

Improving J-divergence of brain connectivity states by graph Laplacian denoising

Tiziana Cattai, Gaetano Scarano, Marie-Constance Corsi, Danielle S.Bassett,
Fabrizio De Vico Fallani, Stefania Colonnese, *Senior Member, IEEE*

Abstract—Functional connectivity (FC) can be represented as a network, and is frequently used to better understand the neural underpinnings of complex tasks such as motor imagery (MI) detection in brain-computer interfaces (BCIs). However, errors in the estimation of connectivity can affect the detection performances. In this work, we address the problem of denoising common connectivity estimates to improve the detectability of different connectivity states. Specifically, we propose a denoising algorithm that acts on the network graph Laplacian, which leverages recent graph signal processing results. Further, we derive a novel formulation of the Jensen divergence for the denoised Laplacian under different states. Numerical simulations on synthetic data show that the denoising method improves the Jensen divergence of connectivity patterns corresponding to different task conditions. Furthermore, we apply the Laplacian denoising technique to brain networks estimated from real EEG data recorded during MI-BCI experiments. Using our novel formulation of the J-divergence, we are able to quantify the distance between the FC networks in the motor imagery and resting states, as well as to understand the contribution of each Laplacian variable to the total J-divergence between two states. Experimental results on real MI-BCI EEG data demonstrate that the Laplacian denoising improves the separation of motor imagery and resting mental states, and shortens the time interval required for connectivity estimation. We conclude that the approach shows promise for the robust detection of connectivity states while being appealing for implementation in real-time BCI applications.

I. INTRODUCTION

Functional connectivity describes how brain areas mutually interact [1]. This information can be modeled as a graph, which is one of the most common formalism to characterize networked data [2], [3]. Many recent studies prove that mental states can be characterized by graph statistics, such as node strength, efficiency, and modularity [4]. Detecting brain connectivity-related features corresponding to different

mental states can enhance several technologies, such as brain-computer interfaces (BCIs). BCI systems allow subjects to communicate and interact without peripheral neuro-muscular activity [5]. The requirement for the BCI functioning is therefore the correct detection of the user's mental states. While research on the subject has significantly advanced over the last decade, there is still a key limitation known as BCI inefficiency [6]. It refers to the fact that there is a percentage of users who cannot be trained to use the interface. This limitation, together with system-user interaction problems [6], motivated us to develop new tools with the intent of having a more robust estimate of brain connectivity with the final goal of better separating two cognitive states. Implementing this estimate from signals acquired at graph nodes (e.g. EEG electrodes) is a difficult task because of the inherent noise, the number of links to estimate, the presence of artifacts, the non-stationarity of the signal.

To address the problem of connectivity estimation together with the improvement of separability between mental states to optimally control a BCI, it is necessary to combine tools from different fields, such as neuroscience and signal processing. For example, graph signal processing (GSP) can be applied in this scenario [7]–[9]. GSP has already been used to deal with biological data, and in particular brain data [10], [11]. Indeed, GSP is potentially able to integrate information regarding brain structure, as represented by the graph itself, with information regarding brain function, as represented by the graph signals.

Another helpful tool to the brain connectivity estimation problem is the signal detection theory. Detection procedures can be applied to investigate statistical differences between the brain connectivity features of two different mental states, which corresponds to motor imagery and resting state for our applications. In this context, widely adopted statistics are the Likelihood Ratio (LR) of the features [12]–[19] as well as the linear detector maximizing the so-called *deflection* [20]–[23]. With the aim of obtaining a distance metric of features under two states, the case of normally distributed observations simplifies the analysis. Indeed, for normally distributed observations with equal conditional variance and different conditional means, the maximum deflection test coincides with the LR test. Moreover, this latter can be extended to a linear quadratic detector so as to cope with observations characterized by different conditional variances. [20], [21]. To obtain a measure of separability between features under the two states, one possibility is the Jensen divergence which reflects the maximum deflection test performance [12], [14], [19].

Corresponding author: Tiziana Cattai (tiziana.cattai@uniroma1.it)

T. Cattai, F. De Vico Fallani and M.-C. Corsi are with Inria Paris, Aramis Project Team, Paris, France. T. Cattai, F. De Vico Fallani and M.-C. Corsi are with Institut du Cerveau et de la Moelle épinière, ICM, Inserm U 1127, CNRS UMR 7225, Sorbonne Université, Paris, France. T. Cattai, S. Colonnese, G. Scarano are with Dept. of Information Engineering, Electronics and Telecommunication, Sapienza University of Rome, Italy. D. S.Bassett is with Department of Bioengineering, University of Pennsylvania, Philadelphia, PA, 19104, USA. D. S.Bassett is with Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA, 19104, USA. D. S.Bassett is with Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA, 19104, USA. D. S.Bassett is with Department of Neurology, Hospital of the University of Pennsylvania, Philadelphia, PA, 19104, USA. D. S.Bassett is with Department of Psychiatry, Hospital of the University of Pennsylvania, Philadelphia, PA 19104, USA. D. S.Bassett is with the Santa Fe Institute, Santa Fe, NM, 87501 USA

In the following sub-section, we present our original contributions.

Paper Contributions

- This paper proposes a novel graph Laplacian denoising algorithm, to enhance the accuracy of brain connectivity estimates. In recent literature, several studies have been conducted to improve the accuracy of link estimation, whereas few studies approach this problem in terms of description of the graph algebraic structure (see Section II). We address this limitation by proposing a subspace-based Laplacian denoising algorithm that preserves relevant connectivity features while rejecting noise-dominated components. In particular, the Laplacian denoising preserves i) the sub-spaces more directly related to the graph topology, as summarized by the eigenvectors corresponding to the smallest Laplacian eigenvalues, and ii) the sub-spaces estimated under a favourable signal-to-noise ratio, as summarized by the eigenvectors corresponding to the largest Laplacian eigenvalues. The noise rejection obtained by this twofold sub-space selection notably improves the separability of two connectivity states. To sake of clarity, we refer to connectivity states as the patterns of connectivity estimated while the brain performs distinct cognitive tasks.
- In order to measure the improvement achieved by the proposed brain connectivity denoising algorithm, we provide an analysis of the J-divergence of the Laplacian coefficients, we explicit their contribution to the states' separability in terms of their first and second order moments of the test statistics, and we show that the proposed Laplacian denoising actually increases the J-divergence of the brain connectivity features *rest (null)* or *motor imagery (alternative)*. The improvement of the J-divergence of the graph Laplacian coefficients under different connectivity states is assessed by numerical simulations on synthetic data.
- Finally, we present experimental results on real EEG data acquired during motor imagery-based BCI experiments, and we prove that the proposed novel denoising algorithm increases the J-divergence of brain connectivity states and paves the way for connectivity estimation time interval reduction. As a relevant by-product of the theoretical J-divergence analysis, we are able to attribute a score to each and every Laplacian coefficient representing its marginal contribution to the J-divergence. The score admits relevant biological interpretation confirming the efficacy of the approach. These results can be assessed by further studies on the brain connectivity features.

The structure of the paper is as follows. Section II reviews the scientific literature related to our work. Section III describes the signal model used in the analysis. Section IV details the novel graph denoising we propose. Section V describes the problem of Gaussian detection, and it presents the novel formulation of J-divergence we use throughout the paper. In section VI, we test our filtering method on synthetic graph to verify its ability to separate two graphs. Section VII applies our method on real EEG data, exhibiting its capacity to

Notation	Description
$\mathbf{A}, \hat{\mathbf{A}}$	adjacency matrix (real, estimated)
\mathbf{V}	set of all nodes
N	total number of nodes
\mathbf{E}	set of all links
$\mathbf{D}, \hat{\mathbf{D}}$	degree matrix (real, estimated)
$\mathbf{L}, \hat{\mathbf{L}}$	Laplacian matrix (real, estimated)
$\lambda, \hat{\lambda}$	eigenvalue (real, estimated)
$\mathbf{u}, \hat{\mathbf{u}}$	eigenvector (real, estimated)
$\mathcal{U}_{\mathcal{L}}, \mathcal{U}_{\mathcal{M}}, \mathcal{U}_{\mathcal{H}}$	subset of smallest, central, larger eigenvalues
$\hat{\mathbf{L}}, \hat{\mathbf{I}}$	filtered graph laplacian matrix and vector
\mathbf{T}	transformation matrix
\mathbf{x}	vectorized laplacian in the transformed domain
J	J-divergence
\mathbf{S}	score

TABLE I
TABLE OF MAIN NOTATION.

estimate graph connectivity during motor imagery tasks and to discriminate between two mental states. We conclude in section VIII. In Table I, we list of the main notation used in the paper.

II. RELATED WORK

The problem of graph connectivity estimation has been well studied in literature in different domains, from neuroscience to signal processing and graph theory [24]–[26]. State-of-the-art graph learning methods have the limitation that they usually present over-simplified models for the signal on graph to overcome problems of computational and memory cost. Some recent works, such as [27], propose different strategies to deal with graph learning problems. Specifically, in the context of mental state identification, authors in [27] present a novel technique to create and modify embeddings associated to each graph node to efficiently compute the adjacency matrix. Since FC computation requires a lot of time and computational power, one possibility consists in clustering FC into relevant communities of synchronous components. One approach, recently proposed in [28], goes in this direction with the application of k-means clustering algorithm followed by a tensor decomposition to reduce the FC data.

FC estimation can leverage the generalization of classical signal processing operations into the graph setting, where signals are localized on graph nodes, giving rise to novel research domain of the graph signal processing (GSP) [7]–[9]. GSP has already showed its potential to describe brain functioning in [29] and [10]. Indeed, GSP representation naturally fits to the brain, where the structure can be described by the graph itself while brain functioning directly corresponds to graph signals. An interesting application is graph filtering [7], [30] which can be useful to extract meaningful brain behaviour [31]. In [32], authors propose a mathematical model for brain fibers able to describe neurophysiological mechanisms. The model, based on GSP techniques, extracts a subset of graph eigenvectors which represent a suitable basis for filtering fiber tracts from brain imaging data.

GSP techniques have been applied also to brain-computer interfaces with NIRS signals [33]. Specifically, GSP analysis is leveraged in [33] in the context of feature extraction to

extract spatial information from the NIRS signals and it has been shown to improve classification performances.

Classical signal processing techniques and eigenvector-based filtering have already been used with brain data [34], [35]. In [36] and [37], eigenvector-based filters are applied to fetal magnetic signals and diffuse optic imaging data to obtain more localized activities and reduce artifacts and noise. Specifically, in [37], classical eigenvector-filtering, i.e. based on larger eigenvectors, is used in diffuse optical imaging with the aim to improving connectivity estimation.

In the following, stemming from the GSP approach to FC estimation, we propose a novel Laplacian denoising algorithm, and we show that it improves the statistical separation of distinguished connectivity states. To this aim, we provide an analysis of the J-divergence, which naturally provides a metric to quantify the distance between two distributions, for the problem under concern. Recently, the J-divergence has been applied in [38] to investigate the time series' irreversibility. Another recent application of the J-divergence is proposed in [39], where authors present a novel approach to vector-skew the J-divergence. This method is able to preserve J-divergence properties and simultaneously to fine tune parameters for specific applications.

