Large Sparse State Estimation based on a Parallel Dual Adaptive FISTA Method

Cristhian D. Molina-Machado, Juan D. Martinez-Vargas, Eduardo Giraldo

***Abstract*—In this paper a novel approach for state estimation of large sparse systems is proposed considering a Parallel Dual Adaptive Fast Iterative Shrinkage-Thresholding Algorithm with spatial and temporal constraints. The proposed method is applied to a large sparse state space model with inherent dynamic activity related to electric model of the brain, which is modeled by a state space sparse equation. The inclusion of the temporal constraint improves the resolution in time of the proposed method but holding the sparseness of the solution. In order to efficiently apply the proposed method over large sparse state space models, the proposed algorithm is developed by considering that the method can be divided into several parallel processes. An experimental analysis is performed for a state space model with $n = 74382$, $n = 10016$ and $n = 2004$ states and $d = 230$ outputs. The performance of the proposed method is evaluated in terms of the normalized estimation error and by comparing with the Tikhonov estimation method with dynamic constraints.**

***Index Terms*—Dual adaptive FISTA, Sparse, Temporal Constraint, Large Scale Systems.**

I. Introduction

State estimation of large scale systems is a task required for several applications such as: brain activity estimation, brain-computer interfaces, state estimation in power systems and distribution systems, system dynamics in mechanical systems, among others [1], [2].

The estimation of states for observable large scale systems is a task that can reduce the cost of the measurement elements since the activity of the states is estimated [3]. In some cases, as in the brain activity estimation, the state estimation can be used to determine how is the activity into the brain. This is performed by using a low number of measurements which is useful for epilepsy detection, reducing the necessity of functional resonances, and therefore diminishing the costs [4]. In addition, for systems where the measurements of each state are available, the state estimation can be used as a robustness issue that can validate the estimation and enhance the performance of the system when the measurement of a sensor fails or by using an optimal filtering stage [5].

Estimation of large scale systems from a reduced number of measurements is an ill-conditioned and ill-posed inverse problem. When the measurements have an inherent temporal dynamic the inverse problem solution must contain temporal and spatial constraints, and also must be solved for each sample. An state space estimation method can be easily designed for low scale systems, but when the scale of the system (number of states) is increased, the selection of closed loop gains for the observer is a high-complex and computer-demanding task which carries on a high computational cost [6]. An additional feature of large scale systems is related with the sparseness of its dynamics, that feature is usually exploited by solving a large number of low scale independent state estimation problems where the interaction among subsystems is neglected.

Several methods for estimation of sparse solutions are used where the Fast Iterative Shrinkage-Thresholding Algorithm (FISTA) and its variations are the most common approach [7], [8]. However, FISTA algorithm only considers spatial constraints [9]. In [10] and approach for l_2 norm with temporal and spatial constraint but with high computational cost. In [11] a mixed l_1 norm and l_2 or l_1 norm is proposed for spatial and temporal constraints, however, even when the authors use a projection matrix to reduce the amount of sources and therefore the computational cost a reduction of the quality of the estimation is obtained.

In this work, a novel approach for state estimation of large sparse systems is proposed considering a Parallel Dual Adaptive Fast Iterative Shrinkage-Thresholding Algorithm with spatial and temporal constraints. The proposed method is applied to a large sparse state space model with inherent dynamic activity related to electric system where the activity of each point state is modeled by a second order differential equation. The inclusion of the temporal constraint improves the resolution in time of the proposed method but holding the sparseness of the solution. In order to efficiently apply the proposed method over large sparse state space models, the proposed algorithm is developed by considering that the method can be divided into several parallel processes. An experimental analysis is performed for a state space model with $n = 74382$, $n = 10016$ and $n = 2004$ states and $d = 230$ outputs. The performance of the proposed method is evaluated in terms of the normalized estimation error and by comparing with the Tikhonov estimation method with dynamic constraints. In section II the theoretical framework for large scales state estimation of sparse systems is proposed based in a FISTA method with a proximal operator that allows parallel computation of the estimation. In III the results of large scale state estimation for an sparse system by using form 1 to 12 parallel processes are evaluated in terms of the relative error and speed-up, in comparison with a Tikhonov estimation method. And finally, in V the conclusions and future work of the proposed method are presented.

