Biomedical Signal Processing and Control 78 (2022) 103886

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Graph denoising of impulsive EEG signals and the effect of their graph representation☆

Anastasia Pentari [a,b,*], George Tzagkarakis [b], Kostas Marias [c,b], Panagiotis Tsakalides [a,b,1]

[a] Computer Science Department, University of Crete, Heraklion, Greece
[b] Institute of Computer Science, Foundation for Research and Technology-Hellas (FORTH-ICS), Greece
[c] Department of Electrical and Computer Engineering, Hellenic Mediterranean University, Heraklion, Greece

ARTICLE INFO

Keywords:
EEG denoising
Graph signal filtering
Fractional lower-order moments
Functional connectivity

ABSTRACT

As the field of brain monitoring is evolving rapidly, there is an increasing demand for innovative approaches to handle relevant signals. Recently, graph signal processing, which enables the treatment of signal ensembles, emerges as a powerful alternative to a per-signal analysis. This is especially the case for electroencephalogram (EEG) signals that naturally admit graph representations, with each electrode corresponding to one graph node. These signals are often corrupted by impulsive noise best characterized by heavy-tailed statistics, thus driving conventional denoising techniques to failure. To address this problem, we propose an efficient regularized graph filtering method based on fractional lower-order moments, which better adapt to heavy-tailed statistics. An experimental evaluation on real EEG measurements, including the publicly available P300 dataset and epilepsy signals, reveals a superior denoising performance of our method when compared against well-established EEG signal denoising methods.

1. Introduction

The acquisition and analysis of electroencephalography (EEG) signals is among the most prominent noninvasive electrophysiological monitoring methods for recording the brain's electrical activity. EEG signal recordings provide useful information about psychological dysfunctions and mental illnesses, even about the way our mind works [1]. However, EEGs are typically corrupted by noise, thus deteriorating subsequent interpretation of the acquired information. This is especially important when addressing distinct medical tasks, such as the automatic categorization (i.e., clustering or classification) of the examined subjects upon a predetermined set of diseases [2].

For the purposes of signal improvement, removing the corrupting noise from an EEG signal ensemble by designing appropriate filtering methods remains a critical task [2]. EEG signal denoising poses a challenge with respect to the selection of a filter that suppresses significantly the noise effects without sacrificing the inherent signal information content [1]. A denoised EEG signal enables a more accurate feature extraction, thus improving the performance of high-level tasks such as classification [3,4]. A major issue in the denoising process is the accurate identification and modelling of the noise component. This problem is especially challenging in the case of impulsive noise, which gives rise to gross samples and abrupt changes in the signal amplitude that may mask the information content of neighbouring samples.

Several methods have been developed for signal denoising tailored to impulsive noise. Most of the existing approaches are per-signal methods, based on the robust wavelet transform (WT) [5], the myriad filter [6], and the independent component analysis (ICA) [7]. Even though these methods are quite effective, they have some critical shortcomings. In particular, the WT aliases the signal structure [8], while the myriad filter's accuracy is affected by the sample size. Among the heavy-tailed noise models, alpha-stable distributions have played a prominent role in describing impulsive noise phenomena, hence they could be valuable tools for describing EEG signals' natural noise. The most prominent sources of noise in EEG come from environmental stimuli or head motions. Addressing the denoising problem of EEG signals with a more appropriate noise distribution, such as the alpha-stable, could improve the quality of the EEG measurements and subsequently help clinicians with their further analyses, such as automatic seizure detection.

☆ "This work is funded by the Hellenic Foundation for Research and Innovation (HFRI) and the General Secretariat for Research and Technology (GSRT) under grant agreement No. 1725 (V4-ICARUS)".

* Corresponding author at: Computer Science Department, University of Crete, Heraklion, Greece.
E-mail addresses: apentari@csd.uoc.gr (A. Pentari), gtzag@ics.forth.gr (G. Tzagkarakis), kmarias@ics.forth.gr (K. Marias), tsakalid@ics.forth.gr (P. Tsakalides).
[1] Member, IEEE.

https://doi.org/10.1016/j.bspc.2022.103886
Received 21 October 2021; Received in revised form 13 May 2022; Accepted 4 June 2022
Available online 16 June 2022

Moreover, per-signal approaches fail to fully exploit the EEG signal structure, such as potential intra- and inter-channel interdependencies. To address this limitation, graph-based methods have been gaining an increasing interest in the biomedical signal processing community. Specifically, graph-based approaches represent a signal ensemble by means of graphs and leverage the extensive analytical techniques developed in the emerging field of graph signal processing (GSP). EEG signals can be represented naturally as graphs with each electrode corresponding to a node of the graph [8]. Graphs are comprehensive data representations that are used for exploiting the geometric structure of data spaces. They are capable of defining the geometric structure of the recorded data by employing appropriate weights or simple binary edges, via the so-called adjacency matrices. In addition, graphs can handle signals which live on irregular grids. Through appropriate procedures, complex problems can be addressed such as denoising of graph signals or the study of signal diffusion [8].

When formulating a graph filter, previous approaches primarily rely on ℓ_2 optimization theory [9,10]. Such ℓ_2 norm based graph filters employ second-order statistics and are well tailored to Gaussian noise. On the other hand, alpha-stable distributions have algebraic tails and they lack finite second-order moments. To address this issue, we introduced in [8] an ℓ_p-regularized graph filter, which is tailored to impulsive noise modelled by means of alpha-stable distributions. The proposed filter showed an improved performance, mainly in highly-impulsive noise cases that deviate significantly from the Gaussian distribution. The solution of this optimization problem was based on an iterative approach through which we achieved to reduce large impulsive noise spikes. However, a main disadvantage of this approach is its difficulty in treating impulsive noise spikes with small amplitudes, a limitation due to the necessary number of iterations for convergence.

At the core of GSP is the design of an appropriate adjacency matrix, which encodes the data interconnectivity structure via weighted edges. In terms of EEG signal analysis, the most well-established computation of the adjacency matrix describes the statistical (i.e., the functional) brain information and depends on the cross-correlation approach. In terms of our analysis, as our EEG signals live in impulsive environments, we aim to approach the statistical brain information via the so-called *covariation matrix*, as a result of the *fractional lower-order moments*. Our main purpose is to evaluate if this matrix provides satisfying results, as an analogous quantity to the covariance matrix.

To this end, in this paper we propose an effective graph filtering method for the denoising of EEG signal ensembles, corrupted by impulsive noise, which is based on the solution of a modified $\ell_{p,\epsilon}$ minimization problem. Our $\ell_{p,\epsilon}$-regularized graph-based filter has the ability to be combined with a variety of adjacency matrices. Our graph-based filter is evaluated in various tasks and is tested for multiple brain structures. Our investigation takes into consideration a wide range of noise scenarios, including the filtering of raw noisy data as well as various cases of impulsive synthetic additive noise.

1.1. Main contributions

The major contribution of this paper is mainly based on the construction of a new graph filtering method as a solution of an $\ell_{p,\epsilon}$-regularized optimization problem, that better adapts to the statistics of the underlying heavy-tailed noise. Moreover, the effect of the EEG signals graph representation is evaluated, as a result of two main approaches, namely, the cross-correlation and the covariation. All our experimental evaluations pass through an automatic categorization task of EEG examinations, which is based on the classification of two important datasets, the P300 and recordings of epilepsy and non-epilepsy. Our classification experiments provide useful evaluation measures where each denoising procedure is tested in the presence of impulsive noise. Thus, even if we do not have a benchmark of the ideally clean signal ensembles, we can still conclude as to the most appropriate filter for the denoising of the EEG signals.

1.2. Paper organization

The rest of the paper is organized as follows: Section 2 includes significant information about the EEG signals. Section 3 introduces the main building blocks of our proposed graph filtering method, namely, the general GSP framework, along with the family of symmetric alpha-stable (SαS) distributions. Section 4 analyses our proposed graph filter, which is formalized as the solution of an approximated ℓ_p-regularized optimization problem. Section 5 describes the classification problem formulation, which is further used as a benchmark to evaluate the performance of our proposed graph filtering method for EEG signal ensembles, against alternative well-established denoising techniques. Section 6 evaluates the performance of our EEG signal denoising method and compares it against its counterparts, in terms of reconstruction quality and classification accuracy. The advantages of our proposed method are discussed in Section 7, whilst Section 8 summarizes the main outcomes and gives directions for further work.

