Single-trial Connectivity Estimation through the Least Absolute Shrinkage and Selection Operator

Yuri Antonacci, *Member, IEEE*, Jlenia Toppi, *Member, IEEE*, Donatella Mattia, Antonio Pietrabissa and Laura Astolfi, *Member, IEEE*

Abstract— **Methods based on the use of multivariate autoregressive models (MVAR) have proved to be an accurate tool for the estimation of functional links between the activity originated in different brain regions. A well-established method for the parameters estimation is the Ordinary Least Square (OLS) approach, followed by an assessment procedure that can be performed by means of Asymptotic Statistic (AS). However, the performances of both procedures are strongly influenced by the number of data samples available, thus limiting the conditions in which brain connectivity can be estimated. The aim of this paper is to introduce and test a regression method based on Least Absolute Shrinkage and Selection Operator (LASSO) to broaden the estimation of brain connectivity to those conditions in which current methods fail due to the limited data points available. We tested the performances of the LASSO regression in a simulation study under different levels of data points available, in comparison with a classical approach based on OLS and AS. Then, the two methods were applied to real electroencephalographic (EEG) signals, recorded during a motor imagery task. The simulation study and the application to real EEG data both indicated that LASSO regression provides better performances than the currently used methodologies for the estimation of brain connectivity when few data points are available. This work paves the way to the estimation and assessment of connectivity patterns with limited data amount and in on-line settings.**

I. INTRODUCTION

Brain connectivity plays an important role in understanding how different brain regions interact. Along the years, different approaches were defined, among which those based on Granger causality [1] are based on the statistical properties of the signals and do not require any a priori knowledge about the brain network structure. Such connectivity estimators rely on multivariate autoregressive models (MVAR), and their application to Electroencephalographic (EEG) data often exploits a frequency version of the estimator (e.g, the Partial Directed Coherence, PDC [2]). However, in order to provide an accurate and reliable estimation, MVAR models require an appropriate amount of data points, confining application to offline analysis [3]. Unfortunately, overcoming this constraint is still a challenging task, since a low amount of data samples implies an increase of the mean squared error in MVAR parameter estimation, obtained by means of the commonly used Ordinary Least Square (OLS) approach. Thus, to ensure the accuracy of OLS estimation, the ratio between the number of data samples available and the number of MVAR parameters (K-ratio) should be greater than 10 [4].

When this condition is not fulfilled, the estimation problem becomes ill-posed and under-determined and the accuracy in estimating MVAR parameters drops dramatically [5]. Moreover, the accuracy of the statistical assessment procedure (necessary to discard spurious links) is also affected by the number of data samples available, as demonstrated for the recent method based on Asymptotic Statistics (AS) in [6].

A theoretical solution to this problem is represented by penalized regression methods. Among the different methods available, the Least Absolute Shrinkage and Selection Operator (LASSO) represents a viable way to regularize the OLS problem with a constraint based on l_1 norm applied on the MVAR parameters [7]. This procedure has a double effect: i) it reduces the mean square error associated with the estimated coefficients; ii) it reduces the error in discarding spurious links, circumventing the statistical assessment, by means of variable selection.

The use of penalized regression methods for connectivity estimation in the case of few data points and real time applications were reported in [3], [8] and [9] for BCI purposes, and applied to simulated and real data. Furthermore, the feasibility of using LASSO regression was demonstrated for fMRI data [10]. To our knowledge, a direct comparison between a classical approach and LASSO regression for EEG data has not yet been performed.

To this purpose, we firstly performed a simulation study to test the performances of the classical OLS/AS approach in comparison to LASSO regression, used for the estimation of MVAR parameters and for the assessment procedure. Secondly, the two methods were compared by means of the application to data obtained during a motor imagery task (MI) performed by a healthy subject, and already used in previous works about real-time connectivity estimation [9].

II. METHODS

A. Ordinary Least Square regression (OLS)

In the framework of linear signal processing, it is possible to represent a multivariate process as follows:

$$Y(n) = \sum_{k=1}^{p} A(k)Y(n-k) + E(n) \qquad (1)$$

Y. Antonacci, J. Toppi, and L. Astolfi, are with the Department of Computer, Control, and Management Engineering, University of Rome Sapienza, Rome, Italy, and with IRCCS Fondazione Santa Lucia, Rome, Italy. D. Mattia is with IRCCS Fondazione Santa Lucia, Rome, Italy.
A. Pietrabissa is with the Department of Computer, Control, and Management Engineering, University of Rome Sapienza, Rome, Italy.

where:

$A(k)$ are the M x M coefficients matrix in which the element $a_{ij}(k)$ describes the dependence of $y_i(n)$ on $y_j(n-k)$; $Y(n) = [y_1(n), \ldots, y_M(n)]^T$, represents the multivariate closed-loop process; M represents the number of time series involved in the analysis; p represents the model order; $E(n)$ represents the innovation process, assumed to be composed of white and uncorrelated noise.

