Filtering by Optimal Projection and application to automatic artifact removal from EEG

Samuel Boudet a,∗,1, Laurent Peyrodie a, Philippe Gallois b, Christian Vasseur c

a Hautes Etudes d’Ingénieur, ERASM-HEI Laboratory, Lille 59046 Fr
b St Vincent Hospital, Lille Fr
c Université de Lille 1, LAGIS Laboratory, Villeneuve d’Asc 59655 Fr

Abstract

A new approach to filter multi-channel signals is presented, called filtering by optimal projection (FOP) in this paper. This approach is based on common spatial subspace decomposition (CSSD) theory. Moreover, an evolution of this method for non-stationary signals is also introduced which is called adaptive FOP (AFOP). As ICA, a filtering matrix is set up in the best way to remove artifacts with linear combination of channels. This filtering matrix is characterized by two subspaces. The first one is determined during a learning phase, by finding components maximizing the ratio signal over noise. The second one will be determined during a filtering phase, by reconstructing signals of a sliding window, by a least square method. These methods are completely automated and enable to filter independently numerous artifact types. Moreover, this filtering can be improved by applying this process on frequency band decomposed signals.

Various tests have been made on electroencephalogram (EEG) signals in order to remove ocular and muscular activity while conserving pathological activity (slow waves, paroxysms). The results are compared with ICA filtering and medical inspection has been carried out to prove that this approach yields very good performance.

Key words: Automatic, Artifact removal, Filtering, EEG, FOP, Adaptative FOP

1. Introduction

In many applications of signal processing like biomedicine (EEG, MEG, etc.), preprocessing is required to remove artifact perturbations and thus improve signal interpretation [1]. The authors have been working for some years on various applications of EEG like anticipated detection of epilepsy seizure [2] and brain computer interface (BCI).

These applications require a completely automated signal processing and it has become of prime necessity to remove non-cerebral activity from these signals.

Filtering method can be broken down into two types. Temporal filter that uses information on the shape of the signal (e.g. FIR, IIR, Kalman filter, wavelet filter, etc.) and spatial filter that uses information of signal distribution on different channels (e.g. PCA [3], common spatial filter (CSP) [4]). Recently, progress has been made in this field with the apparition of independent component analysis (ICA) [5].

This paper presents a new approach to spatially filter multi-channel signals based on CSSD theory [6] which is called filtering by optimal projection (FOP) in this paper. An evolution of this method called adaptive FOP (AFOP) is also described. These methods are also improved by applying them on frequency band decomposition.

EEG signals are filtered with FOP and AFOP and results are compared with automated ICA. The quality of the results is validated by both visual and quantitative inspection.

First, a brief overview of spatial filtering will be introduced by presenting background on ICA. The principle of FOP and AFOP will be introduced by describing projection matrix properties. Next, mathematical aspects concerning subspace decomposition will be studied. Then, algorithms and their applications to artificial signals will be proposed. Finally, results will be presented and compared on electroencephalographic records.

∗ Corresponding author.

Email address: samuel.boudet@hei.fr (Samuel Boudet).

1 Present address: HEI, 13 rue de Toul, 59000, Lille, FR

2. Spatial Filter

Spatial filter aim to find the best filtering matrix to remove artifacts with linear combination of channels. Due to its performance, ICA is the most used method for artifact removal in scientific literature.

2.1. ICA filtering

The ICA problem was introduced by Herault and Jutten [7]. ICA supposes that various signal channels are a linear mixing of sources:

\[ V = MS \]  

(1)

\( V(m, T) \) is the signal matrix where \( m \) lines represent channels and \( T \) columns, time samples. \( S(n, T) \) represents signal matrix of \( n \) sources. The matrix \( M(m, n) \) is called mixing matrix. If the number of sources \( n \) is inferior or equal to the number of channels \( m \), \( M \) will be invertible. The pseudo-inverse matrix \( W = (M^T M)^{-1} M^T \) if number of sources \( n \) is inferior or equal to number of channels \( m \), \( M \) will be invertible. The pseudo-inverse matrix \( W = (M^T M)^{-1} M^T \) is called separating matrix.

ICA aims to estimate this separating matrix \( W(n, m) \) so that sources \( (S = VW) \) would be independent.