Manuscript received December 2, 2019; revised June 4, 2020. This work was carried out under the funding of the Departamento Administrativo Nacional de Ciencia, Tecnología e Innovación (Colciencias). Research project: 111077757982 "Sistema de identificación de fuentes epileptogénicas basado en medidas de conectividad funcional usando registros electroencefalográficos e imágenes de resonancia magnética en pacientes con epilepsia refractaria: apoyo a la cirugía resectiva".

C.D. Molina-Machado and E. Giraldo are with the Department of Electrical Engineering, Universidad Tecnológica de Pereira, Pereira, Colombia, e-mail: cdmolina@utp.edu.co, egiraldos@utp.edu.co

J.D. Martinez is with the Department of Electronical Engineering, Intituto Tecnologico Metropolitano, Medellin, Colombia, e-mail: jdmartinezv@itm.edu.co

II. Theoretical Framework

In this work, we use as an application of state space estimation of large sparse systems the widely studied brain imaging problem based on electroencephalographic (EEG) signals. The distributed solution tho this problem consists of projecting the EEG data at the brain cortical surface populated with thousands of candidate sources, yielding the following discrete state space system described by its output equation:

$$\mathbf{y}_k = \mathbf{M}\mathbf{x}_k + \mu_k, \tag{1}$$

being the output of the system $\mathbf{y}_k \in \mathbb{R}^{m\times 1}$ the electroencephalographic (EEG) activity measured at m sensors, $\mathbf{M} \in \mathbb{R}^{m\times n}$ a gain matrix commonly known as lead field, that relates inputs (brain activity) and outputs (EEG data) of the model. Moreover, $\mu_k \in \mathbb{R}^{n\times 1}$ is the measurement noise.

To parallelize the state estimation, the above system equation can be reformulated as:

$$\mathbf{y}_k = \sum_{i=1}^{N} \mathbf{M}^i \mathbf{x}_k^i + \mu_k \tag{2}$$

being N the total number of parallel processes, $M^j \in \mathbb{R}^{m\times n_j}$ the $j-th$ block of the lead-field matrix, and $x_k^j \in \mathbb{R}^{n_j\times 1}$ the corresponding $j-th$ state vector block describing the neural activity.

By considering the inherent temporal evolution in time of $\mathbf{x}_k$ and the sparseness in the solution, the following cost function is proposed [11]:

$$\hat{\mathbf{x}}_k = \arg\min_{\mathbf{x}_k}(\|\mathbf{y}_k - \mathbf{M}\mathbf{x}_k\|_2^2 + \lambda_k^2 \left\|\mathbf{x}_k - \mathbf{x}_k^-\right\|_1 + \gamma_k^2 \|\mathbf{x}_k\|_1) \tag{3}$$

being $\mathbf{x}_k^-$ the a priori estimation which is defined as $\mathbf{x}_k^- = \mathbf{x}_{k-1}$.

By considering (2), the Parallel Dykstra-like splitting, presented in [12], and the Beck and Teboulle FISTA method, discussed in [13], the following Parallel Dual Adaptive FISTA (PDA-FISTA) method can be proposed for (3), where at each $j-th$ block the following equations are computed:

$$t_k = \frac{1+\sqrt{1+4t_{k-1}^2}}{2} \tag{4}$$

$$\alpha_k = \frac{t_k - 1}{t_k} \tag{5}$$

$$\bar{\mathbf{x}}_k^j = \mathbf{x}_k^{j,p} + \alpha_k\left(\mathbf{x}_k^{j,p} - \mathbf{x}_k^{j,p-1}\right) \tag{6}$$

$$\mathbf{g}_1^j = (\mathbf{M}^j)^T\left(\sum_{j=1}^{N}\mathbf{M}^j\bar{\mathbf{x}}_k^j - \mathbf{y}_k\right) \tag{7}$$

$$\mathbf{z}_k^{j,p} = \hat{\mathbf{x}}_k^j + \mathrm{prox}_{\lambda_k,\|\dot{|}\|_1}\left(\bar{\mathbf{x}}_k^j - \hat{\mathbf{x}}_k^j - \delta_p\mathbf{g}_1^j\right) \tag{8}$$

$$\bar{\mathbf{z}}_k^j = \mathbf{z}_k^{j,p} + \alpha_k\left(\mathbf{z}_k^{j,p} - \mathbf{z}_k^{j,p-1}\right) \tag{9}$$

$$\mathbf{g}_2^j = (\mathbf{M}^j)^T\left(\sum_{j=1}^{N}\mathbf{M}^j\bar{\mathbf{z}}_k^j - \mathbf{y}_k\right) \tag{10}$$