J-divergence has been also applied in the context of BCI design, to tackle one of the most challenging issues of EEG-based BCIs, which is the long calibration time. In general, the number of data required to calibrate the model is really high, because of the noise and the non-stationarity of brain signals. One solution comes from [40], where a subject-to-subject transfer learning is proposed to improve the classification performance with limited training data. J-divergence is used in a transfer learning framework to test their method by comparing the data of the target subject with the data from previous subjects. In the following, we investigate the J-divergence under a different points of view, namely i) we assess the performance of the denoising algorithm in separating connectivity states and ii) we provide criterion for scoring the Laplacian coefficients based on their contribution to the connectivity states separation.

III. SIGNAL MODEL

We are interested in analyzing signals defined on an undirected, connected, weighted graph $G = \{V, E, \mathbf{A}\}$, which consists of a finite set of vertices V with $|V| = N$, a set of edges E and a weighted adjacency matrix \mathbf{A} . If there is an edge $e = (i, j)$ connecting vertices i and j , the element $A_{i,j}$ represents the weight of the edge; otherwise, $A_{i,j} = 0$.

The graph Laplacian, is a real symmetric matrix defined¹ as:

$$\mathbf{L} = \mathbf{D} - \mathbf{A} \quad (1)$$

where the degree matrix \mathbf{D} is a diagonal matrix whose i^{th} diagonal element d_i is equal to the sum of the weights of all the edges incident to vertex i . We denote by $\{\mathbf{u}_i\}_{i=0,1,\dots,N-1}$ set of orthonormal eigenvectors, corresponding to increasingly ordered eigenvalues $0 = \lambda_0 \leq \lambda_1 \leq \lambda_2 \dots \leq \lambda_{N-1} = \lambda_{max}$.

¹We refer here to the non-normalized graph Laplacian, also called the combinatorial Laplacian.

In GSP, the Laplacian eigenvectors are considered as SoGs and provide a basis for the Graph Fourier Transform. For a SoG \mathbf{s} , the GFT is defined as the projection of \mathbf{s} on the eigenvectors of the graph Laplacian: $\hat{\mathbf{s}}(\lambda_l) = \mathbf{s}^H \mathbf{u}_l$. The graph Laplacian eigenvalues $\lambda_l, l = 0, \dots, N-1$ have an analogous meaning to Fourier transform frequency, i.e. smaller eigenvalues are associated to eigenvectors that exhibit smoother variations over connected nodes.

In many SoG application problems, including brain functional connectivity estimation, signal values are actually represented by discrete sequences, obtained by sampling a continuous time signal at each graph node.

Let us denote the sequences of samples acquired over an observation period T_{oss} with sampling pace T_s as $y_n[kT_s]$, $n = 0, \dots, N-1, k = 0, \dots, N_s, N_s = \lfloor T_{oss}/T_s \rfloor$, or in vector form as $\mathbf{y}[kT_s] = [y_0[kT_s] \dots y_{N-1}[kT_s]]$. The vector sequence $\mathbf{y}[kT_s]$, $k = 0, \dots, N_s$ is used to estimate the graph adjacency matrix \mathbf{A} by computing a similarity metric on each and every node pair. There are many state-of-the-art methods to estimate $A_{i,j}$, $i, j = 0, \dots, N-1$, which associate link weights according to different interaction properties [1], [41]–[43]. Thereby, the adjacency matrix \mathbf{A} is actually represented by its estimated version $\hat{\mathbf{A}}$, which contains the connectivity values $\hat{A}_{i,j}$ estimated for each graph node pair $(i, j), i, j = 0, \dots, N-1$. Accordingly, the estimated degree matrix $\hat{\mathbf{D}}$ is computed, so as to derive the estimated laplacian $\hat{\mathbf{L}}$ through Eq. 1, that becomes here:

$$\hat{\mathbf{L}} = \hat{\mathbf{D}} - \hat{\mathbf{A}} \quad (2)$$

Any estimation error on the adjacency matrix affects the Laplacian estimate, and it results into less distinguishable connectivity states. In the following section we address the denoising of the estimated Laplacian for the purpose of improving the separation of connectivity states.

IV. GRAPH CONNECTIVITY DENOISING

In order to introduce the Laplacian denoising algorithm, we consider the eigenvalue decomposition of the estimated Laplacian matrix $\hat{\mathbf{L}}$ as follows:

$$\hat{\mathbf{L}} = \sum_{i=0}^{N-1} \hat{\lambda}_i \hat{\mathbf{u}}_i \hat{\mathbf{u}}_i^H \quad (3)$$

Perturbations affect graph Laplacian estimation in terms of both eigenvalues and/or eigenvectors. To elaborate on the effect of perturbations, we explicit the first, second and third order error contributions to the estimated laplacian $\hat{\mathbf{L}}$ as:

$$\begin{aligned} \hat{\mathbf{L}} &= \sum_{i=0}^{N-1} (\lambda_i + \epsilon_{\lambda_i})(\mathbf{u}_i + \epsilon_{u_i})(\mathbf{u}_i + \epsilon_{u_i})^H \\ &= \sum_{i=0}^{N-1} \underbrace{\lambda_i \mathbf{u}_i \mathbf{u}_i^H}_{\mathbf{L}} + \underbrace{\lambda_i \mathbf{u}_i \epsilon_{u_i}^H + \lambda_i \epsilon_{u_i} \mathbf{u}_i^H + \epsilon_{\lambda_i} \mathbf{u}_i \mathbf{u}_i^H}_{\text{first order error}} \\ &\quad + \underbrace{\lambda_i \epsilon_{u_i} \epsilon_{u_i}^H + \epsilon_{\lambda_i} \mathbf{u}_i \epsilon_{u_i}^H + \epsilon_{\lambda_i} \epsilon_{u_i} \mathbf{u}_i^H}_{\text{second order error}} + \underbrace{\epsilon_{\lambda_i} \epsilon_{u_i} \epsilon_{u_i}^H}_{\text{third order error}} \end{aligned} \quad (4)$$

Thereby, the estimated Laplacian can be approximated at the first order as the sum of N terms:

$$\hat{\mathbf{L}} \approx \sum_{i=0}^{N-1} (\lambda_i + \epsilon_{\lambda_i}) \mathbf{u}_i \mathbf{u}_i^H + \lambda_i (\mathbf{u}_i \epsilon_{u_i}^H + \epsilon_{u_i} \mathbf{u}_i^H) \quad (5)$$

Eq.(5) highlights the first order error contribution due to relative perturbation of the Laplacian eigenvalues as well as of the eigenvectors direction. We are interested in the Laplacian components whose perturbation is contained because either the relative eigenvalue perturbation $\epsilon_{\lambda_i}/\lambda_i$ or the eigenvector perturbation ϵ_{u_i} is (relatively) small. To this aim, we order the set of orthonormal eigenvectors $\mathcal{U}_{ALL} = \{\hat{\mathbf{u}}_l, l = 0, 1, \dots, N-1\}$ with increasingly eigenvalues $0 = \hat{\lambda}_0 \leq \hat{\lambda}_1 \leq \hat{\lambda}_2 \dots \leq \hat{\lambda}_{N-1} := \hat{\lambda}_{max}$, and we consider three subsets of eigenvalues and associated eigenvectors: 1) the subset \mathcal{U}_L containing the N_L smallest eigenvalues; 2) the subset \mathcal{U}_H containing the N_H largest eigenvalues; and 3) the subset \mathcal{U}_M containing the remaining $N_M = N - N_L - N_H$ central eigenvalues, with $\mathcal{U}_L \cup \mathcal{U}_M \cup \mathcal{U}_H = \mathcal{U}_{ALL}$.

Firstly, we remark that the N_H largest eigenvalues are more robust to eigenvalue perturbation; this assumption is usually exploited in classical signal processing, where the subspace \mathcal{U}_H is used for the estimation of the covariance matrix because of its favorable signal-to-noise ratio [44].

Secondly, stemming on recent literature results [45], it can be expected that the subspace \mathcal{U}_L is partially robust in terms of eigenvector perturbation. In fact, in [45] the authors states that a connectivity estimation error on the A_{mn} adjacency matrix element, i.e. on the weight of the link between the m -th and the n -th nodes, induces a perturbation ϵ_{u_i} of the i -th eigenvector depending on the difference between the m -th and the n -th coefficients of \mathbf{u}_i . Thereby, eigenvectors smoothly varying across the m -th and the n -th nodes are less affected by estimation errors on A_{mn} . On the other hand, in GSP, it is well known that \mathcal{U}_L eigenvectors, corresponding to small eigenvalues, represent low frequency basis elements in the Graph Fourier Transform [46], [9] since they are characterized by the smallest variations across strongly connected graph regions. Thereby, the eigenvectors in \mathcal{U}_L are equipped with inherent resilience to connectivity estimation error within these regions. On the other hand, the eigenvectors in \mathcal{U}_L are tightly related to the network connectivity, and therefore they need to be involved in the devised denoising method.

Stemming on these observations, we propose a denoising method based on preserving the contribution to the Laplacian due to the subspaces $\mathcal{U}_L, \mathcal{U}_H$ while discarding those relative to the subspace \mathcal{U}_M . In formulas, given the estimated Laplacian

$$\hat{\mathbf{L}} = \sum_{i \in \mathcal{U}_L \cup \mathcal{U}_M \cup \mathcal{U}_H} \hat{\lambda}_i \hat{\mathbf{u}}_i \hat{\mathbf{u}}_i^H \quad (6)$$

we compute the denoised Laplacian $\tilde{\mathbf{L}}$ as follows:

$$\tilde{\mathbf{L}} = \sum_{i \in \mathcal{U}_L} \hat{\lambda}_i \hat{\mathbf{u}}_i \hat{\mathbf{u}}_i^H + \sum_{i \in \mathcal{U}_H} \hat{\lambda}_i \hat{\mathbf{u}}_i \hat{\mathbf{u}}_i^H \quad (7)$$

We recognize that the proposed, subspace-based, Laplacian denoising approach allows preservation of

- the sub-space \mathcal{U}_L , which is directly related to the graph connectivity features

- the sub-space \mathcal{U}_H , which is estimated with a favourable signal-to-noise ratio.

The proposed Laplacian denoising method, synthetically presented in Algorithm 1, preserves the information relevant for the purpose of graph connectivity identification, while rejecting noisy components. In order to quantify the improvement achieved in terms of connectivity state separability, we resort to the J-divergence as a metric of the distance between connectivity states. In the following we derive a formulation the J-divergence for the problem under concern.