2. EEG data description

This study aims to reduce the impulsive noise, which best characterizes the EEG signals natural noise. A main advantage of the EEG signals, denoted as $\mathbf{S} \in \mathbb{R}^{N\times K}$, is their ability to be represented by graphs. As mentioned before, each one of N electrodes denotes a graph's node, whilst each edge describes an interrelation between two signals of sample length equal to K. In this work, two public datasets are employed for the experimental evaluation, namely, the well-established P300 (ref. [11]) and another one that includes recordings of epileptic and non-epileptic subjects (available in. [12]).

Regarding the P300 dataset, it consists of normal and abnormal subjects. In the subsequent evaluation, 16 subjects were selected, 8 of which belong to "session 1" and the rest of them to "session 2". For each subject, 5 segments of length 256 were randomly selected from each of the 32 channels, with the time intervals being the same for all the channels of a subject. A usual phenomenon in the EEG signals analysis is the necessity of re-referencing. The idea behind the re-referencing is to express the voltage at the EEG scalp channels with respect to another, the reference. P300 is a dataset which requires such a manipulation. The available examinations consist of 34 channels, whose 32 are the electrodes of interest while the rest 2, i.e., the 33th and the 34th, are the mastoids electrodes (denoted as "MA1" and "MA2" in the dataset description). The mastoids channels help us to compute a value, the so-called averaged effect, which will contribute to the signals' re-referencing.

The first quantity we have to compute is the sum of all electrodes activity, ACT, as follows:

$$ACT = \frac{1}{2}\left(2\sum_{n=1}^{32}\sum_{k=1}^{K}\mathbf{S}(n,k) - 32\sum_{k=1}^{K}\mathbf{S}(33,k) - 32\sum_{k=1}^{K}\mathbf{S}(34,k)\right). \quad (1)$$

Then, the averaged effect, R, is given by

$$R = -\frac{ACT}{34}, \quad (2)$$

leading to a unique value. Adding this value to all the channels/electrodes signal amplitudes, the whole examination will be re-referenced. In the subsequent analysis, the original signals are considered to be the re-referenced ones. Notice that, the sampling rate of P300 recordings was 2048 Hz.

Regarding the epilepsy dataset, it consists of real EEG examinations of healthy subjects, as well as patients with epileptic or stroke events. The EEG recordings are acquired by 26 and 32 channels. As the performance of the examined filters is evaluated via a classification task, 20 subjects are selected for training, including 10 normal and 10 abnormal cases, and 10 subjects are selected (5 healthy control and 5 abnormal subjects).

3. GSP and symmetric alpha-stable models

This section presents the main theoretical background behind the family of symmetric alpha-stable distributions. First, let $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_N]^T \in \mathbb{R}^{N\times K}$ be the data matrix where $\mathbf{x}_i \in \mathbb{R}^K$ is the signal recorded by the ith electrode with N and K denoting the number of electrodes and samples per electrode, respectively. We adopt an additive observation noise model as follows,

$$\mathbf{X} = \mathbf{S} + \mathbf{W}, \tag{3}$$

where $\mathbf{S} \in \mathbb{R}^{N\times K}$ denotes the noiseless data matrix and $\mathbf{W} \in \mathbb{R}^{N\times K}$ is the noise term.

The main assumption in our study is that the additive noise matrix, $\mathbf{W}$, has independent and identically distributed (i.i.d.) entries following a SαS distribution. Specifically, SαS distributions are best described by their characteristic function defined by [13],

$$\phi(t) = \exp(i\delta t - \gamma^\alpha |t|^\alpha) \,, \tag{4}$$

where α is the *characteristic exponent*, which varies in $(0, 2]$ and controls the "thickness" of the tails of the density function. The smaller the characteristic exponent, the heavier the density function's tails. The *location parameter* $\delta \in \mathbb{R}$ defines the location of the density function. Without loss of generality, we assume that the noise follows a SαS distribution with $\delta = 0$. Finally, the *dispersion* $\gamma > 0$ determines the spread of the distribution around its location, and plays an analogous role to the variance for Gaussian distributions.

A key characteristic of SαS distributions is the lack of second-order moments. Instead, only moments of order $p < \alpha$ exist and they are called the *fractional lower order moments* (*FLOMs*). In particular, the FLOMs of a random variable following a SαS distribution, $X \sim f_\alpha(\gamma, \delta = 0)$, are given by [13]

$$\mathbb{E}\{|X|^p\} = \big(C(p,\alpha)\cdot\gamma\big)^p \,, \quad 0 < p < \alpha \,, \tag{5}$$

where $\big(C(p,\alpha)\big)^p = \frac{\Gamma\left(1-\frac{p}{\alpha}\right)}{\cos\left(\frac{\pi}{2}p\right)\Gamma(1-p)}$. The S$\alpha$S model parameters (α, γ) can be estimated using the consistent maximum likelihood (ML) method described by Nolan [14], which gives reliable estimates and provides the tightest possible confidence intervals.

Similarly to the second-order-moment theory, where covariance is a fundamental concept, an analogous quantity, the covariation plays an analogous role in the case of SαS distributions. Specifically, let X and Y be jointly SαS random variables (i.e., with the same α), with dispersions γ_X and γ_Y, respectively, and zero location parameters. Then, the covariation of X with Y is defined by

$$[X, Y]_\alpha = \frac{\mathbb{E}\{X Y^{\langle p-1\rangle}\}}{\mathbb{E}\{|Y|^p\}}\gamma_Y^\alpha \,, \tag{6}$$

where $z^{\langle a\rangle} = |z|^a \mathrm{sign}(z)$ for $z \in \mathbb{R}$ and $a \geq 0$, and $\mathbf{z}^{\langle a\rangle} = [|z_1|^a\mathrm{sign}(z_1), \ldots, |z_N|^a\mathrm{sign}(z_k)]$ for a vector $\mathbf{z} \in \mathbb{R}^K$. In the discrete case, let two vectors $\mathbf{x}_i, \mathbf{x}_j \in \mathbb{R}^K$ be two independent realizations of a SαS distribution. Their FLOM-based covariation estimator is defined by

$$c_{ij}^{\mathrm{FLOM}} = \frac{\sum_{k=1}^K x_{i,k}|x_{j,k}|^{p-1}\mathrm{sign}(x_{j,k})}{\sum_{k=1}^K |x_{j,k}|^p}\gamma_{\mathbf{x}_j}^\alpha \,, \tag{7}$$

where $x_{i,k} \in \mathbb{R}$ is the kth element of $\mathbf{x}_i$, whilst α and $\gamma_{\mathbf{x}_j}$ are estimated directly from $\mathbf{x}_j$ based on Nolan's ML estimator.

Notice that, in general, the covariation matrix $\mathbf{C}^{\mathrm{FLOM}}$ of the noisy data matrix $\mathbf{X}$ may have both positive and negative entries. In order to convert it into a valid weighted matrix to be incorporated in our graph filtering framework, we take the absolute values of its elements as follows,

$$\mathbf{F} = \frac{|\mathbf{C}_{\mathrm{FLOM}}| + |\mathbf{C}_{\mathrm{FLOM}}^T|}{2} \,. \tag{8}$$

All the above key expressions depend on the order of the FLOM, p. The selection of an appropriate p is a critical step towards better adapting to the degree of impulsiveness of the corrupting noise, as expressed by the corresponding characteristic exponent α. In our implementation, the optimal value of p is calculated as a function of α by minimizing the standard deviation of a FLOM-based covariation estimator, as described in [15], and specifically by linearly interpolating the entries of a lookup table (ref. Table I in [15]). Notice that regarding the optimal p value, we calculate the average α over all channels, and then we calculate the optimal p. Moreover, the covariation matrix does not necessarily be symmetrized. Specifically, a non-symmetrized covariation matrix was used for the analysis of P300 dataset, whilst a symmetrized one was employed for the epilepsy dataset.

As mentioned before, the EEG signal ensembles can be naturally represented in the form of a graph structure, $G = (V, E)$ with V and E denoting the set of nodes and edges, respectively. The interrelations among the nodes (i.e., the electrodes) are represented by an adjacency matrix $\mathbf{A} \in \mathbb{R}^{N\times N}$, which, in our analysis, is calculated from the noisy data $\mathbf{X}$.

