Structured Sparsity Regularization Approach to the EEG Inverse Problem

Jair Montoya-Martínez  
Antonio Artés-Rodríguez  
Department of Signal Processing and Communications  
Universidad Carlos III de Madrid  
Avda. Universidad 30, Leganés, Madrid, Spain  
Email: {jmontoya,antonio}@tsc.uc3m.es

Lars K. Hansen  
Technical University of Denmark  
DTU Informatics  
Kgs. Lyngby, Denmark  
Email: lkh@imm.dtu.dk

Massimiliano Pontil  
University College London  
Department of Computer Science  
Malet Place London WC1E 6BT, UK  
Email: m.pontil@es.ucl.ac.uk

Abstract—Localization of brain activity involves solving the EEG inverse problem, which is an undetermined ill-posed problem. We propose a novel approach consisting in estimating, using structured sparsity regularization techniques, the Brain Electrical Sources (BES) matrix directly in the source space. We use proximal optimization methods, which are efficient optimization techniques, with good convergence rates and with the ability to handle large nonsmooth convex problems, which is the typical scenario in the EEG inverse problem. We have evaluated our approach under a simulated scenario consisting in estimating a synthetic BES matrix with 5124 sources. The results are compared with conventional applications of $\ell_1$-norm and $\ell_2$-norm regularizers to quantify the improvements in performance.

I. INTRODUCTION

Electroencephalography (EEG) brain imaging is an active research area. It consists in solving the EEG inverse problem: localizing the Brain Electrical Source(s) (BES) distribution based on EEG measurements. EEG provides very high temporal resolution of the brain activity by recording the electrical activity along the scalp. However, its spatial resolution is very low, and makes the task of localizing the BES distribution very challenging: there are very few electrodes (≈ hundreds) with respect to the BES (≈ thousands), and the signal measured by each electrode is affected by all the BES, as well as external noise. Besides, the existence of silent BES (BES that produce nonmeasurable fields on the scalp surface), implies that the EEG inverse problem has no unique solution: a silent BES can always be added to a solution of the inverse problem without affecting the EEG measurements [16]. For all these reasons, the EEG inverse problem is an undetermined ill-posed problem [9].

The standard way of solving this problem is based on the mathematical theory of linear inverse problems. Solutions developed by this theory are stated in terms of a regularization operator, and have the following properties: (i) among all candidate solutions we pick one which satisfies certain prescribed constrains (e.g. smoothness), and (ii) the solution is stable. Concerning the first point, we can use mathematical restrictions (minimum norm estimates) or anatomically, physiologically and functionally based a priori information. Some examples of this kind of neurophysiological information are [9], [6]: the irrotational character of the brain current sources, the (smooth) dynamic of the neural signals, the clusters formed by neighboring or functional related BES, and the smoothness and focality features of the electromagnetic fields generated and propagated within the volume conductor media (brain cortex, skull and scalp).

Several regularization terms have been proposed in the EEG community, including the $\ell_2$-norm (a.k.a Minimum Norm Estimate (MNE)) [9], the $\ell_1$-norm (a.k.a Minimum Current Estimate (MCE)) [17], mixed $\ell_1/\ell_2$-norm in the framework of rotational invariance [10], and also in the framework of time basis, time frequency dictionaries and spatial basis BES decomposition [15], [8], [11], among others (for a more detail overview on inverse methods for EEG see [3], [15] and references therein).

The solutions obtained with MNE are often too diffuse, and tend to estimate sources that spread over a considerable part of the brain, which is not always physiologically meaningful [11]. On the other hand, the solutions obtained with MCE are often too sparse, and tend to be scattered around the true sources [11]. We want to find focal-smooth trade-off solutions, and one alternative to achieve this is through the structured sparsity paradigm [2]. These structured sparsity solutions would be more accurate and neurophysiologically meaningful: they would take into account the smoothness of the BES signals (once a BES is active, it will vary smoothly) and the corresponding functional related clusters in which they group (functional related groups of BES activate together, they work in groups).

The methods based on mixed $\ell_1/\ell_2$-norm [15], [8], [11] used in the framework of the time basis, time frequency dictionaries and spatial basis decomposition look for structured sparsity solutions, but they depends upon decomposing the BES signals as linear combinations of multiple basis functions, e.g. temporal basis functions obtained with SVD [15], time frequency Gabor dictionaries [8] and spatial basis Gaussian functions [11].