V. JENSEN DIVERGENCE OF CONNECTIVITY STATES

Several metrics can be adopted to determine the separability of two connectivity states [47], as represented by the Laplacian matrix L . Herein, we resort to the notion of J-divergence for characterizing the separability of connectivity states, and we reformulate it for the problem under concern. Thus, J-divergence is used to identify the Laplacian coefficients that are most relevant for detection purposes and later on to measure the improvement achieved by the proposed denoising algorithm.

For the purpose of the analysis, we will assume that the Laplacian coefficients obtained at the output of the denoising algorithm are normally distributed. Let us remark that the Gaussian assumption stands in many applications², including the case of connectivity estimates carried out on real brain signals, and thereby it is often assumed in the literature, e.g. for Laplacian estimation purposes [7]. Specifically, we assume that the vector $\tilde{\mathbf{l}} = \text{Vec}(\tilde{L})$ is distributed according to a multidimensional Gaussian probability whose mean vector and covariance matrix differ under two different connectivity states, referred to as the *null* and the *alternative* hypotheses $\mathcal{H}_0, \mathcal{H}_1$ in the following:³

$$\begin{cases} \mathcal{H}_0 : \tilde{\mathbf{l}} \sim \mathcal{N}(\boldsymbol{\eta}_0, \mathbf{K}_0) \\ \mathcal{H}_1 : \tilde{\mathbf{l}} \sim \mathcal{N}(\boldsymbol{\eta}_1, \mathbf{K}_1) \end{cases} \quad (8)$$

As an information theoretic measure of distance between $\tilde{\mathbf{l}}$ under \mathcal{H}_0 and \mathcal{H}_1 , we now compute in analytical form the J-divergence, which is defined as the expected value of the difference of the Log Likelihood Ratio under the two hypothesis \mathcal{H}_0 and \mathcal{H}_1 [14]. The J-divergence formulation will allow us to evaluate to which extent the connectivity states represented by the Laplacian coefficients are distinguishable from each other.

Let us first assume that the Laplacian moments $\boldsymbol{\eta}_0, \boldsymbol{\eta}_1, \mathbf{K}_0, \mathbf{K}_1$ are known. Detection can then be conducted on a linear transformation of the observations:

$$\mathbf{x} = \mathbf{T} (\tilde{\mathbf{l}} - \boldsymbol{\eta}_0)$$

²The reason why this occurs is that the Gaussian assumption tightly models laplacian diagonal elements, computed in each row as the sum of extradiagonal elements in that column, as well as extradiagonal elements which are often computed as the result of correlation estimates.

³The notation $\tilde{\mathbf{l}} \sim \mathcal{N}(\boldsymbol{\eta}_j, \mathbf{K}_j)$, with $j \in \{0, 1\}$ indicates that the random vector $\tilde{\mathbf{l}}$ is Gaussian distributed with mean vector $\boldsymbol{\eta}_j$ and covariance matrix \mathbf{K}_j .

where $\boldsymbol{\eta} \stackrel{\text{def}}{=} \mathbf{T}(\boldsymbol{\eta}_1 - \boldsymbol{\eta}_0)$ and $\mathbf{T} = \mathcal{T}(\mathbf{K}_0, \mathbf{K}_1)$ is an affine transform that simultaneously⁴ whitens the observations in the \mathcal{H}_0 hypothesis obtains uncorrelated observations in the \mathcal{H}_1 hypothesis. An example of the action of the transform T , is shown in Fig. 1 for the case of 2-dimensional Gaussian data whose mean and covariance matrix differ under the $\mathcal{H}_0, \mathcal{H}_1$. The original data points are plotted in Fig. 1(a), whereas their transformed counterparts are represented in Fig. 1(b). The transformed data are unitary variance, zero-centered under \mathcal{H}_0 and are uncorrelated under \mathcal{H}_1 .

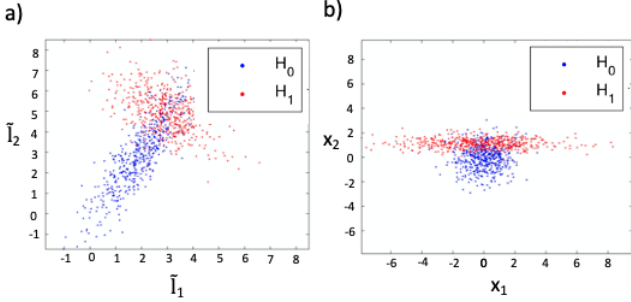


Fig. 1. Example of transformation effect. In a) we have 2-dimensional Gaussian distribution which differ under mean and covariance matrix. In b) we report the same distributions after the T transformation

In real detection systems, the moments $\boldsymbol{\eta}_0, \boldsymbol{\eta}_1, \mathbf{K}_0, \mathbf{K}_1$ can be either estimated from a training set, e.g. during a BCI training phase, or coarsely initialized and tracked throughout the system life, using methodologies, priors, and heuristics related to the application-specific problem under concern [48]–[50]. Besides, the transformed data \mathbf{x} can be obtained even avoiding computation of the moments and of T , by applying the laplacian coefficients $\tilde{\mathbf{I}}$ to a suitably trained network [51], that will enforce the afore-mentioned statistical constraints.

With these position, the observation model becomes:

$$\mathcal{H}_0 : \mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \text{ versus } \mathcal{H}_1 : \mathbf{x} \sim \mathcal{N}(\boldsymbol{\eta}, \boldsymbol{\Sigma}^2) \quad (9)$$

The J-divergence is then defined as:

$$J \stackrel{\text{def}}{=} \mathbb{E}\{\mathcal{R}(\mathbf{x})|\mathcal{H}_1\} - \mathbb{E}\{\mathcal{R}(\mathbf{x})|\mathcal{H}_0\} \quad (10)$$

being $\mathcal{R}(\mathbf{x})$ the Log-Likelihood Ratio⁵:

$$\mathcal{R}(\mathbf{x}) = \mathbf{x}^H (\mathbf{I} - \boldsymbol{\Sigma}^{-2}) \mathbf{x} + 2\boldsymbol{\eta}^H \boldsymbol{\Sigma}^{-2} \mathbf{x} \quad (11)$$

Let us now associate the variables x_n whose variance $\sigma_n^2 \neq 1$ to the first P indexes and the remaining ones to the indexes $n = P + 1, \dots, N$ ⁶ so as to rewrite the LLR as follows:

⁴The matrix \mathbf{T} and the diagonal matrix $\boldsymbol{\Sigma}^2 \stackrel{\text{def}}{=} \text{diag}(\sigma_1^2, \dots, \sigma_n^2, \dots, \sigma_N^2)$ are computed as the *generalized eigenvectors* and the *generalized eigenvalues* matrices of the pencil $(\mathbf{K}_1, \mathbf{K}_0)$, respectively. Given any square root \mathbf{Q}_0 of \mathbf{K}_0^{-1} , i.e. such that $\mathbf{Q}_0^H \cdot \mathbf{K}_0 \cdot \mathbf{Q}_0 = \mathbf{I}$, we may conveniently employ the unitary transformation \mathbf{V}_1 obtained from the eigenanalysis $\mathbf{Q}_0^H \cdot \mathbf{K}_1 \cdot \mathbf{Q}_0 = \mathbf{V}_1 \cdot \boldsymbol{\Lambda}_1 \cdot \mathbf{V}_1^H$; in fact, it is easily proved that the matrix $\mathbf{T} = \mathbf{V}_1 \cdot \mathbf{Q}_0^H$ verifies $\mathbf{T} \cdot \mathbf{K}_0 \cdot \mathbf{T}^H = \mathbf{I}$; $\mathbf{T} \cdot \mathbf{K}_1 \cdot \mathbf{T}^H = \boldsymbol{\Sigma}^2$ with $\boldsymbol{\Lambda}_1 = \boldsymbol{\Sigma}^2$.

⁵The Log-Likelihood Ratio $\mathcal{R}(\mathbf{x})$ is widely adopted classical LLR detection: $\mathcal{R}(\mathbf{x}) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \theta$, being θ selected according to the desired detection versus missing probability tradeoff.

⁶Possibly, we might have $P = N$ or $P = 0$.

$$\begin{aligned} \mathcal{R}(\mathbf{x}) &= \sum_{n=1}^N \frac{1}{\sigma_n^2} [(\sigma_n^2 - 1) x_n^2 + 2\eta_n \cdot x_n] \\ &= \sum_{n=1}^P \frac{(\sigma_n^2 - 1) |x_n|^2 + 2\eta_n \cdot x_n}{\sigma_n^2} \\ &\quad + \sum_{n=P+1}^N 2\eta_n^* \cdot x_n \end{aligned} \quad (12)$$

By adding and subtracting the term $|\eta_n^2|/[\sigma_n^2(\sigma_n^2 - 1)]$ we rearrange the summation (12) as:

$$\begin{aligned} \mathcal{R}(\mathbf{x}) &= \underbrace{\sum_{n=1}^P \frac{\sigma_n^2 - 1}{\sigma_n^2} \left| x_n + \frac{\eta_n}{\sigma_n^2 - 1} \right|^2}_{P \text{ quadratic terms}} + \underbrace{\sum_{n=P+1}^N 2\eta_n \cdot x_n}_{N-P \text{ linear terms}} \\ &\quad - \underbrace{\sum_{n=1}^P \frac{|\eta_n^2|}{\sigma_n^2 (\sigma_n^2 - 1)}}_{\text{constant to be included in the threshold}} \end{aligned} \quad (13)$$

The P variates x_n , $n = 1, \dots, P$, having different conditional variances under the hypotheses $\mathcal{H}_0, \mathcal{H}_1$, contribute to the LLR by the P terms quadratic terms. The $N - P$ variates x_n , $n = P + 1, \dots, N$, having equal unitary conditional variances under the hypotheses $\mathcal{H}_0, \mathcal{H}_1$, contribute to the LLR by the $N - P$ linear terms. To gain further insight on the J-divergence, we resort to the following theorem, whose demonstration is reported in the Appendix.

THEOREM 1 Let $\boldsymbol{\xi}$ be a vector formed by the N statistically independent random variables:

$$\begin{aligned} \xi_n &= \left(x_n + \frac{\eta_n}{\sigma_n^2 - 1} \right)^2, \quad n = 1, \dots, P \\ \xi_n &= x_n, \quad n = P + 1, \dots, N \end{aligned} \quad (14)$$

The LLR is expressed as $\mathcal{R}(\mathbf{x}) = \mathbf{a}_{\text{LLR}}^H \cdot \boldsymbol{\xi}$ being \mathbf{a}_{LLR} constant coefficients defined as in Eq.(25) and the J-divergence in Eq.(10) is computed as follows:

$$\begin{aligned} J &= \sum_{n=1}^P (\sigma_n - \sigma_n^{-1})^2 \left[1 + \frac{|\eta_n^2|}{\sigma_n (\sigma_n - \sigma_n^{-1})^2} \right] + \sum_{n=P+1}^N 2|\eta_n^2| \\ &= \sum_{n=1}^P J_n^{(\sigma, \eta)} + \sum_{n=P+1}^N J_n^{(\eta)} \end{aligned} \quad (15)$$

Theorem 1 generalizes the result in [13], [14] where only the case of variables having equal conditional means and different covariances (i.e. $\boldsymbol{\eta}_1 = \boldsymbol{\eta}_0$, $\mathbf{K}_1 \neq \mathbf{K}_0$) has been addressed.