One of the most important aspects in GSP theory is the computation of an appropriate adjacency matrix that best represents the interconnectivity relations in EEG signal ensembles. The well-established approach is to describe the interrelations between graph nodes via the *correlation matrix* [16], with correlation-based models being considered as appropriate descriptors of the brain's functional connectivity patterns. As a consequence, in our analysis, we evaluate the effect of the EEG signals representation via the FLOMs-based approach.

Analytically, in order to obtain the final adjacency matrix, $\mathbf{F}$ has to be normalized as follows,

$$\mathbf{A} = \frac{\mathbf{F}}{\max(|\lambda|)} \,, \tag{9}$$

a normalization widely used in GSP theory which exploits the eigendecomposition of the examined adjacency matrix and specifically its maximum absolute eigenvalue. Notice that, regarding the P300 dataset's examinations, we had to use the R value not only for re-referencing the signals, but also for normalizing the elements of the individual matrix $\mathbf{F}$. Thus, we had to divide, in an element-wise manner, its values with R, as this is computed via (2). Notice that correlation and related measures (including coherence and perhaps even covariation) are not without flaws when used to assess EEG functional connectivity due to volume conduction.

4. Graph filtering reformulated as an alternative ℓ_p-regularized optimization problem

In this section, we present the state-of-the-art graph-based denoising filters, followed by the detailed description of our proposed approach. The first method, introduced in [9], is based on a closed-form solution of a ℓ_2-norm-based minimization problem, which employs a total variation term. The second method is the graph-based filter we introduced in 2021, employing a ℓ_p-regularized formulation, which proved to be effective in reducing the impulsive noise of signal ensembles. Finally, we present our proposed $\ell_{p,\epsilon}$-regularized graph-based filter, whose iterative procedure depends on an update step, which employs the second derivative of our proposed objective function, in addition to the first.

4.1. ℓ_2 Graph polynomial filter

The graph filter described in [9,17,18] (a.k.a. graph polynomial filter), is obtained as the solution of the following optimization problem,

$$\hat{\mathbf{S}} = \underset{\mathbf{S}\in\mathbb{R}^{N\times K}}{\mathrm{argmin}} \left\{ \frac{1}{2}\|\mathbf{S} - \mathbf{X}\|_2^2 + \frac{1}{2}b\|\mathbf{S} - \mathbf{A}\mathbf{S}\|_2^2 \right\}, \tag{10}$$

where the first norm is a data fidelity term, while the second is a smoothness term controlling how much the graph signals vary with respect to the underlying graph. The solution of (10) is given in closed form,

$$\hat{\mathbf{S}} = (\mathbf{I} + b(\mathbf{I}-\mathbf{A})^H(\mathbf{I}-\mathbf{A}))^{-1}\mathbf{X} \tag{11}$$

where $\mathbf{I}$ is the identity matrix and H denotes the Hermitian transpose. Notice that this approach (hereafter denoted by "TV") is more appropriate for treating Gaussian noise.

4.2. Graph filtering of impulsive noise via ℓ_p-regularized optimization

Our graph-based filter, proposed in [8], addresses the problem of heavy-tailed impulsive noise and it suppresses the effects of the noise $\mathbf{W}$ by considering an appropriate statistical model based on alpha-stable distributions.

Motivated by [9], we employed the ℓ_p norm ($p < 2$) instead of the ℓ_2 norm, as FLOMs of $S\alpha S$ random variables do exist. Our objective function was the following

$$\hat{\mathbf{S}} = \underset{\mathbf{S}\in\mathbb{R}^{N\times K}}{\operatorname{argmin}} \left\{ \frac{1}{2}\|\mathbf{S}-\mathbf{X}\|_p^p + \frac{1}{2}b\|\mathbf{S}-\mathbf{A}\mathbf{S}\|_p^p \right\}, \tag{12}$$

where the first term formulates the denoising procedure while the second, i.e., the smoothness term, shows how much the graph signals vary with respect to the underlying graph. The regularization parameter b controls the smoothness of the denoised signals.

The problem in (12) is highly non-convex and its solution is obtained via an iterative approach comprising two stages, namely an iterative reweighted least squares (IRLS) step and a gradient descent-based update step, as shown in Algorithm 1, which describes the algorithm $\ell_{p,\epsilon}$. At the first stage, an approximate closed-form solution is obtained by applying the IRLS algorithm to the following modified Q_2 optimization problem:

$$Q_2(\mathbf{S}) = \underset{\mathbf{S}\in\mathbb{R}^{N\times K}}{\operatorname{argmin}} \left\{ \frac{1}{2}\|\mathbf{D}(\mathbf{S}-\mathbf{X})\|_2^2 + \frac{1}{2}b\|\mathbf{S}-\mathbf{A}\mathbf{S}\|_2^2 \right\} . \tag{13}$$

In essence, the IRLS algorithm provides an approximate solution of the ℓ_p-norm problem through the formulation of an objective function where the ℓ_p-norms are replaced by ℓ_2-norms.

In non-convex problems, proper initialization is a critical step. In this algorithm, we choose to initialize the denoised signal ensembles via the use of an additive ϵ quantity, in the order of 10^{-3}, to the input noisy signals $\mathbf{X}$. At each iteration of the algorithm, $k = 1, \ldots, K$, a residual vector $\mathbf{r}$ is calculated. The residual is defined as the difference between the current denoised signal kth column $\mathbf{S}(:,k)^{(t)}$ and the corresponding column of the recorded noisy signal $\mathbf{X}(:,k)$. Then, this residual vector $\mathbf{r}$, of length N, is used to create the diagonal matrix $\mathbf{D} \in \mathbb{R}^{\mathbb{N}\times\mathbb{N}}$, which contains the weight values $\mathbf{w}$ in its main diagonal. Notice that, this diagonal matrix is of size $N \times N$, as it must be consistent to the size of the adjacency matrix. To handle divisions by zero, we add a small constant ε, in the order of 10^{-3}, to the residual vector $\mathbf{r}$. Finally, note that the $(p-2)/2$ exponent is applied on each element of vector $\mathbf{w}$ (cf. Algorithm 1).

In order to solve the optimization problem presented in Q_2, we set the derivative of this objective function, with respect to $\mathbf{S}$, equal to zero. The result is a closed-form estimation of the denoised signals, given by:

$$\begin{aligned} \hat{\mathbf{S}} &= \left(\mathbf{D}^T\mathbf{D} + b(\mathbf{I}-\mathbf{A})^H(\mathbf{I}-\mathbf{A})\right)^{-1}(\mathbf{D}^T\mathbf{D})\mathbf{X} \\ \overset{\mathbf{D^T=D}}{\Longrightarrow}\ \hat{\mathbf{S}} &= \left(\mathbf{D}^2 + b(\mathbf{I}-\mathbf{A})^H(\mathbf{I}-\mathbf{A})\right)^{-1}\mathbf{D}^2\mathbf{X} . \end{aligned} \tag{14}$$

Having obtained a first estimate of the denoised signals, $\hat{\mathbf{S}}$, the second step involves the use of a gradient descent scheme applied to the initial objective function Q_1. Specifically, line 11 of Algorithm 1 gives the update signals solution step, a function of the Q_1's gradient, defined by the following equations:

$$\nabla Q_1(\mathbf{S}^{(t)}) = p\mathbf{W}_1^{(t)}(\mathbf{S}^{(t)}-\mathbf{X}) + p\mathbf{W}_2^{(t)}(\mathbf{S}^{(t)}-\mathbf{A}\mathbf{S}^{(t)}) , \tag{15}$$

where $\mathbf{W}_1^{(t)}$ and $\mathbf{W}_2^{(t)}$ are diagonal matrices with elements

$$\mathbf{W}_1^{(t)}(n,n) = \sum \left(|\mathbf{I}(n,:)\mathbf{S}^{(t)} - \mathbf{X}(n,:)|^{(p-2)} \right) , \tag{16}$$

$$\mathbf{W}_2^{(t)}(n,n) = \sum \left(|(\mathbf{I}(n,:) - \mathbf{A}(n,:))\mathbf{S}^{(t)}|^{(p-2)} \right) . \tag{17}$$

The sums correspond to the summation of the K elements of each extracted vector. Moreover, the elements of the diagonal matrices $\mathbf{W}_1^{(t)}$ and $\mathbf{W}_2^{(t)}$ correspond to each one out of the N electrodes (channels), while the exponent in their computation refers to an element-wise manipulation. Notice that, the q value in line 11 of Algorithm 1 corresponds to the learning rate. Finally, the ℓ_p-based algorithm terminates based on a given number of iterations I_{max} that depends on the noise impulsiveness. The smaller the characteristic exponent α, the greater the number of required iterations. It is worth mentioning that, the update step of the ℓ_p-regularized graph filter algorithm (i.e., in line 11) is simply described as: $\mathbf{S}^{(t+1)} = \mathbf{S}^{(t)} + q\nabla Q_1(\mathbf{S}^{(t)})$. Hereafter, we will refer to this ℓ_p-regularized graph-based filter as "*lp*".

