoting intra-class consistency of the source and reduce the spurious sources estimated at each time point.

Define D as a diagonal matrix whose entries are column or row sums of the symmetric matrix M , $D_{ii} = \sum_j M_{ij}$, define $G = D - M$, G is called graph Laplacian [1], The third term of Eq. 4 can be further derived in the following way:

$$\sum_{i,j=1}^N \|s_i - s_j\|_2^2 M_{ij} = \sum_{i,j=1}^N (s_i^T s_i + s_j^T s_j - 2s_i^T s_j) M_{ij} = 2tr(SGS^T) \quad (5)$$

As a result, Eq. 4 is further derived as:

$$\langle S \rangle = \arg \min_S \|X - LS\|_F^2 + \gamma \|S\|_{1,1} + \beta (Tr(SGS^T)) \quad (6)$$

Equation 6 can be efficiently solved using feature-sign search algorithm [1,10].

3.2 Common Sources Decomposition with Voting Orthogonal Matching Pursue (VOMP)

Under the assumption of strong common spontaneous source activation pattern, the contribution of discriminative sources to the EEG recorded data is relatively small, making the solution space for different classes highly correlated and difficult to find discriminative sources. As a result, the convex hull spanned by all the source configuration is limited to a tiny portion of the space [18]. In order to address that, we use the idea of “cross-and-bouquet” model [18] and come up with a useful step that is to decompose of X to find the common sources shared by different classes. The Voting Orthogonal Matching Pursue (VOMP) is proposed and described in Algorithm 1. The aim is to recover the common sources across all classes. The core part of VOMP is Orthogonal Matching Pursue (OMP) which is a very efficient algorithm. After the decomposition of common source, its contribution to the EEG data X is also removed. The new EEG data after removal of the common source is written as $X_{new} = X - LS_c$.

Based on the discussion above, the proposed framework to solve Problem 6 is summarized in Algorithm 2 and illustrated in Fig. 1.

Algorithm 1. Decomposition of Non-discriminative Sources with VOMP

INPUT: Lead field matrix L , EEG data X , maximum number of common sources T_{max} , minimum voting acceptance threshold p

OUTPUT: S_c , result of removed common sources X_{new}

Initialization: $T \leftarrow 1$, $\Omega = \emptyset$, $R = X$, $R_{new} = X$, $S' = 0$

```

while Stopping criteria is not met do
  for  $i \in 1, \dots, N_t$  do
     $s_i \leftarrow \text{OMP}(L, x_i, 1)$ 
     $q_i \leftarrow$  nonzero index of  $s_i$ 
  end for
   $q_{best} \leftarrow$  most frequent  $q_i$ 
  if  $T = T_{max}$  or frequency of  $f(q_{best}) < p$  then
    break;
  else
     $\Omega \leftarrow \Omega \cup q_{best}$ ;  $L' = (L_{:,i} | i \in \Omega)$ ;  $S' \leftarrow \text{pinv}(L')X$ ;  $S' \leftarrow \text{mean}(S')$ ;  $R_{new} \leftarrow X - L'S'$ 
  end if
  for  $k \in 1, \dots, C$  do
     $R_{new}^k = \{R_{new}(i) | i \in \text{class } k\}$ ;
     $R^k = \{R(i) | i \in \text{class } k\}$ 
  end for
  if  $\|R_{new}^k\| < \|R^k\|$  for  $k \in 1, \dots, C$  then
    continue;
  else break;
  end if
   $T \leftarrow T + 1$ ;  $R \leftarrow R_{new}$ 
end while
 $X_{new} = R_{new}$ ;  $S_c = S'$ 
return  $S_c, X_{new}$ 

```

4 Numerical Results

We used a recently developed realistic head model called ICBM-NY or “New York Head” [9]. The dimension of lead field matrix we are using is 108×2004 , representing 108 channels and 2004 voxels. We also assume that source orientation is perpendicular to the cortex surface. In each simulation, noises originate from sensor level and cortex voxel level both contributed to the recorded EEG data. The SNR is calculated as $SNR = 20 \log_{10} \frac{\|S\|_2}{\|N\|_2}$. We show the effectiveness of the graph regularization term in reconstructing the discriminative sources by comparing it with the other eight benchmark algorithms, including ElasticNet, Homotopy, DALM, PDIPA, L1LS, FISTA, sLORETA, MNE. The former 6 algorithms are compared in image reconstruction applications and can be referred to Reference [19] for details.