The J-divergence as formulated in Eq.(15) is a measure of the statistical distance of the graph Laplacian coefficients under two connectivity states \mathcal{H}_1 and \mathcal{H}_0 , and it will be used to quantify the improvement of separability of brain connectivity states achieved by the denoising algorithm described in section IV.

Furthermore, the above analysis sheds a light upon the variables that mostly contribute to the states separability. From Eq.(15), we see a one-to-one correspondence between

the transformed space variables x_n and the terms of the J-divergence J ; besides, the term can be of two kinds

$$J_n^{(\sigma, \eta)} = (\sigma_n - \sigma_n^{-1})^2 \left[1 + \frac{|\eta_n^2|}{\sigma_n} \frac{\sigma_n + \sigma_n^{-1}}{(\sigma_n - \sigma_n^{-1})^2} \right] \quad (16)$$

$$J_n^{(\eta)} = 2|\eta_n^2|$$

depending on whether the variable changes both in conditional mean and standard deviation, or in conditional mean only. The functions $J_n^{(\sigma, \eta)}$, $J_n^{(\eta)}$ are plotted in Fig.2 for η between 0 and 1 and σ^{-1} from 0 to 10. Interestingly, in Fig.2 we recognize

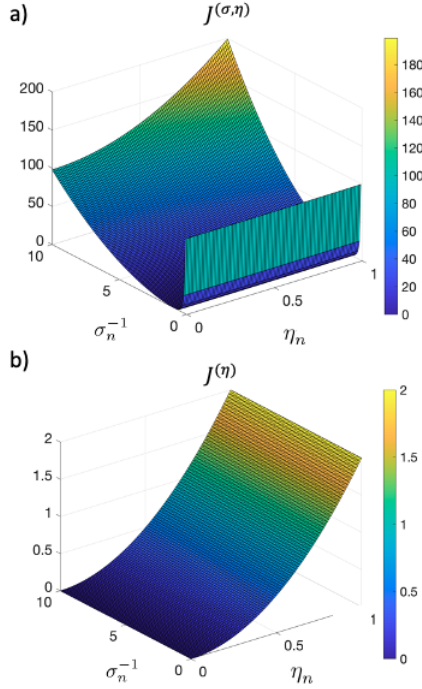


Fig. 2. J-Divergence contributions as function of mean η and standard deviation σ : a) $J^{(\sigma, \eta)}$ for variables whose conditional standard deviation differ under \mathcal{H}_1 and \mathcal{H}_0 , and b) $J^{(\eta)}$ for variables with invariant conditional standard deviation.

that a conditional variance change gives a higher contribution to J than an equal conditional mean change.

To sum up, the above analysis highlights the contribution of each of transformed variable to the separability of the connectivity states, and allows to rank their relevance to J_n , so as to identify the transformed variables which mostly differ under the two hypothesis. This paves the way for an information theoretic scoring of the Laplacian coefficients, described in the following.

A. J-Divergence based Laplacian coefficients scoring

As a by-product of the analysis, we are now able to identify which Laplacian coefficients (i.e. links weights or nodes degrees), contribute mostly to the connectivity states separability. This is obtained by attributing a score to the Laplacian coefficients measuring their contribution to the J-divergence J .

By definition, the Laplacian coefficient $\tilde{l}_{\bar{n}}$, $\bar{n} = 0, \dots, N - 1$, contribute to each transformed variable x_n . Thereby, we

introduce a score $S_{\bar{n}}$, evaluated by suitable backpropagation of the J_n terms on each contributing Laplacian coefficient. Specifically, the \bar{n} -th coefficient score is computed as

$$S_{\bar{n}} = \sum_n J_n \cdot \frac{t_{n\bar{n}}}{\sum_k t_{nk}} \quad (17)$$

where we recognize that the weight $t_{n\bar{n}}$ representing the contribution of the \bar{n} -th Laplacian coefficient to the n -th transformed variable is normalized with respect to the sum $\sum_k t_{nk}$ of the weights of all the contributing coefficients.

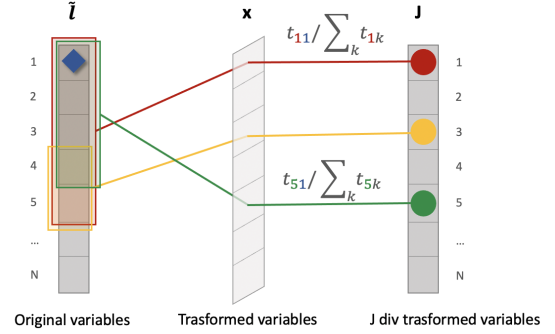


Fig. 3. Graphic interpretation of the score computed for the first element in the vector \tilde{l}

A graphical interpretation of the score is provided in Fig.3, where we represent the set of variables \tilde{l} belonging to the original domain (left), the set of variables x belonging to the transformed domain (center) and the corresponding marginal contributions to J (right). The relationship between \tilde{l} and x is given by the transformation matrix T that blends variables from the original domain to the transformed one. Each J-divergence component J_n (colored circle on the right) is associated to the variable x_n in the transformed domain, which in turn is originated by many \tilde{l}_n (shaded colored box on the left). Thereby, J_n is backprojected to the original domain by weighting its contribution as in Eq. 17. Back-projection and accumulation can also be applied by limiting the summation in Eq.(17) to the largest ranking J_n terms. The score computation allows to quantify the relevancy of the Laplacian coefficients $\tilde{l}_{\bar{n}}$ for state separability, and in the experimental results we show that it leads to meaningful results in case of real BCI data. The Algorithm 2 review the main steps of the J-divergence computation and scoring procedures.

VI. RESULTS ON SYNTHETIC DATA

In this section, we test the performance of the Laplacian denoising presented in section IV in improving the J-divergence of two estimated connectivity states over synthetic SoGs. To this end, we first consider a graph and a model for signals at nodes under two connectivity states, selected to represent an over-simplified model of brain EEG signals functional connectivity; real brain signals are considered in the next section.

We compare our approach with the case of laplacian without filtering (that we refer as \mathcal{U}_{ALL} and with several of

Algorithm 1 Graph Laplacian denoising

Input: Estimated Laplacian \tilde{L}
Output: Denoised Laplacian \tilde{L}

- 1: Compute the eigen-decomposition

$$\hat{L} = \sum_{i=0}^{N-1} \hat{\lambda}_i \hat{\mathbf{u}}_i \hat{\mathbf{u}}_i^H$$

- 2: Compute the denoised Laplacian
- \tilde{L}
- by

- a: Selecting the number
- N_L
- of smallest eigenvalues and the number
- N_H
- of largest eigenvalues to retain
-
- b: Computing

$$\tilde{L} = \sum_{i=0}^{N_L-1} \hat{\lambda}_i \hat{\mathbf{u}}_i \hat{\mathbf{u}}_i^H + \sum_{i=N-N_H}^{N-1} \hat{\lambda}_i \hat{\mathbf{u}}_i \hat{\mathbf{u}}_i^H$$

Algorithm 2 J-divergence and score computation

Input: Conditional means μ_0, μ_1 and covariance matrices K_0, K_1 of $\tilde{\mathbf{l}} = \text{Vec}(\tilde{\mathbf{l}})$ under \mathcal{H}_1 and \mathcal{H}_0
Output: $J_n, S_n, n = 0, \dots, N-1$

- 1: Step 1: Transform computation

- a: Compute the square root matrix

$$Q_0 \leftarrow K_0^{-1/2}$$

and the eigenvectors V_1 and the eigenvalues $\sigma_0^2, \dots, \sigma_{N-1}^2$ of the eigen-decomposition

$$Q_0^H K_1 Q_0 = V_1 \text{diag}(\sigma_0^2, \dots, \sigma_{N-1}^2) V_1^H$$

- b: Compute
- T
- as
- $\leftarrow V_1^H Q_0^H$
- and
- $\Sigma \leftarrow \sqrt{\text{eig}(Q_0^H K_1 Q_0)'}^H$

- 2: Step 2: J-divergence computation

- a: Define a threshold
- θ

- b: Compute
- $J_n, n = 0, \dots, N-1$
- as

$$J_n \leftarrow \begin{cases} 2|\eta_n^2|, & \iff (\sigma_n^2 > \theta) \cup (|\sigma_n^2 - 1| > \theta) \\ (\sigma_n - \sigma_n^{-1})^2 \left[1 + \frac{|\eta_n^2|}{\sigma_n} \frac{\sigma_n + \sigma_n^{-1}}{(\sigma_n - \sigma_n^{-1})^2} \right], & \text{otherwise} \end{cases}$$

- 3: Step 3: Score computation

- a: Compute
- $S_{\bar{n}}, \bar{n} = 0, \dots, N-1$
- as

$$S_{\bar{n}} = \sum_n J_n \cdot \frac{t_{n\bar{n}}}{\sum_k t_{nk}}$$

eigenvector-based filters, i.e. $\mathcal{U}_L, \mathcal{U}_H$). Then, we explain in detail the analysis we performed and the related results.

A. Signal on Graph generation and connectivity estimation

In order to validate our framework on synthetic data, we define signals under the two hypothesis \mathcal{H}_0 and \mathcal{H}_1 to obtain two distinct graph connectivity states.

Under \mathcal{H}_1 , we model the network activity by considering H scalar generator signals $s^{(h)}[kT_s], h = 0, \dots, H-1$. Each generator signal simultaneously contributes to the signals

measured over a subset $\mathcal{G}^{(h)}, h = 0, \dots, H-1$ of nodes identified by the non-zero component of the $N \times 1$ binary vector $\mathbf{g}^{(h)}, h = 0, \dots, H-1$. A noise component $\mathbf{w}[kT_s]$ and a common component across all the nodes $b[kT_s] \cdot \mathbf{1}$ are also present. Under \mathcal{H}_0 , only these latter components are observed. With these positions, we come up with the following expression for the vector of the observed signals $\mathbf{y}[kT_s]$ under the two hypothesis \mathcal{H}_1 and \mathcal{H}_0 :

$$\begin{aligned} \mathcal{H}_1 : \mathbf{y}[kT_s] &= \sum_{h=0}^{H-1} s^{(h)}[kT_s] \cdot \mathbf{g}^{(h)} + \mathbf{w}[kT_s] + b[kT_s] \cdot \mathbf{1} \\ \mathcal{H}_0 : \mathbf{y}[kT_s] &= \mathbf{w}[kT_s] + b[kT_s] \cdot \mathbf{1} \end{aligned} \quad (18)$$

In the simulations, the noise $\mathbf{w}[kT_s]$ is a realization of a discrete, stationary, white Gaussian process, with $E\{\mathbf{w}[k]\} = 0$, $E\{\mathbf{w}[k]\mathbf{w}[k]^T\} = \sigma_w^2 I \forall k$; the samples of discrete sequences $b[kT_s]$ are realizations of a zero mean Gaussian random distribution with variance σ_b^2 ; and $s^{(h)}[kT_s], h = 0, \dots, H-1$ are drawn from a zero mean unit variance Gaussian random variable.