4.3. $\ell_{p,\epsilon}$-Regularized graph filter

Although the ℓ_p-regularized graph-based filter, described in the previous section, efficiently reduces impulsive noisy spikes, it cannot handle low-amplitude noise as effectively because the filter converges too fast. The problem lies in the singularity of the gradient of the ℓ_p^p objective function. Hence, to make convergence slower, we modify the cost function in the optimization problem by using the following regularized version of the ℓ_p^p norm,

$$\|\mathbf{x}\|_{p,\epsilon}^p = \sum_{j=1}^{J} \left(|x_j|^2 + \epsilon\right)^{p/2} , \tag{18}$$

where ϵ is a positive parameter that allows the gradient of the cost function, presented in (19), to be different than zero, for any $p \in (0,1)$.

Based on the above regularized version of the ℓ_p^p-norm, our proposed modified objective function takes the following form,

$$\hat{\mathbf{S}} = \underset{\mathbf{S}\in\mathbb{R}^{N\times K}}{\operatorname{argmin}} \left\{ \frac{1}{2}\|\mathbf{S}-\mathbf{X}\|_{p,\epsilon}^p + \frac{1}{2}b\|\mathbf{S}-\mathbf{A}\mathbf{S}\|_{p,\epsilon}^p \right\} . \tag{19}$$

The above minimization problem has a solution for any $p \in (0,1)$ and $\epsilon > 0$. Again, the first term is necessary for the denoising of the signals, whilst the second, the smoothness term, retains the signals' structure to the underlying graph. Hereafter, the objective function of (19) is denoted by Q_3.

As in the ℓ_p^p case, the solution of (19) also relies on an iterative process. Specifically, the solution of the $\ell_{p,\epsilon}^p$ minimization problem is based on the IRLS algorithm, which approximates $\ell_{p,\epsilon}^p$ norms with the convex ℓ_2^2 norm. The steps of IRLS are described in lines $1-10$ of Algorithm 1.

Modification 1: A first modification of the previous, lp, algorithm is required in order to adapt to the objective function Q_3. More specifically, the update step is carried out based on Newton's gradient descent method, using both the first and second derivatives of Q_3, as follows,

$$\mathbf{S}^{(t+1)} = \mathbf{S}^{(t)} + \left(\nabla^2 Q_3(\mathbf{S}^{(t)})\right)^{-1}\nabla Q_3(\mathbf{S}^{(t)}) . \tag{20}$$

The first derivative of Q_3 is given by

$$\nabla Q_3(\mathbf{S}^{(t)}) = p\mathbf{W}_{1,\epsilon}^{(t)}(\mathbf{S}^{(t)}-\mathbf{X}) + p\mathbf{W}_{2,\epsilon}^{(t)}(\mathbf{S}^{(t)}-\mathbf{A}\mathbf{S}^{(t)}) , \tag{21}$$

where $\mathbf{W}_{1,\epsilon}^{(t)}$ and $\mathbf{W}_{2,\epsilon}^{(t)}$ are diagonal matrices given by

$$\mathbf{W}_{1,\epsilon}^{(t)}(n,n) = \sum \left(\left(\mathbf{I}(n,:)\mathbf{S}^{(t)} - \mathbf{X}(n,:)\right)^2 + \epsilon \right)^{(p-2)/2} , \tag{22}$$

$$\mathbf{W}_{2,\epsilon}^{(t)}(n,n) = \sum \left(\left((\mathbf{I}(n,:) - \mathbf{A}(n,:))\mathbf{S}^{(t)}\right)^2 + \epsilon \right)^{(p-2)/2} , \tag{23}$$

where the summations and exponentiations are performed in an element-wise fashion over the elements of the involved vectors of length K.

Regarding the second derivative of Q_3, after some algebraic manipulation, it is given by

$$\nabla^2 Q_3(\mathbf{S}^{(t)}) = p\left(diag\left(\frac{\partial \mathbf{W}_{1,\epsilon}^{(t)}}{\partial \mathbf{S}^{(t)}}\right)\right)(\mathbf{S}^{(t)} - \mathbf{X}) + p\mathbf{W}_{1,\epsilon}^{(t)}\mathbf{U} + p\left(diag\left(\frac{\partial \mathbf{W}_{2,\epsilon}^{(t)}}{\partial \mathbf{S}^{(t)}}\right)\right)(\mathbf{S}^{(t)} - \mathbf{A}\mathbf{S}^{(t)}) + p\mathbf{W}_{2,\epsilon}^{(t)}(\mathbf{I}-\mathbf{A})^T(\mathbf{I}-\mathbf{A})\mathbf{U}\ , \tag{24}$$

where $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_N]^T$ denotes the all-ones matrix of size $N \times K$. The two matrix derivatives in (24) are given by

$$\frac{\partial \mathbf{W}_{1,\epsilon}^{(t)}}{\partial \mathbf{S}^{(t)}} = (p-2)\sum_{n=1}^{N}\sum_{k=1}^{K}\left((\mathbf{S}^{(t)}(n,k) - \mathbf{X}(n,k))^2 + \epsilon\right)^{(p-4)/2} \cdot |\mathbf{I}(n,:)\mathbf{S}^{(t)} - \mathbf{X}(n,:)|\mathbf{u}_n^T\ , \tag{25}$$

$$\frac{\partial \mathbf{W}_{2,\epsilon}^{(t)}}{\partial \mathbf{S}^{(t)}} = (p-2)\sum_{n=1}^{N}\sum_{k=1}^{K}\left((\mathbf{I}(n,n) - \mathbf{A}(n,n))\mathbf{S}^{(t)}(n,k)^2 + \epsilon\right)^{(p-4)/2} \cdot \left(|\mathbf{I}(n,:) - \mathbf{A}(n,:)|\mathbf{S}^{(t)}\right)(|\mathbf{I}(n,:) - \mathbf{A}(n,:)|)\mathbf{u}_n^T\ . \tag{26}$$

As before, the exponentiations in the above expressions are performed in an element-wise fashion. Notice that, the division of the first to the second derivative of Q_3 in line 11 of Algorithm 1 is in an element-wise fashion.

Modification 2: A main limitation of both graph-based approaches is that they rely on a predefined learning rate q, which may result in algorithm's divergence if not properly chosen. To alleviate this issue, a second modification in line 11 of Algorithm 1 concerns the employment of an adaptive learning rate computed as follows: Let $\mathbf{G}^{(t)}$ denotes the division of the gradient of the first derivative of Q_3 at tth iteration to the gradient of its second derivative. Then, the adaptive learning rate, lr, is given by

$$lr = \frac{1}{\sqrt{\sum_{n=1}^{N} diag\left(\mathbf{G}(n,:)\mathbf{G}(n,:)^T\right) + \varepsilon}} \tag{27}$$

To this end, multiplying the previous learning rate q with the adaptive learning rate lr, we significantly reduce the sensitivity of Algorithm 1 on the manually set rate q improving its convergence. Notice that the ϵ parameter is used to avoid division with zero, and we set its value at the order of 10^{-2}.

Modification 3: Finally, an also important optimization to our approach was to set an appropriate stopping criterion to the algorithm. Specifically, taking the *root mean squared error (RMSE)* between the current denoised signals and the noisy ones, we computed the relative error as a fraction of the $error = RMSE^{(t-1)}/RMSE^{(t)}$. If this ratio is less than a threshold at the order of 10^{-3} then the iterative procedure stops. Hereafter, the graph filter derived based on the $\ell_{p,\epsilon}^p$ formulation will be denoted by "lpe".