The OLS approach to the estimation of MVAR coefficients is based on the minimization of the residuals **E** [11]:

$$\hat{A} = argmin_A \|Y - AX\|_2^2 \tag{2}$$

B. Least Absolute Shrinkage and Selection Operator (LASSO)

LASSO differs from OLS because it is based on a penalized regression. It solves the following problem [7]:

$$\hat{A} = argmin_A \|Y - AX\|_2^2 + \lambda \|A\|_1 \tag{3}$$

where λ is the regularization parameter that controls the amount of penalization to be applied during the coefficients' estimation.

C. Partial Directed Coherence (PDC)

PDC estimator in its squared version is defined as follows:

$$\pi_{ij}(f) = \frac{|A_{ij}(f)|^2}{\sum_{m=1}^{M} |A_{mj}(f)|^2} \tag{4}$$

where $A_{ij}(f)$ is the frequency version of the MVAR parameters (estimated by means of OLS or LASSO) that relate the j-th time series to the i-th one, at the specific frequency f.

D. Asymptotic Statistics (AS)

PDC estimation needs to be followed by a statistical assessment against the null-case, to avoid the detection of spurious causal relationships between signals. Recently, the probability distribution of the null-case for PDC was theoretically derived and used by the asymptotic statistics (AS) assessment method to extract null-case statistical threshold [6]. The AS requires a reduced computational time in comparison with empirical methods, providing the same accuracy [12].

III. SIMULATION STUDY

The simulation study consisted of the following steps:
i. generation of simulated EEG-like datasets, fitting predefined ground truth networks of 10 nodes, under different conditions of K ratio (factor K: 10, 5, 3, 1, 0.8). K ratio is defined as follows:

$$K = \frac{N_{samp}}{N_{CH}p_{opt}} \tag{5}$$

where:
N_{SAMP} represent the number of data samples available; $N_{CH}p_{opt}$ represents the number of MVAR coefficients to be estimated.

The simulated EEG data were generated by means of MVAR filter used as a generator [12]. The structure of the ground-truth networks varied across the iterations with an imposed density equal to 25%, while the value assigned to the non-null connections between signals varied randomly between 0 and 0.9 (in absolute value).

ii. Selection of the optimal regularization parameter (λ) for LASSO regression by means of Generalized Cross Validation method (GCV) [7].
iii. Estimation of the autoregressive parameters by means of OLS and LASSO regression methods and subsequent computation of PDC (factor ALG: OLS, LASSO).
iv. Assessment of the significant links by means of AS and LASSO (factor TYPE: AS, LASSO).
v. Evaluation of the accuracy in the estimation of non-null connection for both OLS and LASSO. We used the Mean Absolute Error (MAE) [13]. Let PDC and $\widehat{PDC}$ denote vectors of the imposed and estimated values of PDC for the entire frequency interval between 1 and 40 Hz. The MAE is defined as:

$$MAE = \frac{1}{N}\frac{1}{F}\sum_{n=1}^{N}\sum_{f=1}^{F}\left|\frac{PDC(f,n) - \widehat{PDC}(f,n)}{PDC(f,n)}\right| \tag{6}$$

where N is the number of connections imposed with a nonzero value and F is the number of frequency bins analyzed. The minimum value that MAE can assumed is 0. It does not have an upper limit.
vi. Evaluation of the performance of AS and LASSO in discarding spurious links through the false positive and false negative rates, summarized by a ROC curve and the related Area Under Curve (AUC) [14]. AUC values are between 0.5 (50%) (random classification) and 1 (100%) (correct classification).

The entire procedure was repeated 50 times. The two indices (MAE and AUC) were subjected to a two-way repeated measures ANOVA considering K and ALG as within main factors for MAE and K and TYPE as within main factors for AUC.