Once SoGs samples $\mathbf{y}[kT_s]$ are obtained, we estimate the adjacency matrix. There are many state-of-the-art methods to perform the estimation, such as spectral coherence [42], imaginary coherence [43], phase-locking value which differently characterize brain interactions [52] of the signals at two nodes i, j as:

$$C_{ij}(\omega_k) = \frac{|\hat{P}_{ij}(\omega_k)|}{\sqrt{\hat{P}_i(\omega_k) \cdot \hat{P}_j(\omega_k)}} \quad (19)$$

In Eq. (19), $\hat{P}_i(\omega_k)$, $\hat{P}_j(\omega_k)$ and $\hat{P}_{ij}(\omega_k)$ are the the estimated auto-spectra and cross-spectrum of the signals $y_i[kT_s], y_j[kT_s]$ at the nodes i and j , computed at the frequency bin⁷ $\omega_k = \frac{2\pi}{N_s}k$. Given $C_{ij}(\omega_k)$ as in Eq. (19), the adjacency matrix $\hat{\mathbf{A}}$, estimated is averaging across the N_s frequency bins as follows:

$$\hat{A}_{ij} = \sum_{k=0}^{N_s-1} C_{ij}(\omega_k) \quad (20)$$

To sum up, our proposed signal model for synthetic data determines a simple graph connectivity under the two hypotheses \mathcal{H}_1 and \mathcal{H}_0 . The model successfully reproduces a network characterized by distinguishable connectivity states in presence of controlled perturbations. In Fig. 4, the estimated adjacency matrix is plotted under the two conditions \mathcal{H}_1 and \mathcal{H}_0 in presence of perturbations. We recognize that under \mathcal{H}_0 there are no links and $\hat{\mathbf{A}}$ fluctuates around zero because of the perturbations. Under \mathcal{H}_1 some connections exist but their values are affected by the perturbations.

Once we have obtained the adjacency matrix estimations under \mathcal{H}_1 and \mathcal{H}_0 , we compute the estimated Laplacians as in Eq. (2) and then, we decompose it with its eigenvalues and eigenvectors as in Eq. (3). In order to recall the eigenvectors' behaviour, we plot in Fig. 5 the first and the 10th eigenvectors on graph under \mathcal{H}_1 hypothesis. We can see that the first

⁷All the power spectral estimates are computed with Welch method, with 1s length Hanning windows and overlap of 50%.

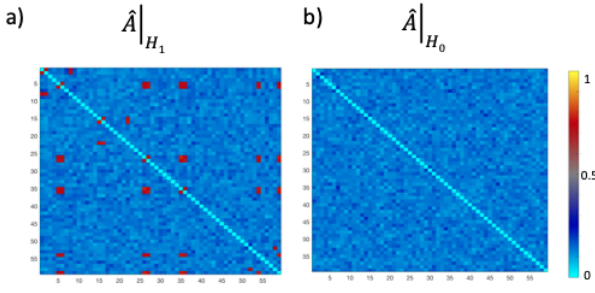


Fig. 4. Adjacency matrix with synthetic data. $\hat{A}_{ij}[k]$ is represented under \mathcal{H}_1 in a) and under \mathcal{H}_0 in panel b).

eigenvector, in Fig. 5(a) perfectly appears smooth on the graph and within a subset of linked nodes. Fig. 5(b) describes the 10th eigenvector on graph. Here, the eigenvector is mostly smooth, it highlights another community, but it shows higher variability than the first eigenvector over linked nodes.

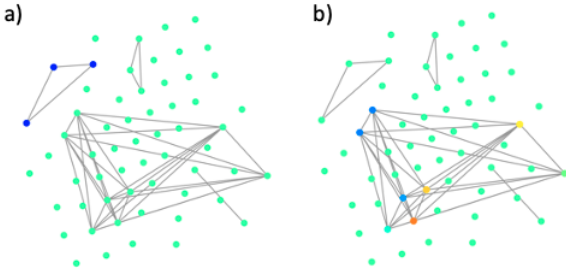


Fig. 5. Eigenvectors on graph. In panel a) there is the first eigenvector; in panel b) the 10th eigenvector

This model will be used to test the proposed Laplacian denoising algorithm. To this goal, we randomly produce 20 repetitions (or trials) for each statistical hypothesis, as i.i.d. realizations of our model with a fixed set of parameters. Different noise and polarization level will be considered.

B. Sub-space robustness on synthetic data

In this sub-section, we compare our laplacian-based filtering based on $\mathcal{U}_L \cup \mathcal{U}_H$, shortly denoted as $\mathcal{U}_{L \cup H}$, with \mathcal{U}_{ALL} , \mathcal{U}_L and \mathcal{U}_H . Specifically, we investigate the robustness of the different sub-spaces with simulated data.

With the final goal of measuring the sub-space robustness, for each sub-space we take into account two cases, namely the absence and the presence of perturbation, to which we refer to as the ground truth (GT) and the noisy cases, respectively. Each ground truth sub-space is compared to different noisy cases, corresponding to $\sigma_w = 0, 1.2$ for noise and $\sigma_b = 0, 2$ for polarization. To quantify sub-space robustness on synthetic data, we firstly measure the Frobenius subspace distance F [53] between the GT case and the noisy configurations, varying the perturbation levels⁸. Results of this analysis are in Fig.

⁸For each sub-space configuration, we compute the Frobenius distance F between its noisy and GT versions. See **Definition 2** in [53] for the mathematical formulation.

6. We plot F versus the trial for the subspace \mathcal{U}_H (red), the subspace \mathcal{U}_L (green), and the subspace $\mathcal{U}_{L \cup H}$ (blue) in several perturbation conditions. In Fig. 6(a) we have the zero perturbation case, in which, not surprisingly, $F = 0$ for every filter and every trial. With a gradual increase of perturbations (i.e. noise only case in Fig.6(b) or polarization only case in Fig.6(c)), the most favorable case is with \mathcal{U}_H for almost every trial. When perturbations dramatically increase Fig.6(d), performances decrease in particular for \mathcal{U}_H filter. In this figure, we do not report results for \mathcal{U}_{ALL} case because $F = 0$ for all the trials with all the eigenvectors.

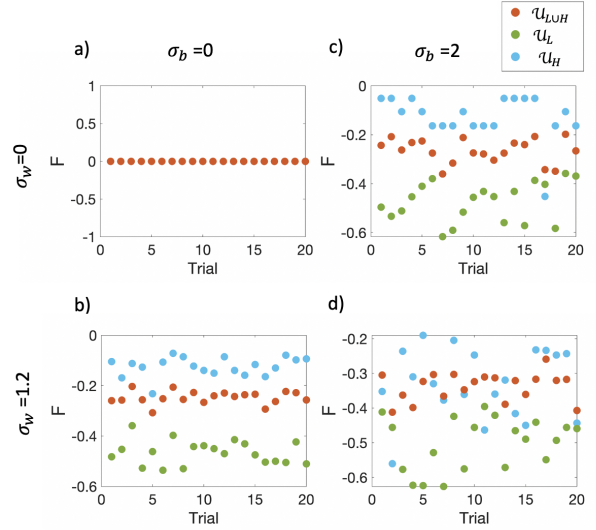


Fig. 6. Results of Frobenius distance on synthetic data. Several perturbation configuration are represented: in panel a) $\sigma_w=0$ and $\sigma_b=0$, in panel b) $\sigma_w=1.2$ and $\sigma_b=0$, in panel c) $\sigma_w=0$ and $\sigma_b=2$ and in panel d) $\sigma_w=1.2$ and $\sigma_b=2$. In the different colors (in the legend) we represent the different sub-spaces.

It is then clear that the eigenvectors in \mathcal{U}_H are significantly robust. This is not surprising, since in classical signal processing \mathcal{U}_H larger eigenvectors are used because of their advantages in signal-to-noise-ratio (SNR). Still, the subspace $\mathcal{U}_{L \cup H}$ maintains the robustness, while being relevant to describe the inherent topology of the graph. In the next results, we show that the Laplacian denoising leveraging the subspace $\mathcal{U}_{L \cup H}$ leads to better distinguishable connectivity states in absence and in presence of perturbation.

C. J-divergence computation on synthetic data

Finally, we test the ability to separate graph Laplacians under the hypotheses \mathcal{H}_1 and \mathcal{H}_0 . The J-divergence analysis presented in section V ends with a measure of the statistical distance J between two states. Here, the analysis of separability between two states is applied to graph laplacian in simulated scenario with the goal of comparing the discriminant ability of our denoising method with respect to the other sub-space configurations, i.e. \mathcal{U}_{ALL} , \mathcal{U}_L and \mathcal{U}_H .

Here the two conditions are the two hypothesis \mathcal{H}_1 and \mathcal{H}_0 and we perform simulations for several perturbation levels. In every case, we compute the total J as a measure of statistical distance between the two conditions and we also evaluate the marginal J_n as measure of the contribution of

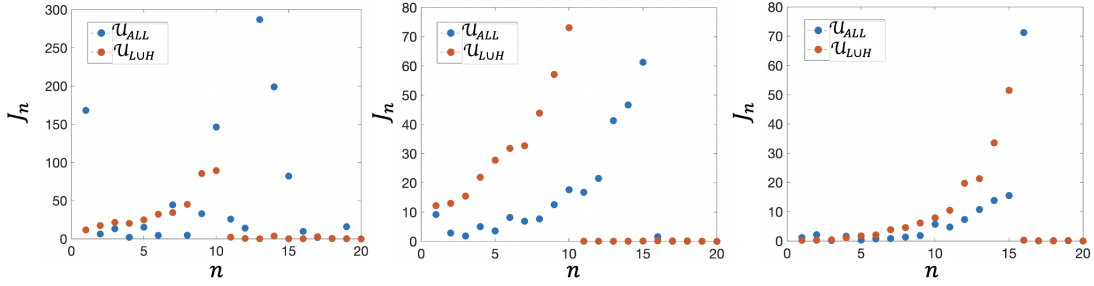


Fig. 7. Results of J-divergence analysis on synthetic data. Several perturbation configuration are represented: in panel a) $\sigma_w=0$ and $\sigma_b=0$, in panel b) $\sigma_w=1.2$ and $\sigma_b=0$, in panel c) $\sigma_w=1.2$ and $\sigma_b=2$. In the different colors (shown in the legend) we represent the different sub-spaces for the filtering.

each n -variable to the final separability. Table II collects J-divergence values for several perturbation sets and for different sub-spaces. Results show that in absence of perturbations the most favorable case is \mathcal{U}_{ALL} . This result is intuitive because in absence of perturbations there is no reason why reduced sub-spaces should better discriminate. Interestingly, increasing perturbations (i.e. noise and polarization), the most favorable case is \mathcal{U}_{LUH} , which gives the highest J , i.e. the best separability. It means that graph laplacian denoising through \mathcal{U}_{LUH} preserves the highest separability between the two hypothesis, even in presence of strong perturbation. Fig 7, shows J_n behavior as function of the first 20 variables only for \mathcal{U}_{ALL} and \mathcal{U}_{LUH} cases. These representations make clear the contribution of n variables to the final J . From Fig 7 we recognize that increasing perturbations, variables in \mathcal{U}_{LUH} generally give higher J_n contributions compared to \mathcal{U}_{ALL} .