We designed the spontaneous common sources with a magnitude of 0.8 with standard deviation to be 0.1 and task related discriminative source with a magnitude of 0.2 with a standard deviation of 0.05 located in different Region Of Interest (ROI)s from the common sources. The ROI we used here are defined in Reference [7]. We sampled 200 time points for each class and did the experiment 5 times to get the average accuracy of the reconstructed source. For the GRDSI parameter, we set β to be 0.05 and α to be 0.06; The noise matrix is designed to affect the EEG recording together with the true source signal. For each time point, 3 random voxels are corrupted randomly with the average value being 0.2,

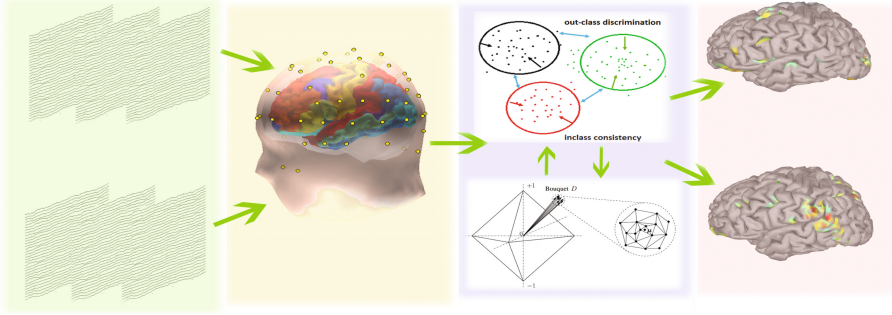


Fig. 1. Procedures of our framework: After gathering labeled EEG recorded data, the brain model is constructed using finite element method (BEM) based on MRI images, the VOMP algorithm is used to decompose the primary common source starting with a high minimum voting percentage, and then solve it using feature-sign search algorithm, the last step is to map discriminative sources to the cortex.

Algorithm 2. Proposed framework of solving Problem 6

INPUT: Lead field matrix L , preprocessed EEG signal matrix X , label matrix H

OUTPUT: Discriminative source S_d

Initialization: $T \leftarrow 1, \Omega = \emptyset, R = X, R_{new} = X, S' = 0$

while Stopping criteria not met **do**

(1) Using VOMP algorithm for common source decomposition;

(2) Solve the following sparse coding problem for $\langle s(i) \rangle = \arg \min_{s(i)} L(s_i) +$

$\gamma \|s_i\|_1$

using the feature-sign search algorithm [10];

(3) Adjust the voting threshold p ;

end while

0.4, 0.6 and variance being 0.05 based on different SNR design. All computations were conducted on a 64-bit Linux workstation with 3.00 GHz i7-5960x CPU and memory of 64 GB.

The reconstruction performance of the proposed method as well as the benchmark methods based on 150 experiments are summarized in Table 1. All of the values in Table 1, except the Time column (in seconds) represents distance in (mm) from ground true source to the reconstructed source calculated from the shortest path along cortex surfaces. PSE represents primary source error, which is the distance of reconstructed primary source to the ground truth primary source. PSE measures the capability of each algorithm to reconstruct the common sources. When the reconstructed location is in a different hemisphere from the ground truth, there is no path connecting those two voxels, so we mark the distance to be 250 mm. EC1 represents error for class 1, which is the distance of the reconstructed discriminative source to the ground truth. EC2 and EC3 are similarly defined.