Once sources are defined, they are identified as artifactual or not and the artifactual ones are canceled:

\[ S' = DS \]  

(2)

where \( D \) is a diagonal matrix with zero on artifactual components and one on the others. The signals are then reconstructed by inverse transformation:

\[ V' = MDS \]  

(3)

\( V' \) represents filtered signals. A filtering matrix is then built using:

\[ F = MDW \]  

(4)

The Filtering matrix properties are described in section 2.2.

The separation criterion of sources used to determine \( W \) is an independence measure optimization. It can be shown that maximizing independence means maximizing the component non-gaussianity. There are several ICA algorithms using various measurements. Most of them aim to maximize the non-gaussianity of component distribution. For example, there is Infomax algorithm which aims to maximize Negentropy [8] and Jade and FastICA which aim to maximize the Kurtosis [9][10]. One of the major problems of ICA is the necessity to manually identify each component as artifactual or not. Various methods have been developed on this subject [11][12]. Vorobyov suggests using the Hurst exponent, considering that this is constant for all artifacts of the same type [14]. Another possibility is to use constrained ICA. The principle of this method is to find sources that are closed to the ones defined by constant bases [15]. The authors also proposed a method in [16], consisting in identifying ICA sources by comparing variance between rest instant and artifact instant. The filtering matrix is then considered as constant for the remaining record. This is the same idea developed in this paper, but the authors directly optimize this criterion of variance using FOP.

This optimization requires the projection matrix construction, which is why the projection principle is mentioned below.

2.2. Projection principle

In the following parts of the document, \( M|A \) (\( M \) is a matrix and \( A \) is a set of \( k \) entires \( k \leq n \)) refers to the sub-matrix formed of columns \( A \) of \( M \). In the same way, \( M|A \) refers to the sub-matrix formed of lines \( A \) of \( M \).

In order to understand better the theoretical aspects of FOP, it is necessary to mention idempotent matrix properties. A square matrix \( F \) of dimension \( n \) is said to be idempotent (or projector) if \( FF = F \) [17]. First the diagonalization of \( F \) is studied.

2.2.1. Diagonalization of \( F \)

In order to diagonalize \( F \), an eigenvector \( x \) associated to the eigenvalue \( \lambda \) is calculated

\[ Fx = \lambda x \]  

(5)

By definition \( FFx = \lambda Fx = \lambda^2 x \) and \( FFx = Fx = \lambda x \). So \( \lambda = 0 \) or 1 and diagonalization of \( F \) can be written as

\[ F = MDM^{-1} \]  

(6)

with

\[ D = \begin{pmatrix} I_{n_1} & 0_{n_1,n_2} \\ 0_{n_2,n_1} & I_{n_2} \end{pmatrix} \]  

(7)

where \( I_{n_i} \) is the identity matrix of dimension \( n_i \times n_i \). The matrix \( F \) then defines two subspaces \( E_1 \) and \( E_2 \) of respective dimensions \( n_1 \) and \( n_2 \) \( (n_1 + n_2 = n) \). These spaces are characterized by \( \forall x \in E_1, \ P x = x \) and \( \forall x \in E_2, \ P x = 0 \). The \( n_1 \) first column vectors of \( M \) are a base of \( E_1 \), the \( n_2 \) followings, a base of \( E_2 \). The matrix \( F \) is called projector on \( E_1 \) parallel to \( E_2 \).

2.2.2. Properties

(i) Matrix \( F^T \) is also a projector whose passage matrix is \( (M^{-1})^T \)

(ii) Matrix \( I - F \) (\( I \) refers to identity matrix) is the projector on \( E_2 \) parallel to \( E_1 \)

(iii) The passage matrix \( M \) is not unique but if \( M_1 \) and \( M_2 \) are two passage matrices corresponding to \( F \), \( M_1|^{1,\ldots,n_1} \) and \( M_2|^{1,\ldots,n_1} \) are bases of the same vectorial subspace \( E_1 \) (likewise for \( M_1|^{n_1+1,\ldots,n} \) and \( M_2|^{n_1+1,\ldots,n} \) with \( E_2 \)). So, the filtering matrix is completely defined by the two subspaces \( E_1 \) and \( E_2 \).