All the classical methods mentioned above, solve the EEG inverse nonsmooth optimization problem using standard mathematical programming techniques. A classical approach used by these methods to deal with the “nonsmoothness” of the
optimization problem is to reformulate it as a general Second-Order Cone Programming (SOCP) problem [11], and use interior point-based SOCP solvers. However, there exist other alternatives to attack nonsmooth optimization problems. One of such alternatives, which has begun to be extensively used by the machine learning community is the proximal optimization methods to solve the corresponding nonsmooth optimization problem, which mainly differs from the methods mentioned before in the following aspects: estimate, using structured sparsity regularization techniques. To accomplish this goal, we will choose $\lambda \Omega(S)$ to be a structured sparsity-inducing (usually nonsmooth) norm.

III. STRUCTURED SPARSITY REGULARIZATION

We can introduce sparsity into the EEG inverse problem by penalizing the reconstruction error by the $\ell_0$-pseudo-norm of the BES matrix $S$ (abusing notation, hereafter we will use $\ell_0$-norm instead of $\ell_0$-pseudo-norm, for simplicity):

$$\hat{S} = \min_S \left\{ \frac{1}{2} ||AS - Y||_F^2 + \lambda ||S||_0 : \lambda > 0 \right\}$$

(4)

However, this leads to hard combinatorial nonconvex problems [2]. One classical convex relaxation to this problem is to replace the $\ell_0$-norm by the square $\ell_2$-norm:

$$\hat{S} = \min_S \left\{ \frac{1}{2} ||AS - Y||_F^2 + \lambda ||S||_2^2 : \lambda > 0 \right\}$$

(5)

where $||S||_p = \left( \sum_{i=1}^N \sum_{j=1}^T |S_{ij}|^p \right)^{1/p}$ denotes the “entry-wise” $\ell_p$-norm of the matrix $S$. This is the so called Minimum Norm Estimate (MNE) Problem [9] (also known as Ridge regression in the context of statistics). The problem (5) is a well-conditioned convex problem, but it does not promote sparsity in the estimated BES matrix $\hat{S}$.

A commonly used alternative to promote sparse solutions is the $\ell_1$-norm:

$$\hat{S} = \min_S \left\{ \frac{1}{2} ||AS - Y||_F^2 + \lambda ||S||_1 : \lambda > 0 \right\}$$

(6)

This is the so called Minimum Current Estimate Problem [17] (also known as the LASSO problem in the context of statistics). We know a priori that during the time window of interest for certain cognitive task not all the BES will be activated, so we would like to obtain a sparse model for the BES matrix $S$, and solving (6) we find such sparse representation of $S$. However, the estimated solution $\hat{S}$ is often too sparse and “spiky” to be biologically plausible [15].

We are interested in sparse estimation under additional conditions on the sparsity pattern of the BES matrix $S$: we expect that $S$ is sparse but also that it is structured sparse, namely certain configurations of its nonzero components are to be preferred to others.

In order to introduce structured sparsity into the EEG inverse problem, we need to take into account the dynamic of the BES and the functional related clusters in which they group. We can expect that groups of BES activate and work together, they will vary smoothly, and depending of the time window in which we acquire the EEG measurements, we can expect that they show the following behavior:

1) A group of BES start active, and remains in that state during all the time window: these patterns could be
induced using a non-overlapping group lasso regularizer, where each row of the BES matrix $S$ forms a group:

$$
\lambda \Omega(S) = \lambda \sum_{i=1}^{N} \|S_{i,:}\|_2
$$

2) If the time window is more or less wide, we can expect that a group of BES change its state (active-inactive) several times during the time window. These patterns could be induced using a sparse group lasso regularizer, where each row of the BES matrix $S$ forms a group, but with sparsity in its columns:

$$
\lambda \Omega(S) = \lambda_1 \|S\|_1 + \lambda_2 \sum_{i=1}^{N} \|S_{i,:}\|_2
$$

**IV. EXPERIMENTAL SETUP**

We used the same EEG database than [7] (it can be downloaded from www.fil.ion.ucl.ac.uk/spm/data/mmfaces/). This database was acquired from a subject who participated in a multimodal study on face perception. Data were acquired on a 128-channel ActiveTwo system (820 sample per channel). From this EEG database and using a three-sphere head model with a cortical mesh with 5124 vertices (each vertex location corresponds to a BES position, whose orientation is fixed perpendicular to the surface), we computed the lead-field matrix, $A \in \mathbb{R}^{128 \times 5124}$. It was computed using routines from SPM software [13]. Given EEG measurements, however, is not enough for validating methods for inverse reconstruction due to the lack of “ground truth” [11] (the underlying BES are unknown). Therefore, a standard way of evaluating EEG inverse methods is to assess their ability to reconstruct simulated BES [7], [11], [10], which we have generated following the same procedure mentioned in [7]. The strategy to obtain simulated BES was to use the EEG data described before to define temporal dynamics of evoked responses (particularly, the N170 evoked response, which is an electrophysiological measure of face perception) and assign these dynamics to distributed but contiguous nodes in the cortical mesh. To accomplish this, following [7], we performed an SVD in the channel space ($Y$), retained the first five singular vectors (the same number of singular vectors used in the synthetic experiment made in [7]), $T_\theta \in \mathbb{R}^{5 \times 820}$, and deployed these over five distributed sources. These sources, $q_\theta \in \mathbb{R}^{5124 \times 5}$, were columns (selected randomly) of a spatial coherence prior matrix $G$ [7]:

$$
G(\sigma) = \exp(\sigma B) \approx \sum_{i=0}^{8} \frac{\sigma^i}{i!} B^i
$$

where $B$ is the adjacency matrix, which encodes the neighborhood relationships, $B_{ij} \in [0, 1]$, between the nodes of the cortical mesh. The ensuing source activity was projected through the lead-field matrix to simulate signals in the channel space, $Y_{syn}$ (synthetic EEG measurements) [7]:

$$
Y_{syn} = A S_{syn} + E
$$

where $S_{syn}$ is the synthetic BES matrix, $S_{syn} = q_\theta^T T_\theta$, and $E$ is white gaussian noise whose variance was chosen such that the resulting Signal-to-Noise-Ratio (SNR) was in the interval $[8, 19]$ dB, which according to [7] were fairly typical SNR values of the ERP that they used to measure. The SNR in our setup is $20 \log_{10} (\|Y_{syn}\|_F/\|E\|_F) = 10.7$ dB. So, our synthetic problem can be stated as follows: Given matrix $Y_{syn} \in \mathbb{R}^{128 \times 820}$ and matrix $A \in \mathbb{R}^{128 \times 5124}$, recover the synthetic BES matrix $S_{syn} \in \mathbb{R}^{5124 \times 820}$.

A. Numerical Simulations

We have tested the structured sparsity-inducing regularizers mentioned in section (III) in the synthetic problem. We have used accelerated proximal methods (FISTA) [4] to solve the respective optimization problems, with stopping criterion

$$
\frac{\|\hat{S}_k - \hat{S}_{k-1}\|_F}{\|\hat{S}_k\|_F} \leq 10^{-3}
$$

and Lipschitz constant

$$
L = 2\|A^\dagger A\| = 1.9202 \times 10^{-4}
$$

where $\|A^\dagger A\|$ is the spectral norm of $A^\dagger A$, and $A^\dagger$ denotes the transpose of matrix $A$, respectively.

B. Performance Evaluation

In order to evaluate the performance of the structured sparsity promoting inverse methods, we study how the objective function varies with $\lambda$. We also show how the number of nonzero components of the estimated BES matrix ($\|S\|_0$) vary with $\lambda$. Finally, we report on how is the inner structure of the estimated BES matrix (concerning the sparsity plots,
blue dots mean nonzero components (nz) and LE means localization error (euclidean distance between the most active source (m.a.s) of the synthetic BES matrix and the m.a.s of the estimated one, in millimeters)). This allows a visual compare the estimated matrix to the true matrix.

According to figures (3) and (4), the Ridge regularizer is able to obtain a low reconstruction error and find the closest m.a.s to the true one, but estimating all the components of the resulting BES matrix as nonzeros, which can be seen in the $\ell_0$-norm and sparsity plots, respectively. The structure of the estimated BES matrix is clearly full, as opposed to true synthetic BES matrix, which presents a structured sparsity pattern.

We can start obtaining sparse solutions using the Lasso regularizer. As we can see in figures (5) and (6), this regularizer can obtain different levels of sparse solutions varying the regularization parameter $\lambda$: “big” values of $\lambda$ shrink all the components of the estimated BES matrix toward zero, resulting in the null solution. On the other hand, “small” values of $\lambda$ allow too many nonzero components in the estimated BES matrix, resulting in the fully dense solution. We have selected three values of $\lambda$ in between (the ones corresponding to fairly low reconstruction error and low $\ell_0$-norm, these are marked with red circles) and plotted their corresponding sparsity patterns. Figure (6) shows emerging structured sparsity patterns, but these are still too sparse respect to structured sparsity pattern of the true synthetic BES matrix. This regularizer has improved the results of the Ridge regularizer with respect to the sparse structure of the solutions, but has lost respect to the localization error, now the estimated m.a.s is farther to the true synthetic one.