	\mathcal{U}_H	\mathcal{U}_L	\mathcal{U}_{LUH}	\mathcal{U}_{ALL}
$\sigma_w=0$ $\sigma_b=0$	85.42	265.15	392.98	1073.93
$\sigma_w=1.2$ $\sigma_b=0$	223.58	294.25	329.19	264.38
$\sigma_w=1.2$ $\sigma_b=2$	188.10	106.77	313.53	150.60

TABLE II

J-DIVERGENCE VALUES ON SYNTHETIC DATA. WE REPORT IN BOLD CHARACTERS THE HIGHEST J-DIVERGENCE VALUE FOR EACH PERTURBATION CONFIGURATION.

Our results with synthetic data show that in presence of perturbations our laplacian filtering succeeds in distinguishing graphs under two conditions. This conclusion remains true if the system is perturbed by noise but also if there is an artefact of different nature, i.e. a common artifact that we indicated as polarization and which can represent real phenomena.

VII. REAL BCI MEASUREMENTS

In this section, we present experimental results of our graph laplacian filtering on real data, recorded during motor-imagery BCI experiments. In this case the \mathcal{H}_1 and \mathcal{H}_0 hypothesis directly correspond to the hypothesis that subject performs motor imagery (\mathcal{H}_1) or he/she is in resting state (\mathcal{H}_0).

A. Experimental Protocol and Preprocessing

The study was conducted on twenty healthy subjects (aged 27.60 ± 4.01 years, 8 women), all right-handed. All the subjects, which did not present any disorder, received financial compensation for their participation and they signed a written informed consent. The ethical committee CPP-IDF-VI of Paris approved the experimental protocol. During the BCI experiments, every subject was in front of a screen with a target. Subjects were instructed to perform right hand - motor imagery task when the target was up, while remaining at rest when the target was down [54]. A 74-channel system was used to record EEG data in a standard 10-10 configuration. The reference for EEG were set to mastoid signals; the ground electrode was located on the left scapula; and impedances were lower than 20 kOhms. Sampling frequency for EEG recordings was 1 kHz, and then downsampled to 250 Hz. For each subject, recorded sequences have been segmented to obtain N_T trials of motor imagery and N_T trials of resting state. The total length of each trial was 5s.

EEG data analysis was preceded by a pre-processing stage. Specifically, an Independent Component Analysis (ICA) was performed to eliminate artifacts, such as ocular and cardiac signals [55]; in particular, the Infomax Algorithm [56] was implemented with Fieldtrip toolbox [57].

B. J-divergence of brain connectivity states

Here, we perform the J-divergence analysis on real motor-imagery data. To this aim, we take EEG signals from one subject and N_T trials for \mathcal{H}_1 and N_T trials for \mathcal{H}_0 , with $N_T = 20$. We use spectral coherence to build the connectivity matrix, as in Eq. (19). Then, the estimated adjacency matrix $\hat{\mathbf{A}}$ is computed as in Eq. (20), and thanks to the Eq. (2), we can derive $\hat{\mathbf{L}}$. As for synthetic data, we compute the filtered graph laplacian $\tilde{\mathbf{L}}$ with the subset \mathcal{U}_{LUH} . We compare results on real data with \mathcal{U}_{ALL} , \mathcal{U}_H , \mathcal{U}_L . In each case, we compute J-divergence J as in Eq. (10) and the marginal contribution J_n associated to the n -th variable as in Eq.(15).

In Table III, we report J-divergence results for each sub-space configuration. Comparing all the methods, the highest J value relates to \mathcal{U}_{LUH} case. This result is very important because it means that the sub-space \mathcal{U}_{LUH} is suitable to separate real EEG data and it is useful to correctly detect the subject mental state.

	\mathcal{U}_H	\mathcal{U}_L	$\mathcal{U}_{L\cup H}$	\mathcal{U}_{ALL}
Real data	155.76	98.89	162.42	155.87

TABLE III
J-DIVERGENCE VALUES ON REAL DATA. WE REPORT IN BOLD
CHARACTERS THE HIGHEST J-DIVERGENCE VALUE.

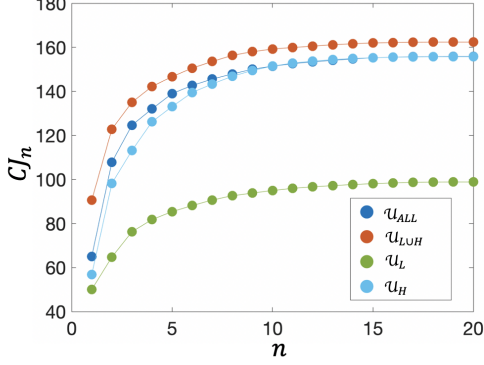


Fig. 8. Results of J-divergence analysis for real data. We report the CJ_n in Eq.(21) as function of the involved variables. In the different colors (shown in the legend) we represent the different sub-spaces used for filtering.

With the aim of understanding the contributions of different variables, we firstly compute the J_n marginal contributions to obtain a weight to each variable in the transformed domain. Then, we compute the cumulative sum of the first n variables. Once the J_n vector is sorted, we evaluate the cumulative J-divergence CJ_n to investigate the impact of the variables to the total J-divergence, as follows:

$$CJ_n = \sum_{k=1}^n J_k \quad (21)$$

Results in Fig. 8 show that the cumulative sum of the first 20 variables, is always higher for $\mathcal{U}_{L\cup H}$ than all the other sub-space configurations. In other words, if a given number of variables are retained, the overall achieved J-divergence is always larger using the proposed denoising algorithm. This confirms the contribution of the proposed Laplacian denoising to discrimination of the two mental states.

C. Scoring of Laplacian coefficients in β band

Given that the filtering with $\mathcal{U}_{L\cup H}$ enables a better discrimination between motor imagery and resting state, we now exploit the above introduced scoring procedure to determine, based on an information theoretic grounded criterion, which connectivity coefficients mostly contribute to separate the resting and motor imagery states.

To proceed, it is important to underline that the brain response to motor tasks in general is not uniform across the frequencies, but it is mostly evident in α (8-13 Hz) or β (14-29 Hz) band [58], depending on the subject.

As a proof of concept, we show results in β band, but in a training BCI scenario, the frequency band of interest can

be tuned according to the subject response. Here, we filter the connectivity matrix in the selected frequency band, as follows:

$$\hat{A}_{ij} = \sum_{\omega_k/T_s \in \beta} C_{ij}(\omega_k) \quad (22)$$

Thereby, having stated that the denoising based on $\mathcal{U}_{L\cup H}$ sub-space provides best results in separating Laplacians under \mathcal{H}_1 and \mathcal{H}_0 , we restrict the analysis to $\mathcal{U}_{L\cup H}$ and \mathcal{U}_{ALL} for score analysis in β band. We compute the scores as described in section V(A) and we report the score results in Fig. 9. In the first row, we collect the results referring to extra diagonal elements of \tilde{L} , i.e. links, and in the second row, we report results for diagonal elements, i.e. nodes weights. Besides, on the left and right columns we provide the results achieved without and with application of the proposed Laplacian denoising.

The first interesting observation is that, also when the analysis is restricted to the β band, application of the proposed denoising improves the J-divergence of the observed connectivity states (from 79.47 to 160.77).

As far as the score analysis is concerned, different remarks are in order.

In Fig. 9(c-d), the score associated to nodes weights (Laplacian diagonal elements) is represented. The score range in absence of denoising is smaller, i.e. maximum values are 0.4 and 0.7. Furthermore, the scoring obtained without denoising is larger on nodes located in frontal, temporal or parietal area, such as FP_Z and P_4 . After denoising, the scores are more pronounced on sensory-motor areas, and we are able to pinpoint some more relevant nodes, such as C_2 and FC_5 . Let us now analyze the score associated to links' weights (Laplacian extra-diagonal elements). In absence of denoising, recognizing contributions of different brain areas is difficult because all the link weights are generally low, i.e. between 0 and 0.035. Besides, we can observe that the 20 links with highest score do not involve sensory-motor nodes. On the contrary, when the denoised Laplacian is considered, link scores achieve higher values, i.e. 0.42, the strongest links are localized in sensory-motor areas, and links connecting contralateral motor areas, such as $CP_3 - C_3$ rank highest.

Thereby, the scoring based on the denoised Laplacian provides a mean for analysis and interpretation of the observed connectivity states.

D. Fast estimation of Laplacian coefficients in β band

BCIs aim to provide real time interaction between the subject and the interface [54], [59]; thereby, reducing the observation time T_{oss} for Laplacian estimation is beneficial for potential applicability to online motor-imagery BCI. With this application framework in mind, we test the Laplacian denoising when the observation time window length is reduced to $T_{oss} = 1s$. In the following we consider a moving window of length $T_{oss} = 1s$ and shift it by $m\Delta t, m = 0 \dots M - 1$, with $M = 9$ $\Delta t = 0.5s$, so as to analyze the total available recording length of 5s over nine 50% overlapping temporal intervals. [60].

For each of the 20 subjects of the experimental study, we compute the spectral coherence on the m -th temporal interval,

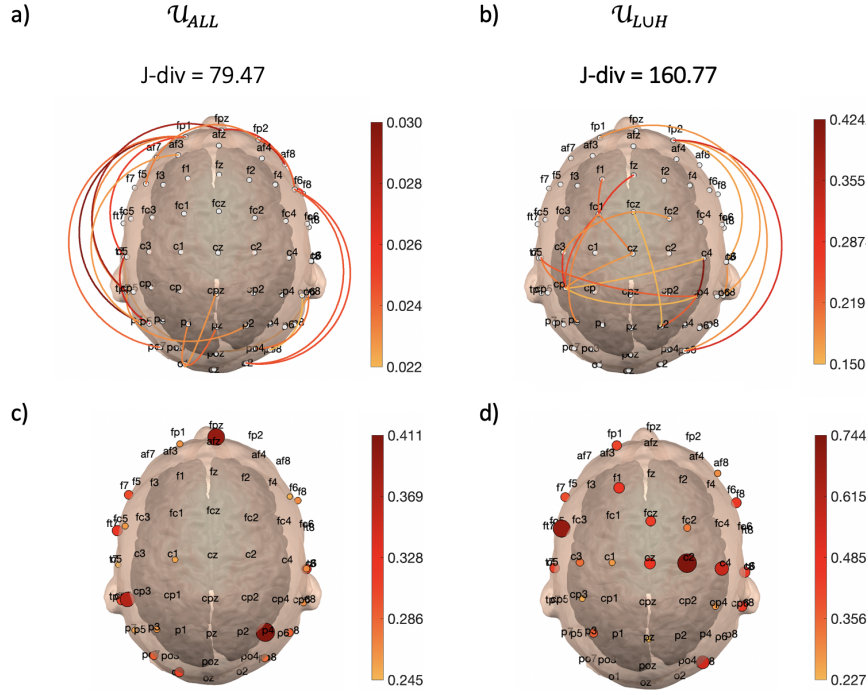


Fig. 9. Results of score computation for real data. We report results for \mathcal{U}_{ALL} subspace filtering in panels a,c); for \mathcal{U}_{LUH} subspace filtering in panels b,d). In the first line, score values refer to links (i.e. extra-diagonal elements) and in the second line, they relate to elements in the principal diagonal (i.e. nodes). For sake of clarity, in all the figures we plot the 20 nodes or links with highest score.