(iv) \( \text{Rank}(F) = \text{trace}(F) = \text{dim}(E_1) \)
(v) \( \mathbf{F} = \mathbf{M}^{1,\ldots,n_1} \mathbf{W}_{1,\ldots,n_1} \)

(vi) Vectorial subspace generated by base line vectors \( \mathbf{W}_{1,\ldots,n_1} \) is the orthogonal subspace to the one generated by base column vectors \( \mathbf{M}^{1,\ldots,n_1} \).

(vii) In the same way, vectorial subspace generated by base line vectors \( \mathbf{W}_{n_1+1,\ldots,n} \) is the orthogonal subspace to the one built by base column vectors \( \mathbf{M}^{n_1+1,\ldots,n} \). This is the fundamental property used in AFOP.

When the two subspaces \( E_1 \) and \( E_2 \) are orthogonal, the idempotent matrix is then called orthogonal projector. It is important to notice that the non-orthogonality of spaces can increase the norm of the filtering matrix sources. Indeed \( \|\mathbf{F}\| = \max_{\|\mathbf{x}\|=1}(\|\mathbf{F}\mathbf{x}\|) \) is minimal (= 1) when the angle between \( E_1 \) and \( E_2 \) is 90°. This non-orthogonality can then cause some stability problems, because a small noise can significantly change the position of this projection. Fig. 1 illustrates this property. Distance \( OA \) represents a noise and distance \( OB \) is the projected noise which is more important due to the fact that projection is not orthogonal.

![Fig. 1. Illustration of projection instability.](image)

### 2.3. Principle of FOP and adaptative FOP

Filtering by linear combination consists in calculating

\[
\mathbf{V}' = \mathbf{FV}
\]

Matrix \( \mathbf{F} \) is a projector. This property is verified in Section 3.2.2. Under the criteria defined in Section 3, the optimal matrix \( \mathbf{F} \) is a projector. It can be decomposed as Eq. (6). The purpose of FOP and AFOP is to find the two vectorial subspaces \( E_1 \) and \( E_2 \) which define \( \mathbf{F} \). \( E_1 \) refers to interesting activity repartition (\( \mathbf{M}^{1,\ldots,n_1} \)) and \( E_2 \), repartition of artifact activity (\( \mathbf{M}^{n_1+1,\ldots,n} \)). Properties 6 and 7 in Section 2.2.2 show that the orthogonal of \( E_1 \) denotes the subspace for constructing artifactual sources (\( \mathbf{W}_{1,\ldots,n_1} \)) and \( E_2 \), the subspace for constructing interesting sources (\( \mathbf{W}_{n_1+1,\ldots,n} \)).

The approach requires a learning process, consisting in comparing signals with few artifacts and with many artifacts. In order to define the problem well, hypotheses must be taken into account:

(i) First, as ICA, the signal is considered to be a linear mixture of \( n \) sources (\( n \leq m \)) that can be artifactual or not. FOP considers that this linear mixture is constant the whole time. Although, AFOP considers that this linear mixture is constant on the learning step but not for the remaining signal. It considers that only the artifact mixture is constant.

(ii) The second hypothesis similar to ICA is that artifactual sources are uncorrelated to others.

(iii) Variance of artifactual sources increases between rest periods (with few artifacts) and artifactual periods (with many artifacts). On the other hand, variance of non-artifactual sources does not increase significantly. However, it is important to notice that it is not necessary that the rest period does not contain artifacts at all. It is just required to contain significantly fewer artifacts than the artifact period.

FOP then considers the hypothesis that both \( E_1 \) and \( E_2 \) are constant and then the filtering matrix is constant. AFOP requires only hypothesis that repartition of artifacts (\( E_2 \)) is constant and so the construction space of interesting sources is considered also as constant. The principle of this method is then to detect this space by the learning process by comparing signals during rest and during artifacts, and to find sources that variance increases the least (Section 3.1). Once this step is over, the filtering process consists in finding the best distribution of these interesting sources on a sliding window thus characterizing \( E_1 \) (Section 3.2).

The number of artifactual sources is automatically defined by the sources where the variance increased more than a threshold.

### 3. Theory

#### 3.1. Detection of less artifacted sources

In order to execute the learning process, a subspace of non-artifactual sources (dimension \( n_1 \)) must be determined. This method, used here, can be also used to emphasize sources that are characterized to an increasing of variance on a defined time.