$$\mathbf{x}_k^{j,p+1} = \mathrm{prox}_{\gamma_k,\|\dot{|}\|_1}\left(\bar{\mathbf{z}}_k^j - \delta_p\mathbf{g}_2^j\right) \tag{11}$$

being $\mathrm{prox}_{\lambda_k,\|\dot{|}\|_1}$ and $\mathrm{prox}_{\gamma_k,\|\dot{|}\|_1}$ the proximal operators of l_1 norm defined as [12]:

$$\mathrm{prox}_{\omega,\|\dot{|}\|_1}(x) = \begin{cases} x-\omega^2 & \text{if } x<\omega^2 \\ 0 & \text{if } -\omega^2 \le x \le \omega^2 \\ x+\omega^2 & \text{if } x>\omega^2 \end{cases} \tag{12}$$

and being $\delta_p = \frac{1}{L}$ with L defined as the Lipschitz constant and computed as the maximum eigenvalue of $(\mathbf{M}^j)(\mathbf{M}^j)^T$, and $\hat{\mathbf{x}}_k^j$ the $j-th$ block of $\mathbf{x}_k^-$, and p the refinement iteration.

III. Experimental set-up

A large scale model is considered for evaluation of the dynamic state estimation. The model considers a simulated sparse state activity, and an output equation corresponding to a real head model obtained from a high resolution structural Magnetic Resonance Imaging.

A. Forwad model

As volume conductor model (gain matrix), we use the New York (NY) head that is frequently used in neuroimage studies in lack of individual MRIs. NY was constructed based on the segmentation of a symmetric head template (ICBM-152 v2009) into two tissue types (gray matter (GM) and white matter (WM)) and partition of another symmetric template (ICBM-152 v6) into non-brain tissues (CSF, skull, and scalp). The segmentation is shown in top of Fig. 1.

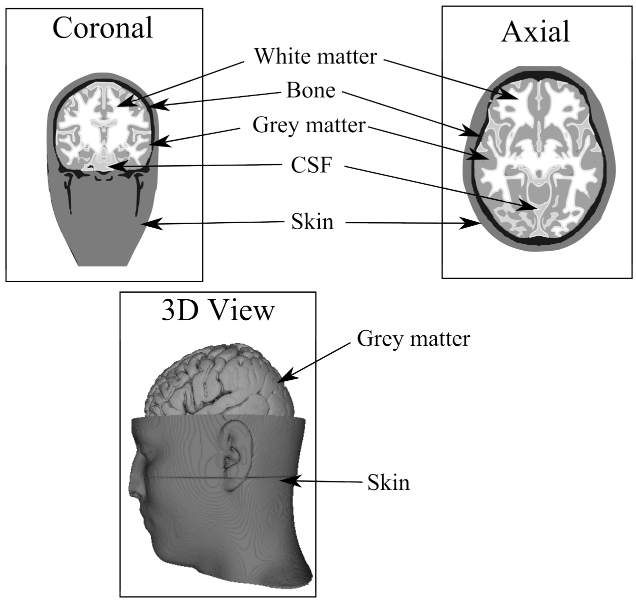

Fig. 1. New York head model. Top: Segmented tissues. Bottom: 3D view of the cortical surface.

For applying distributed solutions to the EEG inverse problem, the New York provides cortical meshes with different numbers of candidate sources (source space), as seen in Fig. 2. Summarizing, our large scale model comprises $m = 230$ outputs (number of EEG channels), and $n = 74382, 10016, 2004$ states (possible places of source activity). A further description of the NY model can be found in [14].

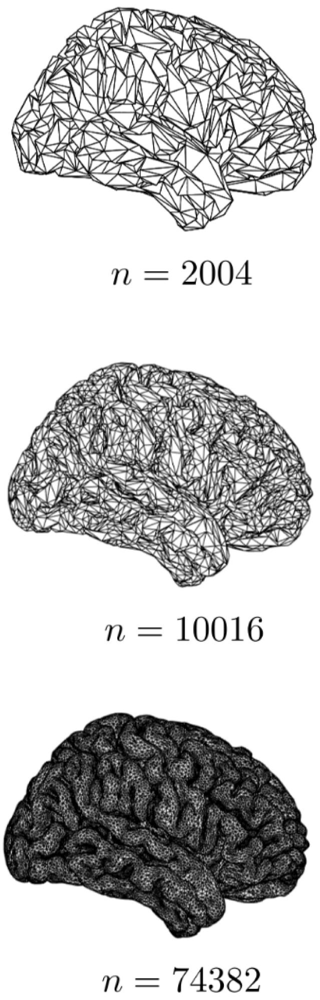

Fig. 2. Brain model for computing $\boldsymbol{M}$

B. Activity simulation

A common approach to assessing the quality of inverse solutions is to use simulated state activity, for which the underlying brain activity is known. Thus, an sparse state space equation is used to described the dynamical of the EEG simulation model. The state space equation for the $i-th$ state is defined as:

$$\begin{aligned} x_k =& w_1 x_{k-1} + w_2 x_{k-2} + w_3 x_{k-\tau} \\ &+ w_4 x_{k-1}^2 + w_5 x_{k-1}^3 + \eta_k \end{aligned} \tag{13}$$

being $w_1 = 1.0628$, $w_2 = -0.42857$, $w_3 = 0.008$, $w_4 = 0.000143$ y $w_5 = -0.000286$, $\tau = 20$ and by considering a parameter variation at sample 50 to the following parameters $w_1 = 1.3$, $w_2 = -1$.

A sparse distance matrix $\mathbf{Q} \in \mathbb{R}^{n \times n}$ is used to define the set of active sources. The structure of the state transition matrix for the $n = 74382$ states model is presented in Fig. 3. Here, the number of non-zero elements is 15605938.

The set of active sources around source 20000 from distance matrix Q is presented in Fig. 4.

Later, using eq. (1), state activity is mapped to the measurements considering a Signal-to-Noise-Ratio of 5dB. Simulation results for $T = 100$ samples an a sample time of 100 milliseconds are shown in Fig. 5. It can be seen that the simulated activity exhibit from the measurements point of view of a combination of the activity in the $n = 74382$ states.

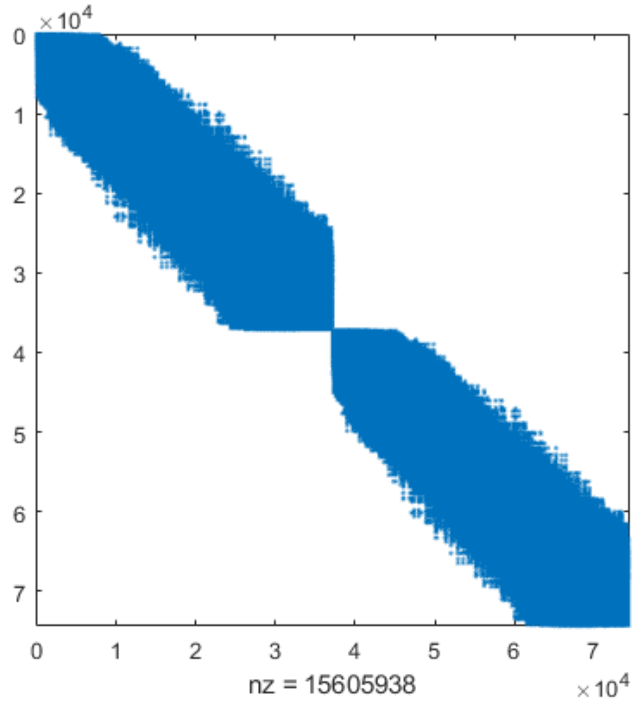

Fig. 3. Distance matrix of $n = 74382$ states where the number of non-zero elements is 15605938.

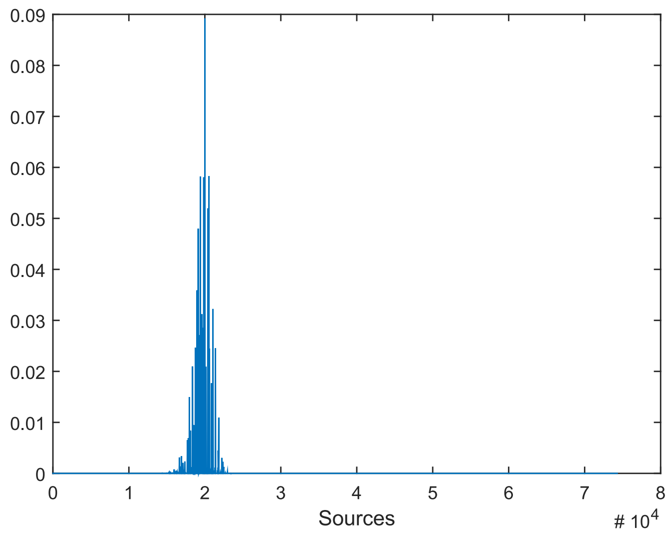

Fig. 4. Set of active sources around source 20000 for a $n = 74382$ states.