$m = 0 \dots M - 1$ as in Eq. (19) both for resting (\mathcal{H}_1) and motor imagery (\mathcal{H}_0) state. Then we derive the conditional ($\mathcal{H}_1, \mathcal{H}_0$) estimated adjacency matrix \hat{A} as in Eq.(22), the estimated graph laplacian \hat{L} as in Eq. (3), and its denoised version \tilde{L} as in Eq.(7). Then, we evaluate the J-divergence between the two hypotheses as in Eq. (15). Finally, we average the J obtained on the m -th window $m = 0 \dots M - 1$ in each time-interval across subjects. For comparison sake, we repeat the above computations in absence of denoising.

Fig. 10 reports results of this analysis by plotting the J-divergence, averaged across subject, as a function of the time window index m , $m = 0 \dots M - 1$. Our findings show that in the majority of the considered time intervals, i.e. on 7 intervals out of 9, the denoised Laplacian \tilde{L} , leveraging the \mathcal{U}_{LUH} subspace, leads to higher J-divergence than the estimated Laplacian \hat{L} (\mathcal{U}_{ALL} subspace). This result is really interesting because it shows that, even with short time-interval, our method succeeds in separating the two mental states. The above findings on real EEG data show that the proposed Laplacian denoising applies also on short time-windows and improves the potential to correctly detect motor imagery state. This paves the way to application of the proposed Laplacian denoising to real BCI applications.

VIII. CONCLUSION AND FURTHER WORK

This work has proposed a Laplacian denoising algorithm for the purpose of graph connectivity states detection. A novel formulation of the Jensen divergence has been derived. The J-divergence formulation is used to quantify the performance of the denoising algorithm, as well as to attribute a score

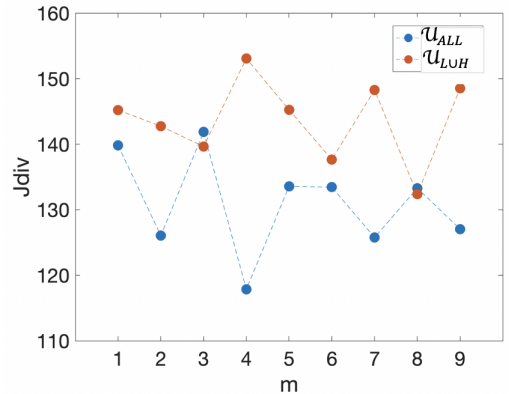


Fig. 10. Results of J-divergence analysis over a moving window on real data. We plot the J-divergence over $M = 9$, 1s long, time intervals with 50% overlapping, versus the time interval index. The J-divergence is computed in β band and averaged across subjects.

to the Laplacian coefficients in terms of their contribution to the connectivity states separability. The Laplacian denoising algorithm performances are assessed by numerical simulations on synthetic data. Furthermore, the Laplacian denoising algorithm has been applied to real EEG data acquired within motor imagery BCI experiments. The proposed Laplacian denoising improves the separation of the two mental states of motor imagery and resting state, even under restrained observation time intervals. Besides, the J-divergence based scoring sheds light over the contribution of different connectivity coefficients to motor imagery state detection. Thereby, the proposed ap-

proach is promising for the robust detection of connectivity states while being appealing for implementation in real-time BCI applications.

APPENDIX A THEOREM 1

Let us consider the problem of binary classification of Gaussian variables $\mathcal{H}_0 : \mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, $\mathcal{H}_1 : \mathbf{x} \sim \mathcal{N}(\boldsymbol{\eta}, \boldsymbol{\Sigma}^2)$, corresponding to the uncommon mean, uncommon covariance case, by means of the LLRT formulation in Eq.(11). By simple algebraic manipulation, we recognize that the test $\mathcal{R}(\mathbf{x}) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} t'$ corresponds to:

$$\mathcal{R}'(\mathbf{x}) = \underbrace{\sum_{n=1}^P \frac{\sigma_n^2 - 1}{\sigma_n^2} \left| x_n + \frac{\eta_n}{\sigma_n^2 - 1} \right|^2}_{P \text{ quadratic terms}} + \underbrace{\sum_{n=P+1}^N 2\eta_n \cdot x_n}_{N-P \text{ linear terms}} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} t'' \quad (23)$$

with $t'' = t' + \sum_{n=1}^P |\eta_n|^2 [\sigma_n^2 (\sigma_n^2 - 1)]^{-1}$.

Let us consider the linear-quadratic observation space Ξ of the N -dimensional random vector $\boldsymbol{\xi} \stackrel{\text{def}}{=} [\xi_1 \dots \xi_N]^T$ defined as (see Eq. (14))

$$\begin{aligned} \xi_n &= \left(x_n + \frac{\eta_n}{\sigma_n^2 - 1} \right)^2 \\ &= x_n^2 + 2 x_n \frac{\eta_n}{\sigma_n^2 - 1} + \left(\frac{\eta_n}{\sigma_n^2 - 1} \right)^2; n = 1, \dots, P \\ \xi_n &= x_n, n = P + 1, \dots, N \end{aligned} \quad (24)$$

In the space Ξ the LLRT $\mathcal{R}'(\mathbf{x}) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} t''$ rewrites as follows:

$$\sum_{n=0}^{N-1} a_{\text{LLR},n} \xi_n = \mathbf{a}_{\text{LLR}}^H \cdot \boldsymbol{\xi} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} t'' \quad (25)$$

where the elements of $\mathbf{a}_{\text{LLR}} \stackrel{\text{def}}{=} [a_{\text{LLR},1}, \dots, a_{\text{LLR},N}]^T$ are:

$$a_{\text{LLR},n} \stackrel{\text{def}}{=} \begin{cases} \frac{\sigma_n - \sigma_n^{-1}}{\sigma_n} & \text{for } n = 1, P \\ 2\eta_n & \text{for } n = P + 1, N \end{cases} \quad (26)$$

With these positions,

$$\begin{aligned} J &\stackrel{\text{def}}{=} \mathbb{E} \{ \mathcal{R}(\mathbf{x}) | \mathcal{H}_1 \} - \mathbb{E} \{ \mathcal{R}(\mathbf{x}) | \mathcal{H}_0 \} \\ &= \mathbb{E} \{ \mathcal{R}'(\mathbf{x}) | \mathcal{H}_1 \} - \mathbb{E} \{ \mathcal{R}'(\mathbf{x}) | \mathcal{H}_0 \} \\ &= \sum_{n=0}^{N-1} a_{\text{LLR},n} (\mathbb{E} \{ \xi_n | \mathcal{H}_1 \} - \mathbb{E} \{ \xi_n | \mathcal{H}_0 \}) \end{aligned} \quad (27)$$

By computing the above expectations it can be straightforwardly shown that the n -th term $a_{\text{LLR},n} (\mathbb{E} \{ \xi_n | \mathcal{H}_1 \} - \mathbb{E} \{ \xi_n | \mathcal{H}_0 \})$ of the above sum equals to

$$\begin{aligned} J_n(\sigma_n, \eta_n) &= \frac{\sigma_n - \sigma_n^{-1}}{\sigma_n} \left(\sigma_n^2 + \eta_n^2 + 2 \frac{\eta_n^2}{\sigma_n^2 - 1} - 1 \right) \\ &= (\sigma_n - \sigma_n^{-1}) \left[(\sigma_n - \sigma_n^{-1}) + \frac{\eta_n^2 \sigma_n^2 + 1}{\sigma_n \sigma_n^2 - 1} \right] \\ &= (\sigma_n - \sigma_n^{-1})^2 \left[1 + \frac{\eta_n^2}{\sigma_n (\sigma_n - \sigma_n^{-1}) (\sigma_n^2 - 1)} \right] \\ &= (\sigma_n - \sigma_n^{-1})^2 \left[1 + \frac{\eta_n^2}{\sigma_n (\sigma_n - \sigma_n^{-1})^2} \right], n = 1, \dots, P \\ J_n(\eta_n) &= 2\eta_n^2, n = P, \dots, N - 1. \end{aligned} \quad (28)$$

and

$$J_n(\eta_n) = 2\eta_n^2, n = P, \dots, N - 1. \quad (29)$$

QED.

ACKNOWLEDGEMENTS

FD acknowledges support from the Agence Nationale de la Recherche through contract number ANR15-NEUC-0006-02; and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 864729). TC acknowledges Juliana Gonzalez-Astudillo for useful discussions and suggestions.

REFERENCES

- [1] A. M. Bastos and J.-M. Schoffelen, "A tutorial review of functional connectivity analysis methods and their interpretational pitfalls," *Frontiers in systems neuroscience*, vol. 9, p. 175, 2016.
- [2] M. Newman, "Networks: an introduction: Oxford university press, inc," *Nielsen, M., Elmes, G. and Kipyatkov*, vol. 1999, 2010.
- [3] L. Torres, A. S. Blevins, D. S. Bassett, and T. Eliassi-Rad, "The why, how, and when of representations for complex systems," *arXiv preprint arXiv:2006.02870*, 2020.
- [4] J. Gonzalez-Astudillo, T. Cattai, G. Bassignana, M.-C. Corsi, and F. D. V. Fallani, "Network-based brain computer interfaces: principles and applications," *Journal of Neural Engineering*, 2020.
- [5] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control," *Clinical neurophysiology*, vol. 113, no. 6, pp. 767-791, 2002.
- [6] M. C. Thompson, "Critiquing the concept of bci illiteracy," *Science and Engineering Ethics*, vol. 25, no. 4, pp. 1217-1233, 2019.
- [7] A. Ortega, P. Frossard, J. Kovačević, J. M. Moura, and P. Vandergheynst, "Graph signal processing: Overview, challenges, and applications," *Proceedings of the IEEE*, vol. 106, no. 5, pp. 808-828, 2018.
- [8] A. Sandryhaila and J. M. Moura, "Discrete signal processing on graphs," *IEEE transactions on signal processing*, vol. 61, no. 7, pp. 1644-1656, 2013.
- [9] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE signal processing magazine*, vol. 30, no. 3, pp. 83-98, 2013.
- [10] W. Huang, T. A. Bolton, J. D. Medaglia, D. S. Bassett, A. Ribeiro, and D. Van De Ville, "A graph signal processing perspective on functional brain imaging," *Proceedings of the IEEE*, vol. 106, no. 5, pp. 868-885, 2018.
- [11] J. D. Medaglia, W. Huang, E. A. Karuza, A. Kelkar, S. L. Thompson-Schill, A. Ribeiro, and D. S. Bassett, "Functional alignment with anatomical networks is associated with cognitive flexibility," *Nature human behaviour*, vol. 2, no. 2, pp. 156-164, 2018.