IV. APPLICATION TO REAL EEG DATA

The data used to test the different methods can be found in [9]. They were recorded from a healthy volunteer (male, under 30 and right-handed) performing a motor imagery task (MI). The task consisted in performing 90 trials of right-hand MI and 90 trials of foot MI trials. Signals were recorded by a 45-channels EEG, with electrodes located according to the international 10-20 system. Data were filtered with a 0.5-100 Hz bandpass and a 50 Hz notch filter. EOG components were reduced by means of a regression approach [9]. We further resampled the data at 100 Hz.

A. Single trial analysis

The single trial analysis consisted of the following steps:
i. Selection of 11 of the 45 EEG channels available (C5, C3, C1, C2, C4, C6, CP3, CP4, Cz, CPz, FCz) to match the simulation study conditions.
ii. Segmentation of the EEG signals, including the 100 samples between the third and the fourth second of each trial for each condition.
iii. Estimation of the model order for each condition by means of Final Prediction Error criterion (as suggested in [15]).
iv. Estimation of the optimal value for the regularization parameter (for each trial and for each condition) by means of GCV.
v. Estimation of the autoregressive parameters (for each trial and for each condition), by means of OLS and LASSO regression methods, and subsequent computation of PDC.

vi. PDC assessment by means of AS and LASSO (for each trial and condition).

V. RESULTS

A. Simulation study

Table 1 reports the results of the two-way interaction between factor K and ALG for MAE index.

TABLE I.   ANOVA RESULTS - MAE INDEX

Parameter	ANOVA within factors		
	ALG	*K*	*ALG x K*
DoF	(1, 49)	(4, 196)	(4, 196)
MAE	99.9**	63.4**	29.1**

Result of the two-way ANOVA for MAE index. Symbol ** means p<0.001

Table 1 shows a significant effect of all the within factors and their interactions to the MAE. Figure 1 shows the related plot of means.
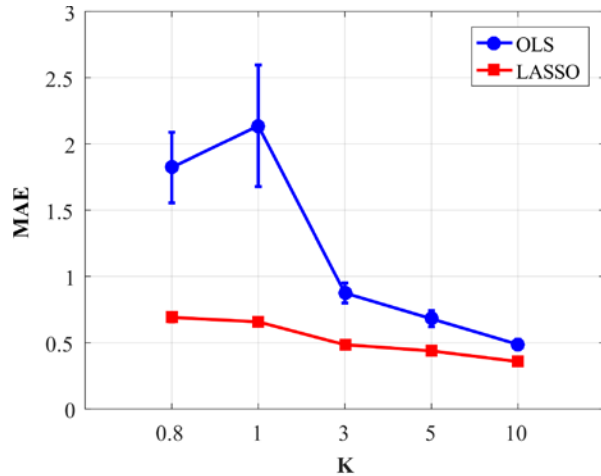

Figure 1- Plot of means of MAE index for the interaction ALG x K. The diagram shows the mean values of the index for different values of K and different methods. The bars represent the 95% confidence interval.

Regardless the method used for the estimation of PDC, the value of K-ratio strongly influences the estimation performances, with a decrease of the estimation error with the increase of K-ratio. While the performances of the two methods are superimposable for values of K equal to 10, in the case of small amount of data (K<3) the error level can rise up to 2 (200%) for OLS, while LASSO shows a mean value of MAE between 0.5 and 1 in the same conditions.

Table 2 reports the ANOVA results for the two-way interaction between factor K and TYPE for AUC, showing a significant effect of all the within factors and their interactions on the AUC index. Figure 2 shows the plot of means of AUC computed after the validation of PDC by means of AS and LASSO, suggesting how the AUC index strongly depends on all the factors analyzed. In particular, AUC increases with the increase of K-ratio, while, with the decrease of K-ratio (i.e. when few data samples are used), AS does not represent a reliable method for assessing brain connectivity. In fact, when K-ratio is lower than 1, the value of AUC is around 50%, corresponding to chance level. This is not the case for LASSO in the same condition. In fact, when K is equal to 0.8 and 1, AUC is still greater than 0.8. Even when K is equal to 10,

Tukey's post-hoc test highlights a statistical difference between AS and LASSO in discarding spurious links.