C. Comparison approaches

The proposed model is compared with a regularized state space estimation method called Tikhonov with dynamic constraints. The estimation is computed as follows:

$$\hat{\mathbf{x}}_k = (\mathbf{M}^T\mathbf{M} + \lambda_k^2 \mathbf{I}_n)^{-1}(\mathbf{M}^T\mathbf{y}_k + \lambda_k^2 \mathbf{x}_{k-1}), \tag{14}$$

where the $\mathbf{x}_{k-1}$ is the estimated activity at sample $k-1$. It can be seen that this method requires the inverse of an $n \times n$ matrix, and for implementation the inverse is computed by using a Singular Value Decomposition.

The selection of the regularization parameters of (14) and (4) are computed by using the generalized cross-validation.

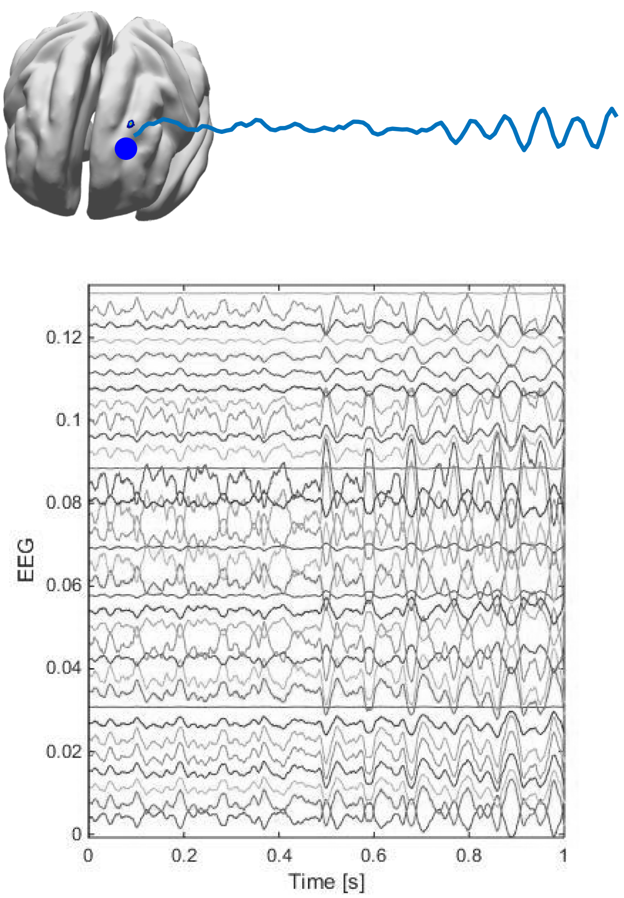

Fig. 5. Simulated brain activity. Top left: Simulated source around $n = 20000$ in the $n = 74382$ model. Top right: Simulated time series according to the state-space model described above. Bottom: Simulated EEG $\boldsymbol{y}_k$ for the first $m = 18$ channels (measurements)