- [12] L. Scharf and B. Van Veen, "Low rank detectors for gaussian random vectors," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 35, no. 11, pp. 1579–1582, 1987.
- [13] L. L. Scharf, *Statistical signal processing*, vol. 98. Addison-Wesley Reading, MA, 1991.
- [14] A. Pezeshki, L. L. Scharf, J. K. Thomas, and B. D. Van Veen, "Canonical coordinates are the right coordinates for low-rank gauss–gauss detection and estimation," *IEEE Transactions on Signal Processing*, vol. 54, no. 12, pp. 4817–4820, 2006.
- [15] V. Trees and L. Harry, *Detection, Estimation, and Modulation Theory-Part I-Detection, Estimation, and Linear Modulation Theory*. John Wiley & Sons New York, 2001.
- [16] H. V. Poor, *An introduction to signal detection and estimation*. Springer Science & Business Media, 2013.
- [17] S. Kullback, *Information theory and statistics*. Courier Corporation, 1997.
- [18] S. Theodoridis and K. Koutroumbas, "Pattern recognition, edition," 2009.
- [19] J. Schürmann, *Pattern classification: a unified view of statistical and neural approaches*. John Wiley & Sons, Inc., 1996.
- [20] B. Picinbono and P. Duvaut, "Detection and contrast," in *Stochastic processes in underwater acoustics*, pp. 181–203, Springer, 1986.
- [21] B. Picinbono and P. Devaut, "Optimal linear-quadratic systems for detection and estimation," *IEEE Transactions on Information Theory*, vol. 34, no. 2, pp. 304–311, 1988.
- [22] B. Picinbono, "On deflection as a performance criterion in detection," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 31, no. 3, pp. 1072–1081, 1995.
- [23] P. Chevalier and B. Picinbono, "Complex linear-quadratic systems for detection and array processing," *IEEE transactions on signal processing*, vol. 44, no. 10, pp. 2631–2634, 1996.
- [24] E. Bullmore and O. Sporns, "Complex brain networks: graph theoretical analysis of structural and functional systems," *Nature reviews neuroscience*, vol. 10, no. 3, pp. 186–198, 2009.
- [25] S. Segarra, S. P. Chepuri, A. G. Marques, and G. Leus, "Statistical graph signal processing: Stationarity and spectral estimation," in *Cooperative and Graph Signal Processing*, pp. 325–347, Elsevier, 2018.
- [26] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, "Complex networks: Structure and dynamics," *Physics reports*, vol. 424, no. 4-5, pp. 175–308, 2006.
- [27] N. Ghoroghchian, D. M. Groppe, R. Genov, T. A. Valiante, and S. C. Draper, "Node-centric graph learning from data for brain state identification," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 6, pp. 120–132, 2020.
- [28] G. Frusque, J. Jung, P. Borgnat, and P. Gonçalves, "Multiplex network inference with sparse tensor decomposition for functional connectivity," *IEEE transactions on Signal and Information Processing over Networks*, vol. 6, pp. 316–328, 2020.
- [29] W. Huang, L. Goldsberry, N. F. Wymbs, S. T. Grafton, D. S. Bassett, and A. Ribeiro, "Graph frequency analysis of brain signals," *IEEE journal of selected topics in signal processing*, vol. 10, no. 7, pp. 1189–1203, 2016.
- [30] M. Rabbat, M. Coates, and S. Blouin, "Graph laplacian distributed particle filtering," in *2016 24th European Signal Processing Conference (EUSIPCO)*, pp. 1493–1497, IEEE, 2016.
- [31] L. Rui, H. Nejati, and N.-M. Cheung, "Dimensionality reduction of brain imaging data using graph signal processing," in *2016 IEEE International Conference on Image Processing (ICIP)*, pp. 1329–1333, IEEE, 2016.
- [32] J. Wang, D. B. Aydogan, R. Varma, A. W. Toga, and Y. Shi, "Topographic regularity for tract filtering in brain connectivity," in *International Conference on Information Processing in Medical Imaging*, pp. 263–274, Springer, 2017.
- [33] P. C. Petrantakis and I. Kompatsiaris, "Single-trial nirs data classification for brain–computer interfaces using graph signal processing," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 9, pp. 1700–1709, 2018.
- [34] M. E. Spencer, R. M. Leahy, J. Mosher, and P. Lewis, "Adaptive filters for monitoring localized brain activity from surface potential time series," in *ASLOMAR CONFERENCE ON SIGNALS SYSTEMS AND COMPUTERS*, pp. 156–156, COMPUTER SOCIETY PRESS, 1992.
- [35] P. Strobach, K. Abraham-Fuchs, and W. Harer, "Event-synchronous cancellation of the heart interference in biomedical signals," *IEEE transactions on biomedical engineering*, vol. 41, no. 4, pp. 343–350, 1994.
- [36] M. Chen, R. T. Wakai, and B. Van Veen, "Eigenvector based spatial filtering of fetal biomagnetic signals," *Journal of perinatal medicine*, vol. 29, no. 6, pp. 486–496, 2001.
- [37] Y. Zhang, D. H. Brooks, M. A. Franceschini, and D. A. Boas, "Eigenvector-based spatial filtering for reduction of physiological interference in diffuse optical imaging," *Journal of biomedical optics*, vol. 10, no. 1, p. 011014, 2005.
- [38] J. Li, P. Shang, and X. Zhang, "Time series irreversibility analysis using jensen–shannon divergence calculated by permutation pattern," *Nonlinear Dynamics*, vol. 96, no. 4, pp. 2637–2652, 2019.
- [39] F. Nielsen, "On a generalization of the jensen–shannon divergence," *arXiv preprint arXiv:1912.00610*, 2019.
- [40] J. Giles, K. K. Ang, L. S. Mihaylova, and M. Arvaneh, "A subject-to-subject transfer learning framework based on jensen–shannon divergence for improving brain–computer interface," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3087–3091, IEEE, 2019.
- [41] K. J. Friston, "Functional and effective connectivity: a review," *Brain connectivity*, vol. 1, no. 1, pp. 13–36, 2011.
- [42] G. C. Carter, "Coherence and time delay estimation," *Proceedings of the IEEE*, vol. 75, no. 2, pp. 236–255, 1987.
- [43] G. Nolte, O. Bai, L. Wheaton, Z. Mari, S. Vorbach, and M. Hallett, "Identifying true brain interaction from eeg data using the imaginary part of coherency," *Clinical neurophysiology*, vol. 115, no. 10, pp. 2292–2307, 2004.
- [44] E. Ollila, D. E. Tyler, V. Koivunen, and H. V. Poor, "Complex elliptically symmetric distributions: Survey, new results and applications," *IEEE Transactions on signal processing*, vol. 60, no. 11, pp. 5597–5625, 2012.
- [45] E. Ceci and S. Barbarossa, "Graph signal processing in the presence of topology uncertainties," *IEEE Transactions on Signal Processing*, vol. 68, pp. 1558–1573, 2020.
- [46] U. Von Luxburg, "A tutorial on spectral clustering," *Statistics and computing*, vol. 17, no. 4, pp. 395–416, 2007.
- [47] M. Basseville, "Divergence measures for statistical data processing—an annotated bibliography," *Signal Processing*, vol. 93, no. 4, pp. 621–633, 2013.
- [48] Y.-M. Chang, N.-J. Hsu, and H.-C. Huang, "Semiparametric estimation and selection for nonstationary spatial covariance functions," *Journal of Computational and Graphical Statistics*, vol. 19, no. 1, pp. 117–139, 2010.
- [49] P. Guttorp and A. M. Schmidt, "Covariance structure of spatial and spatiotemporal processes," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 5, no. 4, pp. 279–287, 2013.
- [50] M. Greco, S. Fortunati, and F. Gini, "Maximum likelihood covariance matrix estimation for complex elliptically symmetric distributions under mismatched conditions," *Signal Processing*, vol. 104, pp. 381–386, 2014.
- [51] A. Ali, M. Testa, T. Bianchi, and E. Magli, "Biometricnet: deep unconstrained face verification through learning of metrics regularized onto gaussian distributions," *arXiv preprint arXiv:2008.06021*, 2020.
- [52] T. Cattai, S. Colonnese, M.-C. Corsi, D. S. Bassett, G. Scarano, and F. D. V. Fallani, "Characterization of mental states through node connectivity between brain signals," in *2018 26th European Signal Processing Conference (EUSIPCO)*, pp. 1377–1381, IEEE, 2018.
- [53] O. M. Baksalary and G. Trenkler, "On subspace distances determined by the frobenius norm," *Linear Algebra and its Applications*, vol. 448, pp. 245–263, 2014.
- [54] J. R. Wolpaw, D. J. McFarland, T. M. Vaughan, and G. Schalk, "The wadsworth center brain–computer interface (bci) research and development program," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 11, no. 2, pp. 1–4, 2003.
- [55] A. Delorme, T. Sejnowski, and S. Makeig, "Enhanced detection of artifacts in eeg data using higher-order statistics and independent component analysis," *Neuroimage*, vol. 34, no. 4, pp. 1443–1449, 2007.
- [56] A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural computation*, vol. 7, no. 6, pp. 1129–1159, 1995.
- [57] R. Oostenveld, P. Fries, E. Maris, and J.-M. Schoffelen, "Fieldtrip: open source software for advanced analysis of meg, eeg, and invasive electrophysiological data," *Computational intelligence and neuroscience*, vol. 2011, p. 1, 2011.
- [58] Y. Meirovitch, H. Harris, E. Dayan, A. Arieli, and T. Flash, "Alpha and beta band event-related desynchronization reflects kinematic regularities," *Journal of Neuroscience*, vol. 35, no. 4, pp. 1627–1637, 2015.
- [59] J. Wolpaw and E. W. Wolpaw, *Brain–computer interfaces: principles and practice*. OUP USA, 2012.
- [60] P. Shenoy, M. Krauledat, B. Blankertz, R. P. Rao, and K.-R. Müller, "Towards adaptive classification for bci," *Journal of neural engineering*, vol. 3, no. 1, p. R13, 2006.