TABLE II.   ANOVA RESUTLS - AUC

Parameter	ANOVA within factors		
	TYPE	*K*	*TYPE x K*
DoF	(1, 49)	(4, 196)	(4, 196)
AUC	3163**	1996**	365**

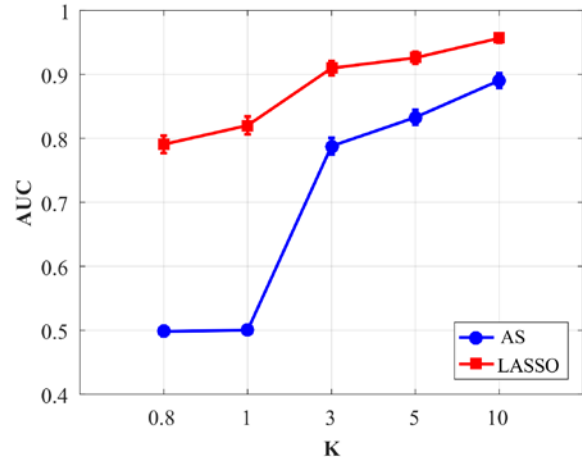Result of the two-way ANOVA on AUC index. Symbol ** means p<0.001

Figure 2- Plot of means of AUC index for the interaction TYPE x K. The diagram shows the mean values of the index for different values of K and different methods. The bars represent the 95% confidence interval.

B. Real EEG data application

Figure 3 shows the plots of the PDC-based connectivity patterns obtained by OLS-AS and LASSO, for the Foot and Hand experimental conditions (upper and lower part of the figure, respectively) at 10 Hz.
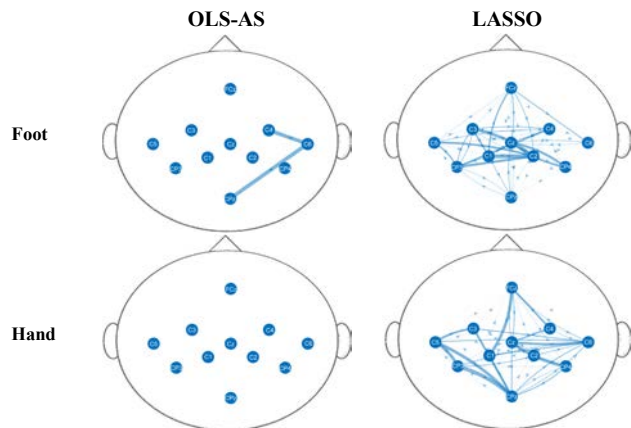

Figure 3- Connectivity patterns obtained for a single trial (randomly selected among the 90 available) at 10 Hz. The arrows thickness encodes the strength of connections, that range between 0 and 1. Scalp representation, nose to the upper part of the page.

Figure 3 shows how in the Foot condition the OLS regression followed by the AS only few connections survive the assessment procedure (only two connections in the right hemisphere for this specific trial). On the other hand, the estimation and assessment obtained by LASSO highlights a strong involvement of the electrode Cz (which is located above the feet primary motor areas) in the network. In the

Hand condition, the combination of OLS and AS cannot estimate any connection between the analysed signals in a single trial. In the same condition, the estimation-assessment procedure performed by LASSO shows the presence of strong connections in the left hemisphere (contralateral to the MI task). Moreover, it is possible to identify a sub-network composed by the electrodes: C1, C3, C5, CP3, Cpz and Fcz.

VI. DISCUSSION AND CONCLUSION

The aim of the present work was to determine if the LASSO regression can be a viable solution for the estimation of PDC-based brain connectivity when few data samples are available, as in EEG single trial analysis.

Simulation study

As expected, PDC values estimated with OLS and LASSO were strongly influenced by the value of K-ratio. A decrease in the accuracy of the estimation with the decrease of K-ratio was investigated in previous works, that highlighted a direct relationship between the MVAR estimation accuracy (in term of MSE) and the number of data samples available [4]. Our results (Figure 1) show that, when K≤1, the PDC values estimated by OLS exhibit a strong bias (MAE > 1.5), while those estimated by means of LASSO regression showed no brisk discontinuities in their accuracy, indicating its tolerance to the reduction of data samples.

Since the network structure and the consequent assumptions are more influenced by the assessment procedure than by the PDC values, we further analysed the performances in term of AUC. LASSO regression does not require an assessment procedure, due to the automatic selection of the MVAR parameters in the time domain (and consequently in the frequency domain). Such feature allows to unify the process of PDC estimation and the consequent statistical assessment in a unique step. As can be seen in Figure 2, LASSO regression provided a higher value of AUC than AS, for all the conditions analysed. In particular, for low values of K-ratio, the performances of AS break down to the chance level, as expected, since we are operating outside the conditions required by the classical estimation of MVAR parameters [12].