Overall, the main difference between the lp and lpe filters lies in the update step in line 11 of Algorithm 1. Specifically, in the case of lp method, this step relies on a gradient descent scheme adapted to the ℓ_p^p formulation in (12). On the other hand, regarding the lpe method, the graph filter utilizes Newton's gradient descent on the $\ell_{p,\epsilon}^p$ approximation of (19). Finally, we employed an adaptive learning rate in our implementation, so as to improve its convergence.

Algorithm 1 $\ell_{p,\epsilon}$-regularized graph filter

1: **Inputs**: $\mathbf{X}$, $\mathbf{A}$, I_{max}, b, ϵ
2: **Outputs**: $\hat{\mathbf{S}}$
3: **Initialization**: $\mathbf{S}^{(0)} = \mathbf{X} + \epsilon$
4: **for** t = 1:I_{max} **do**
5: **for** k = 1:K **do**
6: $\mathbf{r} = \mathbf{S}(:,k)^{(t)} - \mathbf{X}(:,k)$
7: $\mathbf{w} = |\mathbf{r} + \varepsilon|^{(p-2)/2}$
8: $\mathbf{D} = \text{diag}(\mathbf{w})$
9: $\mathbf{S}(:,k)^{(t)} = \left((\mathbf{D}^2 + b(\mathbf{I}-\mathbf{A})^H(\mathbf{I}-\mathbf{A}))^{-1}\mathbf{D}^2\right)\mathbf{X}(:,k)$
10: **end for**
11: $\mathbf{S}^{(t+1)} = \mathbf{S}^{(t)} + qlr\frac{\nabla Q_3(\mathbf{S}^{(t)})}{\nabla^2 Q_3(\mathbf{S}^{(t)})}$
12: **if** $error < 10^{-3}$ **then**
13: $\hat{\mathbf{S}} = \mathbf{S}^{(t)}$
14: break
15: **end if**
16: **end for**
17: $\hat{\mathbf{S}} = \mathbf{S}^{(Imax)}$

5. Medical diagnosis tasks

The performance of our proposed graph filtering framework, and its comparison against the state of the art, is evaluated indirectly in a classification scenario as the targeted medical diagnosis application. Classification consists of a feature extraction step, which is responsible for the extraction of informative signal characteristics, followed by the assignment of the examined signals into predefined classes through a learning process. The reason we rely on classification as an indirect means of performance evaluation is due to our lack of knowledge about the original, noise-free signals, since the recorded raw measurements are typically corrupted by distinct noise sources (e.g. electronic, head motion, etc.). In the following, we provide a brief description of the extracted features, along with the classifier we employ in the subsequent experiments.

5.1. Feature extraction

In our implementation, the feature extraction step depends on the calculation of five variables per EEG channel, namely, the mean and standard deviation of each channel, the mean value of the samples which belong in the position range $[i-10 : i+10]$ of the position i where the maximum amplitude of the signal is located (if they exist), and the maximum and minimum channel values. The final feature vector is obtained as the average of the feature vectors of the N channels.

5.2. Classification process

In our implementation, the k-nearest neighbours (k-NN) algorithm is employed to perform the classification task. We emphasize that the overall performance of a medical diagnosis system depends on the proper selection of the classifier. Nevertheless, classification is used as an application scenario in this work, while finding the optimal classifier for the datasets considered herein is beyond the scope of the paper.

The k-NN algorithm classifies unlabelled observations by assigning each observation to the nearest class according to the utilized (labelled) training samples (i.e., features in our case). Regarding the P300 dataset, the corresponding features are assigned the labels {"session 1", "session 2"} depending on the session each window segment belongs to. As for the epilepsy data, the features corresponding to a given window are assigned the labels {"normal", "abnormal"} depending on whether a window corresponds to a normal or epileptic seizure, respectively. Then, the "unknown" class of a given test sample is estimated by

calculating the Euclidean distance between its feature vector and the feature vectors of the training samples, and then selecting the class where the k nearest neighbours belong to.

6. Experimental evaluation

6.1. Performance evaluation on P300 dataset

In this section, the effectiveness of the lpe graph-based filter is evaluated and compared with its competitors, i.e., the lp, TV and a per-signal denoising method, the wavelet transform ("WT"), in terms of statistical and classification metrics.

Towards this direction, the first evaluation was done on the public EEG dataset, namely, the P300 dataset, described in Section 2. The experimental assessment of P300 was based on the corruption of the raw data with synthetic impulsive additive noise. Specifically, symmetric alpha-stable noise is generated by varying the degree of impulsiveness $\alpha \in \{1.1, 1.3, 1.7, 2\}$, and the dispersion $\gamma \in \{1.5, 4.5\}$. The smaller the characteristic exponent and the larger the dispersion, the higher the probability of occurrence of high-amplitude spikes.

The main purpose of the evaluation on P300 is to verify the effect of the distinct denoising methods presented in this work, under a controlled degree of noise impulsiveness, and how the noise impulsiveness affects the classification accuracy of the examined subjects. Four denoising methods are compared below, namely, WT, TV (based on ℓ_2 norm), lp and lpe. Furthermore, in order to investigate the effect of the adjacency matrix in representing different brain anatomies and functionalities, two distinct adjacency matrices are compared, namely, the cross-correlation and the covariation-based one. Before the analysis of P300 dataset, some preprocessing is required, as described in the next paragraph.

Preprocessing of P300. The first step is the re-referencing step, which was described in Section 2 and ensures that all the EEG signals are transferred to the same amplitude scale with a common starting point. Since the graph filters handle the whole EEG signal ensemble simultaneously, this re-referencing is critical so as to eliminate the effects of different scales among the channels. The second step concerns the examination of the impulsive noise effects on the raw EEG recordings. To this end, a histogram of the estimated characteristic exponent values α of all EEG channels is constructed (ref. Fig. 1). Clearly, we observe that the EEG recordings present a broad range of impulsive behaviours, from light-tailed ($\alpha = 2$) up to extremely heavy-tailed ($\alpha = 1.2$).

Adjacency matrices normalization. Regarding the two examined adjacency matrices, some additional normalization is required. Specifically, the cross-correlation matrix (hereafter denoted as "cross-corr") is normalized by its maximum absolute eigenvalue, similarly to (9). Regarding the covariation matrix, the following normalization steps are applied: (i) the absolute value of the estimated covariation matrix is taken; (ii) a normalization is performed by dividing with the average effect R, so as to rescale its entries in the range $[0, 1]$. Finally, (iii) a further normalization is required by dividing its values with the maximum absolute eigenvalue.

In the subsequent experimental evaluation, the performance of each method and adjacency matrix is averaged over 30 Monte Carlo runs, 100 classification different partitions, all channels, extracted segments and over all subjects examined. The performance metric employed is the signal-to-error ratio (SER) defined by $\mathrm{SER}(\mathbf{s}, \hat{\mathbf{s}}) = 10\log_{10}\left(\frac{\|\mathbf{s}\|_2^2}{\|\mathbf{s}-\hat{\mathbf{s}}\|_2^2}\right)$ (in dB), where $\mathbf{s}$ and $\hat{\mathbf{s}}$ denote the original and denoised signals, respectively.

Table 1 summarizes the parameters setting for each one of the compared filters, selected upon a cross-validation process. Specifically, we let the regularization parameter vary in $b = 0.01, 0.1, 1$ and the learning rate vary in $q = 0.1, 0.01, 0.001$. Then, the best classification scores were calculated and the corresponding pair of (b, q) parameters was chosen. Among all the evaluations, we observed that for $b \geq 0.1$ and $q \leq 0.01$, the classification results of the selected data were not affected significantly. Regarding the ϵ parameter, since it is used to avoid divisions by zero, we empirically set its value at the order of 10^{-2}.