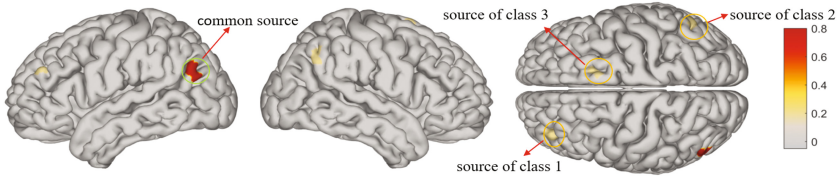


Fig. 2. Ground truth for all 3 classes

Table 1. Reconstruction accuracy summary

Methods	SNR = 10					SNR = 22				
	Time	PSE	EC1	EC2	EC3	Time	PSE	EC1	EC2	EC3
ElasticNet	0.001	43.4	142.3	159.6	159.2	0.001	8.87	172.5	195.0	13.0
Homotopy	0.12	3.43	53.2	42.5	40.8	0.09	0	0.28	0.70	8.00
DALM	0.07	4.59	53.0	43.1	39.6	0.08	0	0.28	1.73	7.98
PDIPA	0.29	3.43	53.4	45.0	40.4	0.26	0	0.28	0.63	7.98
LILS	3.89	0.69	51.6	67.4	37.1	3.92	0.069	0	0	4.36
FISTA	0.95	0.63	61.0	95.2	47.6	0.96	40.1	66.1	73.5	54.5
sLORETA	0.015	10.2	131.7	178.2	142.8	0.02	2.62	194.1	164.2	123.5
MNE	$3e-5$	29.3	131.8	157.7	131.7	$3e-5$	4.30	119.8	136.2	113.5
GRDSI (Proposed)	0.15	1.85	14.4	4.13	3.67	0.10	0	0	0	2.12

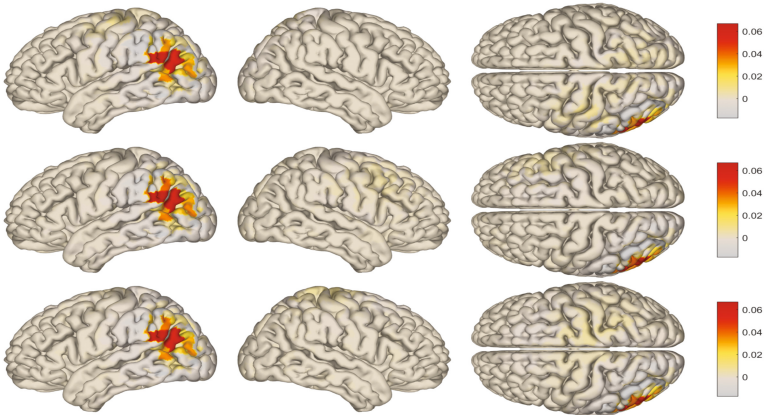


Fig. 3. MNE solution: The above row is the MNE solution for class 1; Class 2 and class 3 is illustrated in the middle and bottom row. The solution MNE gives is not sparse, with too many spurious sources of small magnitude.

To illustrate the effect of the proposed framework, the ground truth of the activated pattern is given in Fig. 2, with the reconstructed source by MNE, sLORETA and our method given in Figs. 3, 4 and 5. We can see from Table 1 and the Figs. 3, 4 and 5 that when the SNR is large, all the algorithms

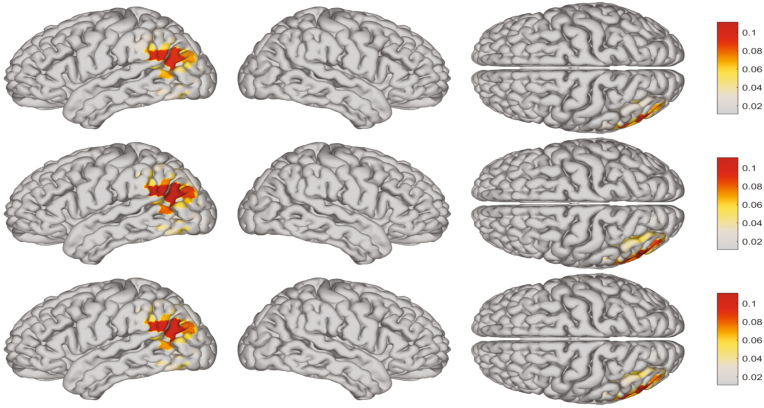


Fig. 4. sLORETA inverse solution: sLORETA can successfully reconstruct the primary source, however the secondary source is not successfully reconstructed. Compared to the solution of MNE, sLORETA can suppress spurious sources with small magnitude.