Application on real EEG data

By applying PDC estimation to a single trial case, we tested the methods in a challenging condition. Figure 3 shows how, even if K is equal to 1 (which means that we have few data), LASSO procedure allowed to recognize two different networks, one for each task, showing a strong involvement of Cz electrode as a main hub in the Foot condition and a network involving the electrodes C1, C3, C5 and CP3 for the Hand condition. This is in line with previous knowledge on the task. In fact, Hand/Foot motor imagery results in a modulation of the power spectral density in the alpha band [8-12 Hz], detectable in C3 (Hand), Cz and FCz (Foot) [16]. Furthermore, previous studies used connections that involve the electrodes Cz, FCz, C3, C1 and CP3 to discriminate the two MI tasks [15].

The use of a classical OLS/AS approach, on the other hand, had a very different outcome, showing no meaningful patterns in any of the two conditions. This was also expected as we were operating outside the conditions usually accepted to perform an MVAR estimation [4].

Even if previous works reported the effectiveness of LASSO regression when few data samples are available in fMRI context [10], to our knowledge this is the first time that a classical approach is compared with LASSO regression for different conditions of K-ratio on (simulated and real) EEG data.

In summary, the results of this study suggest that the use of LASSO regression can improve the accuracy of the entire estimation process (including the validation), allowing to obtain meaningful results even with a single trial approach. Consequently, this approach could open the way to future applications ranging from the analysis of single EEG trials when the hypothesis of stationarity along trials is not verified, to the real time estimation of brain connectivity for Brain Computer Interfaces (BCI) applications.

REFERENCES

[1] C. W. J. Granger, «Investigating Causal Relations by Econometric Models and Cross-spectral Methods», 1969.

[2] L. A. Baccalá e K. Sameshima, «Partial directed coherence: a new concept in neural structure determination», mag. 2001.

[3] T. R. Mullen *et al.*, «Real-Time Neuroimaging and Cognitive Monitoring Using Wearable Dry EEG», *IEEE Trans Biomed Eng*, 2015.

[4] A. Schlögl e G. Supp, «Analyzing event-related EEG data with multivariate autoregressive parameters», *Prog. Brain Res.*,2006.

[5] D. E. Farrar e R. R. Glauber, «Multicollinearity in Regression Analysis: The Problem Revisited», *The Review of Economics and Statistics*, 1967.

[6] D. Y. Takahashi, L. A. Baccalà, e K. Sameshima, «Connectivity Inference between Neural Structures via Partial Directed Coherence», vol. 34, n. 10, pagg. 1259–1273, 2007.

[7] R. Tibshirani, «Regression Shrinkage and Selection via the Lasso», *Journal of the Royal Statistical Society. Series B (Methodological)*, 1996.

[8] Y. Antonacci, J. Toppi, S. Caschera, A. Anzolin, D. Mattia, e L. Astolfi, «Estimating brain connectivity when few data points are available: Perspectives and limitations», in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2017, pagg. 4351–4354.

[9] M. Billinger, C. Brunner, e G. R. Müller-Putz, «Single-trial connectivity estimation for classification of motor imagery data», *J Neural Eng*, 2013.

[10] P. A. Valdés-Sosa *et al.*, «Estimating brain functional connectivity with sparse multivariate autoregression», vol. 360, n. 1457, 2005.

[11] H. Lütkepohl, *New Introduction to Multiple Time Series Analysis*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005.

[12] J. Toppi, D. Mattia, M. Risetti, R. Formisano, F. Babiloni, e L. Astolfi, «Testing the Significance of Connectivity Networks: Comparison of Different Assessing Procedures», *IEEE Trans Biomed Eng*, 2016.

[13] S. Kim e H. Kim, «A new metric of absolute percentage error for intermittent demand forecasts», *International Journal of Forecasting*, 2016.

[14] J. Toppi *et al.*, «Measuring the agreement between brain connectivity networks», in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2016, pagg. 68–71.

[15] D. Rathee, H. Cecotti, e G. Prasad, «Single-trial effective brain connectivity patterns enhance discriminability of mental imagery tasks», *Journal of neural engineering*, vol. 14, n. 5, pag. 056005, 2017.

[16] F. Pichiorri *et al.*, «Sensorimotor rhythm-based brain-computer interface training: the impact on motor cortical responsiveness», vol. 8, n. 2, pag. 025020, apr. 2011.