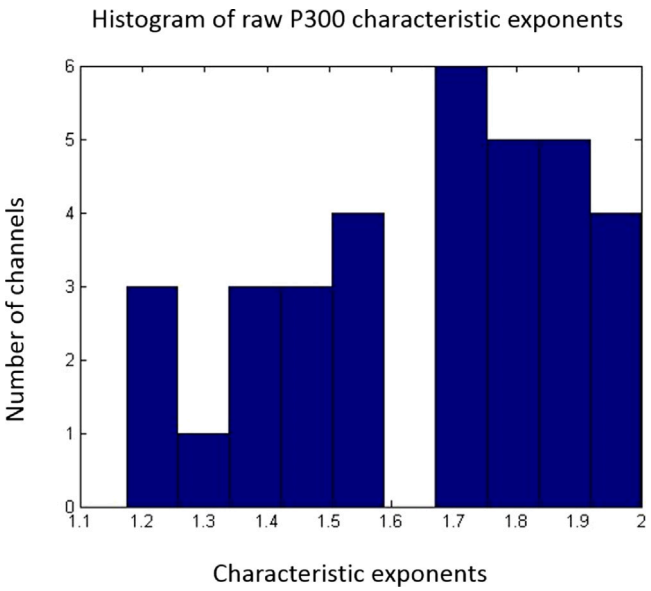

Fig. 1. Histogram of characteristic exponents estimated from the raw P300 data.

Table 1
Denoising methods specifications.

WT	TV	lp	lpe
db8	$b = 0.01$	$b = 0.1$	$b = 0.1$
2 levels	–	$q = 0.001$	$q = 0.001$
–	–	$\epsilon = 0.01$	$\epsilon = 0.01$
–	–	$I_{max} = 20$	$I_{max} = 20$

Finally, the selected number of total iterations I_{max} was chosen to be equal to 20.

Denoising Performance. Figs. 2 and 3 illustrate the averaged SERs, as well as the standard deviations, for the compared adjacency matrices and for all the denoising methods, corresponding to $\gamma = 1.5$ and $\gamma = 4.5$, respectively. First, we observe that as the impulsiveness increases (i.e., small α values), and for the moderate dispersion $\gamma = 1.5$, our proposed lpe filter outperforms significantly its counterparts, demonstrating a better adaptation to the underlying heavy-tailed statistics of the noise part. Regarding the adjacency matrices, cross-correlation is shown to be slightly better against the other. This can be attributed to the fact that accurate estimation of the covariation adjacency matrix requires more samples than the cross-correlation matrix. Nevertheless, for larger dispersion ($\gamma = 4.5$) the difference in performance between the cross-correlation and the covariation matrices decreases. Finally, as the noise statistics are close to a Gaussian ($\alpha \approx 2$), the WT achieves the best performance among all the denoising methods.

It is important to note that, as we do not have knowledge of the true (noiseless) signals, comparison of SER values does not fully guarantee the effectiveness of a denoising method. Furthermore, the WT method is independent of the choice of an adjacency matrix, that is why the SER values are equal for the two distinct matrices. Motivated by this result, we investigate the performance of each denoising method in an indirect way. Specifically, a classification task is carried out upon filtering the synthetically corrupted data using the four denoising methods.

Classification of P300. Regarding the classification task, the dataset includes 40 windows from "session 1" and 40 windows from "session 2". The 80% of them is kept for training, and the remaining 20% for testing. Fig. 4 shows the classification rates, and the standard

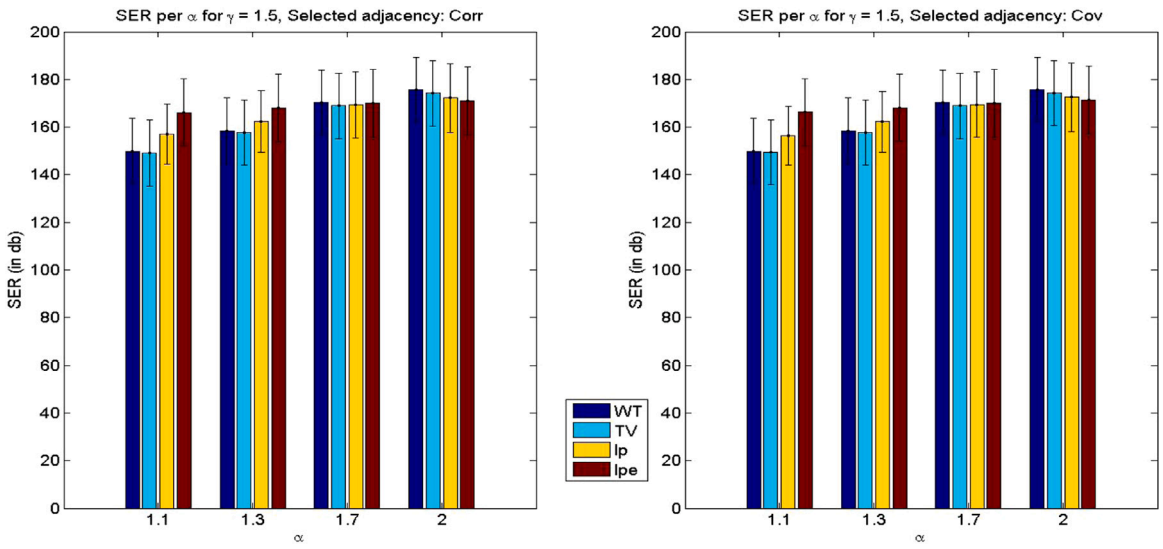

Fig. 2. Average SER versus α for the four denoising methods and the two adjacency matrices ($\gamma = 1.5$).

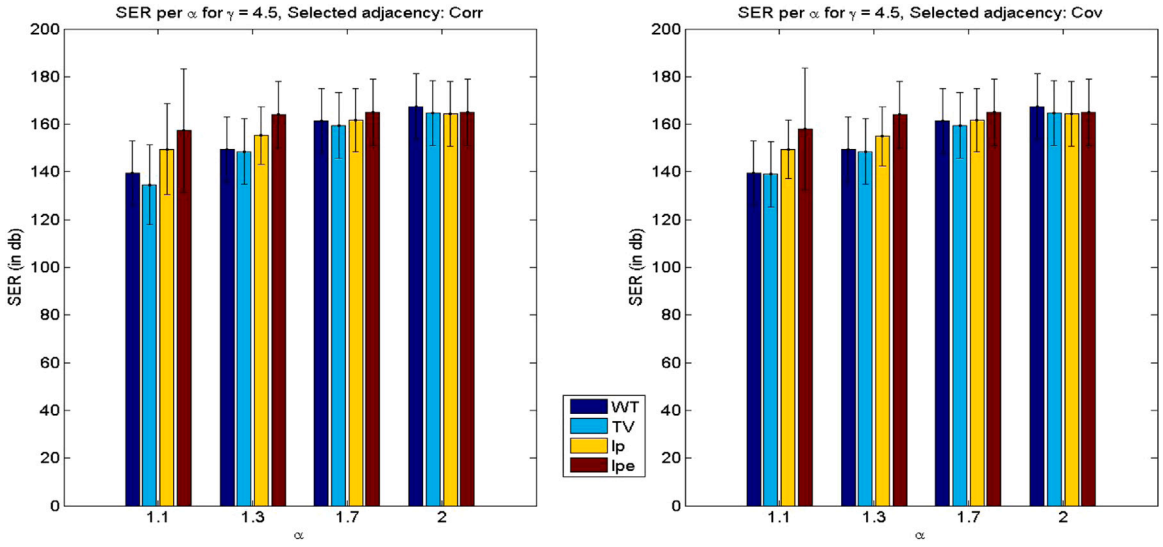

Fig. 3. Average SER versus α for the four denoising methods and the two adjacency matrices ($\gamma = 4.5$).

deviations, for each denoising method and adjacency matrix, by varying the degree of noise impulsiveness. First, we observe that for both dispersion cases, the lpe filter outperforms its counterparts. Regarding the two adjacency matrices, we do not observe significant differences in the performance. Notice that, as the value of α increases (i.e., the noise distribution tends to a Gaussian) all the denoising methods achieved a similar classification accuracy about 0.92, as in the original classification case.

Finally, in Fig. 5, we present an experimental result of the P300 dataset. Specifically, this concerns selected merged segments of one subject out of the 16 selected, and one electrode of the examination. It is observed that, the most well-denoised signal is achieved via the lpe graph-based filter. Notice that, this result concerns the most impulsive noise case, (i.e., for $\alpha = 1.1$ and for $\gamma = 4.5$) and the selected adjacency matrix is the covariation-based.