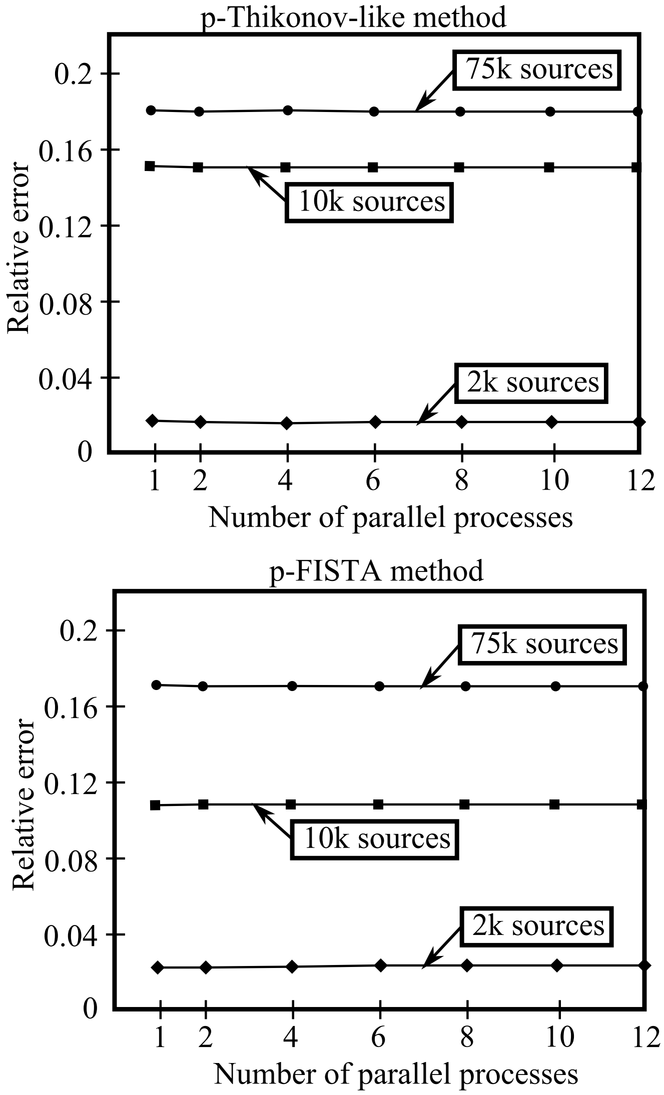

Fig. 6. Comparison of Tikhonov and FISTA estimation results in terms of the relative error

The comparison is performed in terms of relative estimation error computed as follows:

$$e_k = \frac{\|\mathbf{x}_k - \hat{\mathbf{x}}_k\|}{\|\mathbf{x}_k\|} \tag{15}$$

IV. Results and Discussion

The relative error calculation by using the FISTA method and Thikonov method in terms of the relative error are shown in Fig. 6. The experiment is repeated 50 times to account for variations in the simulated additive noise. It is noticeable that for $n = 74382$, $n = 10016$ and $n = 2004$ states the estimation results under several parallel processes (1 to 12) is hold. That means that the estimation is not affected by the number of parallel processes used for the computation of the dual FISTA algorithm based on the proximal operator. In addition, the results obtained by the Tikhonov method are also hold, which validated the implementation of the method. It is remarkable that for the Tikhonov method the mathematical library used for linear algebra operations is compiled by using the number of parallel processes detailed in Fig. 6. In addition it can be seen that the relative error is lower for the FISTA algorithm (proposed method) in comparison with the Tikhonov method.

Figures 7 and 8 show an example of the achieved reconstruction using the methods under comparison. Top left figure (brain activity reconstruction) shows that both FISTA and Tikhonov methods are able to identify the simulated source (shown in Fig. 5) in the frontal left lobe of the brain. However, it is noticeable that the sparseness achieved by the FISTA method, allows to obtain a more focal states reconstruction unlike the Tikhonov solution which is not able to follow the sparse pattern of the simulated activity. Consequently, the Tikhonov solution shows activity almost in the entire brain lobe. The same patter can be identified in the bottom of both Figures (source distribution).

Top right panels of Figs. 7 and 8 show the reconstructed time series around state 20000 (simulated state location). It can be seen that regardless the sparse penalty for states transition (second term of Eq. 3), the FISTA temporal pattern resembles to the simulated one, alike to the Tikhonov solution. Sumarizing, FISTA properly identifies both the spatial sparse pattern and the temporal changes in the simulated states activity.

The speedup result of the proposed method by comparing the performance for several parallel processes is presented in Fig. 9. It can be seen that the acceleration in the computation of the estimation method is increased when the number of parallel processes is increased. Therefore, a reduction in the computational time can be obtained without reducing the

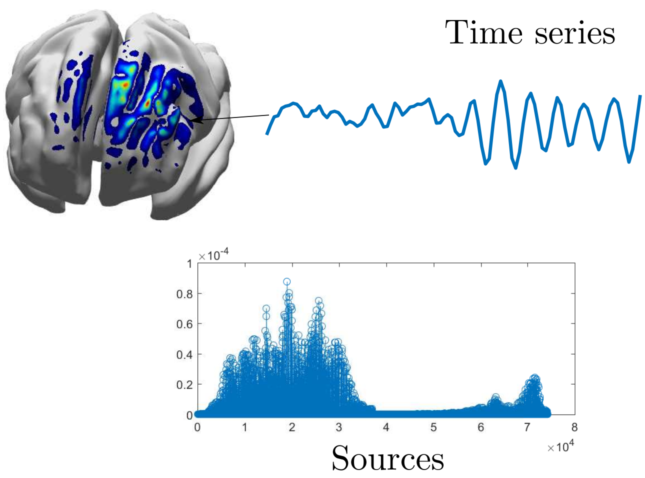

Fig. 7. Achieved solution using Tikhonov estimation. Top left: Brain activity reconstruction.Top right: Reconstructed time series around state 20000. Bottom: Source distribution.