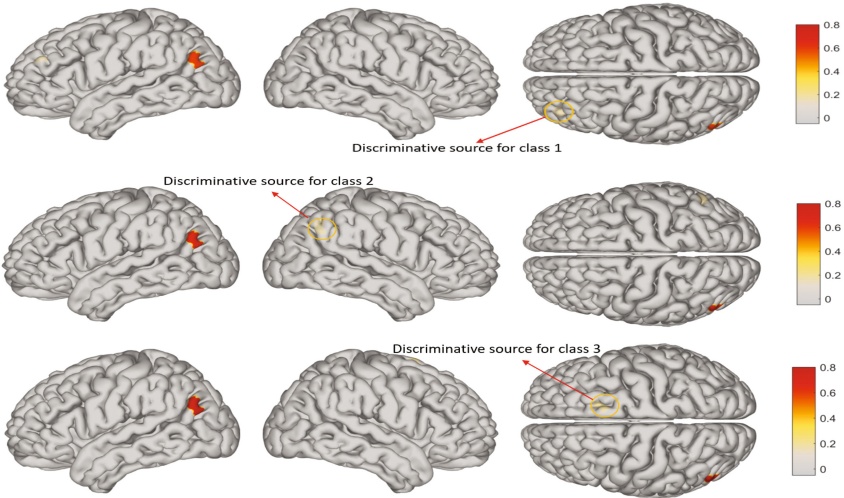


Fig. 5. GRDSI reconstructed source: The reconstruction solutions for 3 classes are given in each row. The discriminative source can be successfully reconstructed.

performs well in reconstructing primary source, as for the discriminative sources for different classes, our method can achieve almost perfect reconstruction. All other algorithms' performances are also acceptable when SNR is large, except for sLORETA, MNE and ElasticNet. When we increase the noise, all of the algorithms can still achieve high accuracy in finding the primary source. For the discriminative source, our algorithm performs much better. We also validated that, to solve a pure ℓ_1 EEG inverse problem, the Homotopy algorithm performs

better in most cases than other algorithms in the EEG inverse problem, which is in line with Reference [19].

5 Conclusion

In this paper, we proposed to use label information to retrieve discriminative sources corresponding to different brain status. A graph regularized EEG inverse formulation called GRDSI that implicitly uses the label information was presented that can boost the intra-class consistency and eliminate spurious sources. We bring up the idea of “cross-and-bouquet” in the inverse problem and present an efficient algorithm to address the high coherence problem of the reconstructed signals given high background spontaneous source signal. An efficient algorithm called feature-sign search algorithm is used to solve the GRDSI problem. We illustrated the superior of our algorithm in retrieving discriminative sources while traditional algorithms failed given certain level of noises.