6.2. Performance evaluation on epilepsy dataset

In this section, the effectiveness of our graph filters, lp and lpe, is evaluated and compared with the WT and TV denoising methods in a different test case. Specifically, the performance of the four denoising methods compared herein is evaluated in an indirect way, by first applying the filters on raw EEG signals and then solving a classification problem.

Classification of epileptic versus non-epileptic subjects is a challenging problem. This is because epilepsy, apart from being one of the most important neurological disorders, has a unique behaviour. In particular, signals corresponding to epileptic seizures typically present many fluctuations and their amplitudes are greater than those of non-epileptic periods. However, the associated EEG signals are typically contaminated by higher noise than the conventional EEGs, due to the head motions. such a noise has proven to be better represented by members of the alpha-stable distributions family [19]. Thus our main scope is to examine how the graph filters treat the noise that is inherent to the original recordings, without adding synthetic noise as in the P300 dataset. Since we do not have access to the groundtruth (i.e., noiseless) EEG signals, a proper evaluation of the filtering efficiency is obtained by calculating the classification ratio.

Experimental Setup. Regarding the selected data, 7 non-overlapping segments of length 256 are randomly selected from each EEG examination. Each segment is filtered through the four denoising methods, followed by the training of a kNN classifier using the five

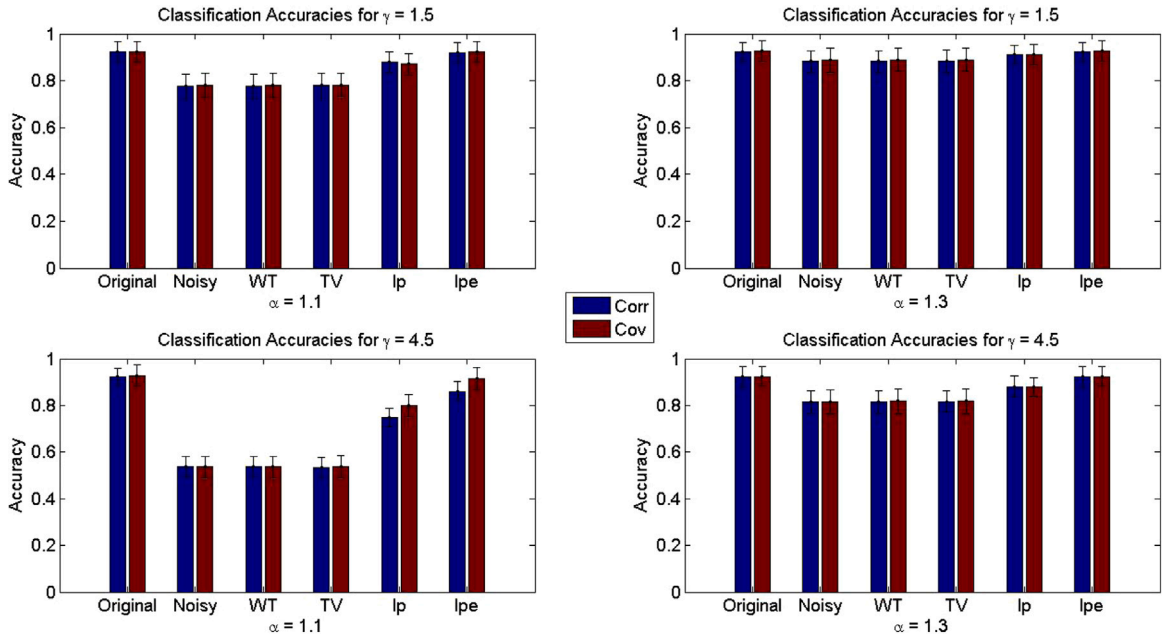

Fig. 4. Classification rates per adjacency matrix, for the four denoising methods, for $\gamma = 1.5$ experimental case and for the most impulsive noisy cases, i.e., for $\alpha = 1.1$ and $\alpha = 1.3$.(top-left, top-right figures);Classification rates per adjacency matrix, for the four denoising methods, for $\gamma = 4.5$ experimental case and for the most impulsive noisy cases, i.e., for $\alpha = 1.1$ and $\alpha = 1.3$.(bottom-left, bottom-right figures).

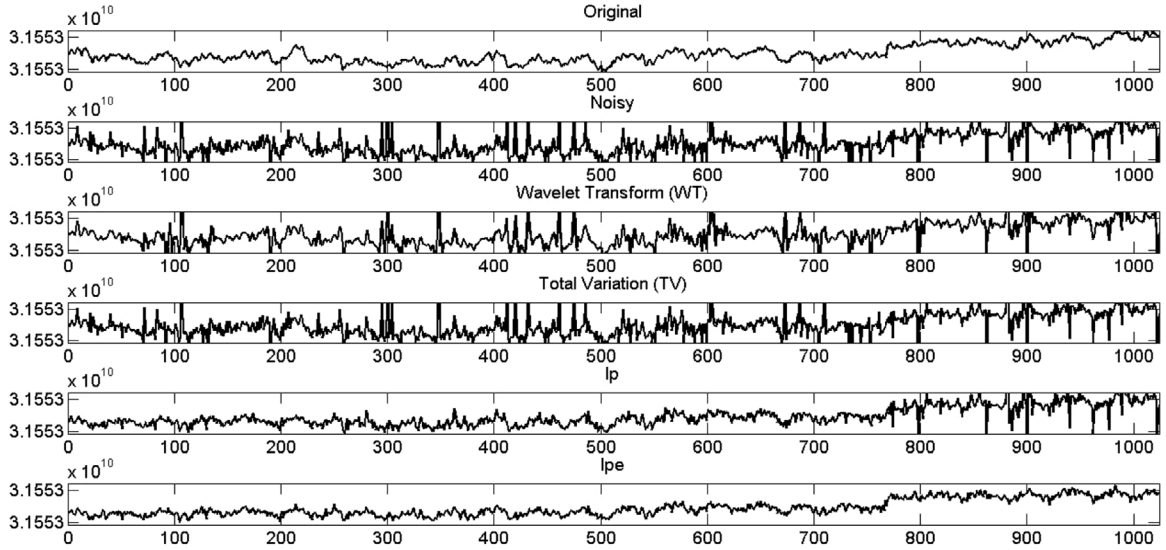

Fig. 5. Example of P300 dataset denoising. The original and noisy versions are presented as well as the denoising results per filtering method.

Table 2
Filtering methods parameters setting.

WT	TV	lp	lpe
db8	$b = 0.01$	$b = 0.1$	$b = 0.1$
2 levels	–	$q = 0.001$	$q = 0.001$
–	–	$\epsilon = 0.01$	$\epsilon = 0.01$
–	–	$I_{max} = 20$	$I_{max} = 20$

Table 3
Classification accuracy for the epilepsy dataset.

	Original	WT	TV	lp	lpe
corr	5/10	5/10	6/10	6/10	7/10
cov	5/10	5/10	6/10	6/10	**9/10**

features described in Section 5.1, that are extracted from each filtered signal. We note that a different model is trained for each filtering method. Given, the trained kNN classifiers, the classification ratio is calculated for each denoising method on the testing segments. Table 2 summarizes the parameters values used for each denoising method.

Classification Performance. For the calculation of the classification ratios, a majority voting scheme was employed, where the final label ("normal" or "abnormal") of a testing subject is the one with the highest frequency among its corresponding testing segments. Table 3 presents the classification accuracy for each denoising method and the two adjacency matrices, in terms of the ratio of the 10 testing subjects that are classified correctly.

Inspection of the results reveals that the $\ell_{p,\epsilon}$-regularized graph filter improves the classification accuracy by better adapting to the underlying heavy-tailed statistics of the raw data. Furthermore, for this

Table 4
Computational time (in sec) per denoising method.

	WT	TV	lp	lpe
corr	0.0593	0.0002	0.2045	0.2806
cov	0.0586	0.0002	0.2131	0.2864

test case, the choice of the adjacency matrix is shown to be critical. Indeed, the use of FLOM-based adjacency matrix can represent the EEG signals' graph structure more efficiently.

Finally, in this experimental case, as it was the most realistic scenario, we computed the computational time (in seconds) as an average over all the selected segments and the subjects (for training and testing) and in Table 4 we present them, for each denoising method.

It is observed that, the less demanding filtering method is the TV whilst, the most computational expensive are the lp and the lpe.