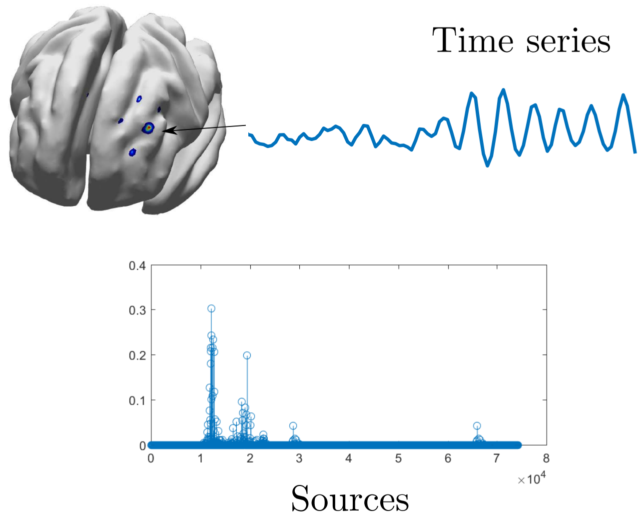

Fig. 8. Achieved solution using FISTA estimation. Top left: Brain activity reconstruction.Top right: Reconstructed time series around state 20000. Bottom: Source distribution.

estimation accuracy.

In Fig. 10 is presented the comparison of Simulated, Tikhonov and FISTA estimation results in terms of brain mapped activity and their corresponding time series. It can be seen, that the mapped activity by using the FISTA method reduces the dispersion and estimate adequately the time series.

V. Conclusions

In this work a novel method for state estimation of large sparse systems is proposed considering a Parallel Dual Adaptive Fast Iterative Shrinkage-Thresholding Algorithm with spatial and temporal constraints. The proposed method is applied to a large sparse state space model with inherent dynamic activity related to electric system where the activity of each point state is modeled by a second order differential equation. The inclusion of the temporal constraint improves the resolution in time of the proposed method but holding the sparseness of the solution. According to the presented results the proposed method reduces the relative estimation

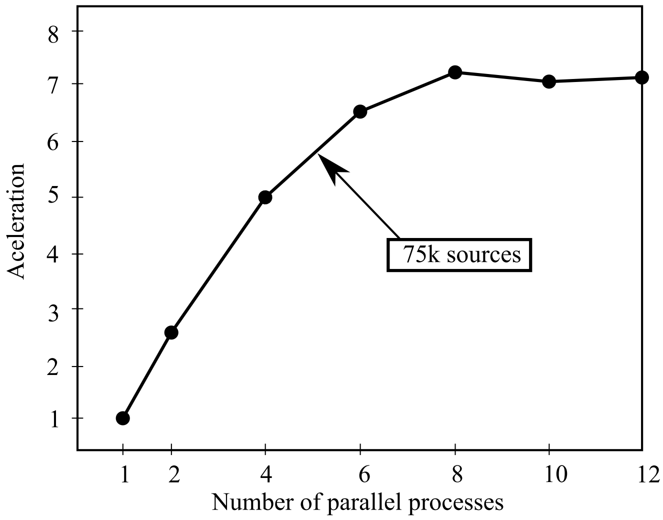

Fig. 9. Speedup performance for several parallel processes

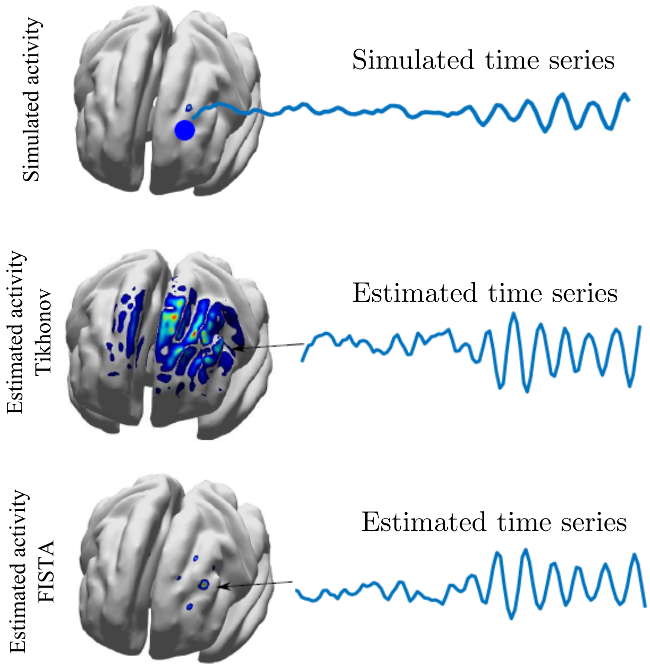

Fig. 10. Comparison of Simulated, Tikhonov and FISTA estimation results in terms of brain mapped activity and their corresponding time series