From figures (7) and (8), we can see that if we introduce a priori knowledge into the problem, using structured sparsity inducing norms (in this case through the $\ell_1/\ell_2$-norm or Group Lasso), structured sparse solutions similar to the synthetic BES matrix starts to emerge. Each group corresponds to one row of the estimated BES matrix. Varying $\lambda$ we determine the sparsity level in the rows of the estimated BES matrix. We have selected three values of $\lambda$ in between (the ones corresponding to estimated BES matrix with low reconstruction error and low $\ell_0$-norm). Figures (7) and (8) show us that Group Lasso regularizer outperforms both Ridge and Lasso regularizers: it is able to find the closest m.a.s to the synthetic true one, the structure of the estimated BES matrix is very similar to the synthetic BES matrix and the reconstruction error and $\ell_0$-norm are quite low.

The sparsity pattern exhibited by the synthetic BES matrix tells us that each BES starts activated and remains in that state during the observation window. This is a very particular case, in general, we could expect changes in the states of several BES during the observation window. In this case, it would be valuable and appropriate to use the Sparse Group Lasso regularizer. We decide to test this regularizer to recover the synthetic BES matrix, even though we know that this matrix does not exhibit the most appropriate structured sparsity pattern to be tackled using this regularizer. From figures (9) and (10), we can see that this regularizer also outperforms the standard EEG inverse methods MNE and MCE, and its performance is comparable to Group Lasso regularizer: the structure of the estimated BES matrix is similar to the synthetic BES matrix, although it also introduces sparsity in the columns. It estimates the m.a.s so good as the Group Lasso regularizer, but using a lower number of nonzero components, and their corresponding reconstruction errors are pretty similar.

C. Ridge Regularizer

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{ridge_regularization.png}
\caption{Performance of Ridge Regularizer.}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{ridge_sparsity_pattern.png}
\caption{Sparsity pattern induced by Ridge Regularizer.}
\end{figure}
D. Lasso Regularizer

\[ \Omega(S) = \lambda ||S||_1 \]

\[ \lambda = 10^{-8}, 10^{-7}, 10^{-6}, 10^{-5}, 10^{-4}, 0 \]

Fig. 5. Performance of Lasso Regularizer.

Fig. 6. Sparsity pattern induced by Lasso Regularizer.

E. Group Lasso Regularizer

\[ \Omega(S) = \lambda \sum_{i} ||S(i,:)||_2 \]

\[ \lambda = 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 0 \]

Fig. 7. Performance of Group Lasso Regularizer.

Fig. 8. Sparsity pattern induced by Group Lasso Regularizer.
F. Sparse Group Lasso Regularizer

\[ \Omega(S) = \lambda_1 ||S||_1 + \lambda_2 \sum_{i} ||S(i,:)||_2, \quad \lambda_2 = 10^{-4}. \]

Fig. 9. Performance of Sparse Group Lasso Regularizer.

\[ |\Omega(S)| \]

\[ \lambda_1 \]

\[ \lambda_2 \]

\[ \lambda_1 = 10^{-3}, \lambda_2 = 10^{-4}, \lambda_1 = 10^{-4}, \lambda_2 = 10^{-4}, \lambda_1 = 10^{-5}, \lambda_2 = 10^{-5}, \lambda_1 = 10^{-6}, \lambda_2 = 10^{-6}, \lambda_1 = 10^{-7}, \lambda_2 = 10^{-7}. \]

Fig. 10. Sparsity pattern induced by Sparse Group Lasso Regularizer.

V. CONCLUSION

We have presented a novel approach to tackle the EEG inverse problem: estimate, using structured sparsity regularization techniques, the BES matrix directly in the source space, without needing to rely on selecting good basis to perform sparse basis decomposition. Our approach makes use of neurophysiological a priori information, which helps to obtain more accurate and interpretable solutions. We have used proximal optimization methods, which are efficient optimization techniques, with good convergence rates and with the ability to handle large nonsmooth convex problems, which is the typical scenario in the EEG inverse problem.

ACKNOWLEDGMENT

The authors would like to thank Dr. Carsten Stahlhut, from DTU Informatics, for valuable discussions about EEG forward and inverse problems, standard algorithms employed in these areas, as well as classical EEG software packages. This work has been partly supported by Ministerio de Educación of Spain (projects 'DEIPRO', id. TEC2009-14504-C02-01, and 'COMONSENS', id. CSD2008-00010). Lars K. Hansen and Massimiliano Pontil was funded by Banco Santander and Universidad Carlos III de Madrid’s Excellence Chair programme.

REFERENCES