7. Discussion

EEG signals are highly susceptible to noise and this inherent problem hampers the efficient analysis, interpretation and feature/biomarker extraction both in clinical routine and in emerging techniques such as Brain–Computer Interface (BCI) applications [20]. This paper investigated the performance of well-established filters for the denoising of EEG signals corrupted by impulsive noise, and compared against our $\ell_{p,\epsilon}$-regularized graph filtering method based on fractional lower-order moments. The derived experimental results reveal the incapability of some of the pre-existing filters to treat properly the presence of impulsive noise, as opposed to our iterative graph filters. Specifically, all the denoising methods compared herein are applied on real EEG data corrupted by synthetic heavy-tailed noise generated by drawing samples from symmetric alpha-stable distributions as well as on their original versions (i.e., raw). We emphasize that the denoising process remains exactly the same both for synthetic and real noise conditions. Nevertheless, an advantage of performance evaluation using synthetic noise is that it allows us to simulate and examine a broad range of impulsive noise behaviours, thus deriving useful statistical results.

Regarding the synthetic noise scenario applied on P300 dataset, the impulsive noise simulation constitutes a realistic and satisfying scenario due to the evaluation of multiple impulsive noise cases (i.e., from very to less impulsive experimental cases) and the large number of Monte-Carlo repetitions. Additionally, the heavy-tailed distributions characterize the EEG noise effect more appropriately, as mentioned, rendering the proposed GSP alternative approach more robust for EEG denoising. Nevertheless, the results regarding the epilepsy scenario, i.e., the filtering of the raw epilepsy dataset, strongly support the hypothesis that our proposed graph filter combined with our the covariation matrix can be applied on real EEG data and provide the most robust results. Since EEG is still one of the most prominent techniques of examining brain function, the presented method can add value as a preprocessing tool for adding precision in clinical neuroscience or experimental analysis tasks. In particular, when the clinical task concerns complicated disease examinations, as in the case of epilepsy, the denoising of EEG signals is necessary.

Concluding, overall, the modified $\ell_{p,\epsilon}$-regularized graph-based EEG filtering method combined with an appropriate brain graph representation, i.e., via the cross-correlations or even the covariation adjacency matrix, is a powerful preprocessing step, not only for adding precision in classification tasks but also in problems which require a more thorough signal analysis, such as automatic seizure detection and emerging BCI applications.

8. Conclusions and future work

In this study, we examined the effectiveness of an alternative approach of a recently introduced graph filtering methodology and we introduced an $\ell_{p,\epsilon}$-regularized graph filter, which can better handle the effects of (i) additive impulsive noise on different EEG tasks and problems and (ii) the natural impulsive noise which exists on raw EEG recordings. Specifically, our proposed $\ell_{p,\epsilon}$-regularized graph-based filter, which depends on an iterative solution of a $\ell_{p,\epsilon}$-optimization problem through FLOMs combined with FLOMs-based adjacency matrix, is compared with well-established graph-based denoising techniques. Comparisons of several performance metrics prove that our proposed method is more effective, for a large range of impulsive noise, than the majority of the existing ones which cannot handle the heavy-tailed noise. Moreover, this alternative approach of the pre-existing ℓ_p-regularized graph-based filter proved to be more effective than its previous version, not only on very impulsive noise scenarios but also on less impulsive, i.e., when the noise distribution is close to Gaussian. Moreover, the multiple adjacency matrices evaluation proved that the graph-based representation of the EEG signals plays an important role, especially when the noise tends to be very impulsive. Finally, the classification procedure on raw and synthetic noisy EEG signals show that the denoised signals obtained by means of our proposed method are grouped in a more accurate way, than by the WT, TV and lp examined filtering cases, making so the filtering of signals a necessary pre-processing part. Overall, our proposed filtering method proved to be more robust and effective, among its counterparts, in treating the real noise of EEG epilepsy signals, which essentially is the most realistic scenario. The encouraging results presented call for more extended studies in diverse EEG applications to further prove the clinical potential of our approach. Finally, a challenging task could be how we could introduce the causality into our optimization problem.

CRediT authorship contribution statement

Anastasia Pentari: Software, Methodology, Writing. **George Tzagkarakis:** Co-Advisor, Writing. **Kostas Marias:** Biomedical evaluation. **Panagiotis Tsakalides:** Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] A. Kawala-Sterniuk, et al., Comparison of smoothing filters in analysis of EEG data for the medical diagnostics purposes, Sensors 20 (3) (2020) 807, http://dx.doi.org/10.3390/s20030807.
[2] M. Jadav, et al., Adaptive filtering and analysis of EEG signals in the time-frequency domain based on the local entropy, EURASIP J. Adv. Signal Process. (2020) http://dx.doi.org/10.1186/s13634-020-00667-6.
[3] Amin, et al., Classification of EEG signals based on pattern recognition approach, Front. Comput. Neurosci. 11 (2017) 103, http://dx.doi.org/10.3389/fncom.2017.00103.
[4] L. Ke, et al., Classification of EEG signals by multi-scale filtering and PCA, in: IEEE International Conference on Intelligent Computing and Intelligent Systems. Shanghai, 2009, pp. 362–366, http://dx.doi.org/10.1109/ICICISYS.2009.5357825.
[5] A. Turnip, et al., Artefacts removal of EEG signals with wavelet denoising, MATEC Web Conf 135 (6) (2017) 00058, http://dx.doi.org/10.1051/matecconf/201713500058.
[6] J.G. Gonzalez, et al., Weighted myriad filters: A robust filtering framework derived from alpha-stable distributions, in: IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings. Atlanta, GA, USA, Vol. 5, 1996, pp. 2833–2836, http://dx.doi.org/10.1109/ICASSP.1996.550143.
[7] R. Vigario, et al., Independent component approach to the analysis of EEG and MEG recordings, J. IEEE Trans. Biomed. Eng. 47 (5) (2000) 589–593, http://dx.doi.org/10.1109/10.841330.

[8] A. Pentari, et al., Graph-based denoising of EEG signals in impulsive environments, in: Proc. 28th European Signal Processing Conference (EUSIPCO '20), Amsterdam, The Netherlands, 2021.
[9] S. Chen, et al., Signal denoising on graphs via graph filtering, in: IEEE Global Conf. Signal & Inf. Process. Atlanta, GA, 2014, pp. 872–876, http://dx.doi.org/10.13140/2.1.4186.2727.
[10] Y. Schoenenberger, et al., Graph-based denoising for time-varying point clouds, in: 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON), 2015, pp. 1–4, http://dx.doi.org/10.1109/3DTV.2015.7169366.
[11] Ulrich Hoffmann, et al., An efficient P300-based brain–computer interface for disabled subjects, J. Neurosci. Methods 167 (1) (2008) 115–125.
[12] https://www.isip.piconepress.com/projects/tuh_eeg/downloads/tuh_eeg_abnormal/.
[13] G. Samorodnitsky, M. Taqqu, Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance, Chapman & Hall, New York, 1994.
[14] J. Nolan, Numerical calculation of stable densities and distribution functions, Commun. Statist.-Stoch. Models 13 (1997) 759–774.
[15] G. Tzagkarakis, et al., Compressive sensing using symmetric alpha-stable distributions for robust sparse signal reconstruction, J. IEEE Trans. Signal Process. 67 (3) (2019) 808–820, http://dx.doi.org/10.1109/TSP.2018.2887400.
[16] A. Pentari, et al., A study on the effect of distinct adjacency matrices for graph signal denoising, in: Proc. 20th BioInformatics And BioEngineering Conference (BIBE '20). Virtual conference, USA, 2020.
[17] W. Waheed, et al., Graph polynomial filter for signal denoising, IET Signal Process. 12 (3) (2018) 301–309.
[18] Y. Chen, Y, et al., Image denoising with adaptive weighted graph filtering, CMC-Comput. Mater. Continua 64 (2) (2020) 1219–1232.
[19] S. Supriya, et al., Analyzing EEG signal data for detection of epileptic seizure: Introducing weight on visibility graph with complex network feature, in: 27th Australasian Database Conference (ADC), 2016.
[20] R. Mamunur, et al., Current status, challenges, and possible solutions of EEG-based brain-computer interface: A comprehensive review, Front. Neurorobot. 14 (2020) http://dx.doi.org/10.3389/fnbot.2020.00025.