error in comparison with the Tikhonov estimation method for large scale systems. It can be seen that the relative estimation error for large scale sparse systems is hold even when the number of sources of the model is increased, in contrast with the Tikhonov method where the error is increased when the number of states is increased. As a result an estimation method for large scale systems is validated for several number of states where the selection of the parameters of the observer are computed by using generalized cross-validation.

References

[1] J. Mota, N. Deligiannis, A. C. Sankaranarayanan, V. Cevher, and M. Rodrigues, "Dynamic sparse state estimation using l_1-l_1 minimization: Adaptive-rate measurement bounds, algorithms and applications," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2015, pp. 3332–3336.

[2] W. Pan, Y. Yuan, J. Goncalves, and G. Stan, “A sparse bayesian approach to the identification of nonlinear state-space systems,” *IEEE Transactions on Automatic Control*, vol. 61, no. 1, pp. 182–187, Jan 2016.

[3] J. D. Martinez-Vargas, E. Giraldo, and G. Castellanos-Dominguez, “Enhanced spatio-temporal resolution using dynamic sparse coding for eeg inverse problem solutions,” *IAENG International Journal of of Computer Science*, vol. 46, no. 4, pp. 564–574, 2019.

[4] M. Bueno-Lopez, P. A. Munoz-Gutierrez, E. Giraldo, and M. Molinas, “Electroencephalographic source localization based on enhanced empirical mode decomposition,” *IAENG International Journal of of Computer Science*, vol. 46, no. 2, pp. 228–236, 2019.

[5] M. Bueno-Lopez, E. Giraldo, M. Molinas, and O. Fosso, “The mode mixing problem and its influence in the neural activity reconstruction,” *IAENG International Journal of of Computer Science*, vol. 46, no. 3, pp. 384–394, 2019.

[6] C. D. Molina-Machado, J. D. Martinez-Vargas, and E. Giraldo, “Dynamic state estimation for large scale systems based on a parallel proximal algorithm,” *Engineering Letters*, vol. 28, no. 2, pp. 347–351, 2020.

[7] Z. Peng, M. Yan, and W. Yin, “Parallel and distributed sparse optimization,” in *2013 Asilomar Conference on Signals, Systems and Computers*, Nov 2013, pp. 659–646.

[8] C. Y. Lin and J. A. Fessler, “Efficient dynamic parallel mri reconstruction for the low-rank plus sparse model,” *IEEE Transactions on Computational Imaging*, vol. 5, no. 1, pp. 17–26, March 2019.

[9] C. D. Molina-Machado, E. Cuartas, J. D. Martínez-Vargas, and E. Giraldo, “Comparative analysis of parallel brain activity mapping algorithms for high resolution brain models,” *TecnoLógicas*, vol. 22, no. 46, pp. 177–187, 2019.

[10] E. Giraldo, *Nonlinear time varying model identification in ill-posed problems*. Germany: Scholar’s Press, 2014.

[11] E. Giraldo-Suarez, J. Martinez-Vargas, and G. Castellanos-Dominguez, “Reconstruction of neural activity from EEG data using dynamic spatiotemporal constraints,” *International Journal of Neural Systems*, vol. 26, no. 07, pp. 1–15, 2016.

[12] P. L. Combettes and J.-C. Pesquet, *Proximal Splitting Methods in Signal Processing*. New York, NY: Springer New York, 2011, pp. 185–212. [Online]. Available: https://doi.org/10.1007/978-1-4419-9569-8-10

[13] J. Liang and C.-B. Schönlieb, “Improving FISTA: Faster, Smarter and Greedier,” *arXiv e-prints*, p. arXiv:1811.01430, Nov 2018.

[14] Y. Huang, L. C. Parra, and S. Haufe, “The new york head: A precise standardized volume conductor model for eeg source localization and tes targeting,” *NeuroImage*, vol. 140, pp. 150 – 162, 2016.