References

1. Cai, D., He, X., Han, J., Huang, T.S.: Graph regularized nonnegative matrix factorization for data representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **33**(8), 1548–1560 (2011)
2. Costa, F., Batatia, H., Chaari, L., Tourneret, J.Y.: Sparse EEG source localization using Bernoulli laplacian priors. *IEEE Trans. Biomed. Eng.* **62**(12), 2888–2898 (2015)
3. Gramfort, A., Kowalski, M., Hämäläinen, M.: Mixed-norm estimates for the M/EEG inverse problem using accelerated gradient methods. *Phys. Med. Biol.* **57**(7), 1937 (2012)
4. Guan, Z.H., Liu, F., Li, J., Wang, Y.W.: Chaotification of complex networks with impulsive control. *Chaos Interdisc. J. Nonlinear Sci.* **22**(2), 023137 (2012)
5. Guo, H., Jiang, Z., Davis, L.S.: Discriminative dictionary learning with pairwise constraints. In: Lee, K.M., Matsushita, Y., Rehg, J.M., Hu, Z. (eds.) *ACCV 2012*. LNCS, vol. 7724, pp. 328–342. Springer, Heidelberg (2013). doi:[10.1007/978-3-642-37331-2_25](https://doi.org/10.1007/978-3-642-37331-2_25)
6. Hämäläinen, M.S., Ilmoniemi, R.J.: Interpreting magnetic fields of the brain: minimum norm estimates. *Med. Biol. Eng. Comput.* **32**(1), 35–42 (1994)
7. Haufe, S., Ewald, A.: A simulation framework for benchmarking EEG-based brain connectivity estimation methodologies. *Brain Topogr.* 1–18 (2016). doi:[10.1007/s10548-016-0498-y](https://doi.org/10.1007/s10548-016-0498-y)
8. Hipp, J.F., Hawellek, D.J., Corbetta, M., Siegel, M., Engel, A.K.: Large-scale cortical correlation structure of spontaneous oscillatory activity. *Nat. Neurosci.* **15**(6), 884–890 (2012)
9. Huang, Y., Parra, L.C., Haufe, S.: The New York Head—a precise standardized volume conductor model for EEG source localization and tES targeting. *NeuroImage* **140**, 150–162 (2016). transcranial electric stimulation (tES) and Neuroimaging
10. Lee, H., Battle, A., Raina, R., Ng, A.Y.: Efficient sparse coding algorithms. In: *Advances in Neural Information Processing Systems*, pp. 801–808 (2006)
11. Li, Y., Qin, J., Hsin, Y.L., Osher, S., Liu, W.: s-SMOOTH: sparsity and smoothness enhanced EEG brain tomography. *Front. Neurosci.* **10**, 543 (2016)

12. Liu, F., Wang, S., Rosenberger, J., Su, J., Liu, H.: A sparse dictionary learning framework to discover discriminative source activations in EEG brain mapping. In: AAAI, pp. 1431–1437 (2017)
13. Liu, F., Xiang, W., Wang, S., Lega, B.: Prediction of seizure spread network via sparse representations of overcomplete dictionaries. In: Ascoli, G.A., Hawrylycz, M., Ali, H., Khazanchi, D., Shi, Y. (eds.) BIH 2016. LNCS, vol. 9919, pp. 262–273. Springer, Cham (2016). doi:[10.1007/978-3-319-47103-7_26](https://doi.org/10.1007/978-3-319-47103-7_26)
14. Newman, M.E.: The structure and function of complex networks. *SIAM Rev.* **45**(2), 167–256 (2003)
15. Pascual-Marqui, R.D., et al.: Standardized low-resolution brain electromagnetic tomography (sLORETA): technical details. *Methods Find. Exp. Clin. Pharmacol.* **24**(Suppl D), 5–12 (2002)
16. Uutela, K., Hämäläinen, M., Somersalo, E.: Visualization of magnetoencephalographic data using minimum current estimates. *NeuroImage* **10**(2), 173–180 (1999)
17. Watts, D.J., Strogatz, S.H.: Collective dynamics of ‘small-world’ networks. *Nature* **393**(6684), 440–442 (1998)
18. Wright, J., Ma, Y., Mairal, J., Sapiro, G., Huang, T.S., Yan, S.: Sparse representation for computer vision and pattern recognition. *Proc. IEEE* **98**(6), 1031–1044 (2010)
19. Yang, A.Y., Sastry, S.S., Ganesh, A., Ma, Y.: Fast ℓ_1 -minimization algorithms and an application in robust face recognition: a review. In: 2010 17th IEEE International Conference on Image Processing (ICIP), pp. 1849–1852. IEEE (2010)