First, the case of \( n_1 = 1 \) is treated, and then, the results are generalized to other cases. Theoretical aspects are tested in Section 4.3, by the construction of a theoretical signal.

##### 3.1.1. Case of \( n_1 = 1 \)

The first purpose is to find the component maximizing the signal over noise ratio. This component is defined by \( \mathbf{w} \) maximizing \( R(\mathbf{w}) \) with

\[
R(\mathbf{w}) = \frac{||\mathbf{V}_r^T \mathbf{w}||^2}{||\mathbf{V}_a^T \mathbf{w}||^2}
\]

\( \mathbf{V}_r \) is the signal matrix during rest and \( \mathbf{V}_a \) is the signal matrix during artifacts. This expression can be simplified as follow

\[
R(\mathbf{w}) = \frac{(\mathbf{w}^T \mathbf{C}_r \mathbf{w})}{(\mathbf{w}^T \mathbf{C}_a \mathbf{w})}
\]

with \( \mathbf{C}_r \) and \( \mathbf{C}_a \) are respectively covariance matrices of rest instants and artifact instants. So, \( \mathbf{C}_r = \mathbf{V}_r \mathbf{V}_r^T \) and \( \mathbf{C}_a = \mathbf{V}_a \mathbf{V}_a^T \).

Maximizing \( R(\mathbf{w}) \) is equivalent to maximizing \( \ln R(\mathbf{w}) \).

\[
\ln(R(\mathbf{w})) = \ln w^T \mathbf{C}_r \mathbf{w} - \ln w^T \mathbf{C}_a \mathbf{w},
\]

\[
\nabla \ln R(\mathbf{w}) = 2 \frac{\mathbf{C}_r \mathbf{w}}{w^T \mathbf{C}_r \mathbf{w}} - 2 \frac{\mathbf{C}_a \mathbf{w}}{w^T \mathbf{C}_a \mathbf{w}}.
\]
\[ R(w) \text{ is maximal for } \nabla \ln R(w) = 0 \text{ that is to say:} \]
\[ \frac{w^T C_r w}{w^T C_r w} = C_r^{-1} C_r w \]

(11)

So, if \( w \) is solution, then \( w \) is eigenvector of \( C_r^{-1} C_r \) corresponding to eigenvalue \( \frac{w^T C_r w}{w^T C_r w} \). It must be verified that all eigenvectors are solution of (eq. 11). Supposing that \( \lambda_0 \) is an eigenvalue of \( C_r^{-1} C_r \) and \( w_0 \) is the corresponding eigenvector, the expression \( C_r^{-1} C_r w_0 = \lambda_0 w_0 \) is verified and \( \frac{w_0^T C_r w_0}{w_0^T C_r w_0} = \lambda_0 \)

(12)

and then the equation 11 is verified. Local extrema of \( R \) are given by eigenvectors of \( C_r^{-1} C_r \). In addition, on this vector \( R \) is equal to the corresponding eigenvalue. So, the maximum of \( R \) is given by the higher eigenvalue (\( \lambda_0 = R(w_0) \)). In practical problems all eigenvalues are different and positive (since \( C_a \) and \( C_r \) are positive definite), so the matrix \( C_r^{-1} C_r \) can then be diagonalized:

\[ C_r^{-1} C_r = \text{PDP}^{-1} \]

(13)

\( D \) is diagonal matrix composed of decreasing ordered eigenvalue \( \lambda_1, \lambda_2, \ldots, \lambda_n \). The columns of passage matrix \( P \) are eigenvectors normalised by norm \( ||.||_2^2 = < C_a, > \) (\( < \ldots \) designs the standard scalar product).

It can be proved by the principle of simultaneous orthogonalisation [17] that \( P \) is orthonormal concerning scalar product \( < C_a, > \) and orthogonal concerning the scalar product \( < C_r, > \). In other words \( P^T C_a P = I \) and \( P^T C_r P = D \) (This is due to the fact that \( C_r \) and \( C_a \) are symmetrical). However, it is important to notice that \( P \) is not orthogonal with the standard scalar product. The ratio \( R(w) \) can be then written as follow:

\[ R(w) = \frac{w^T P^{-1} DP^{-1} w}{w^T P^{-1} P^{-1} w} \]

(14)

Remark : In order to reduce calculation time, it is possible to not calculate \( C_a^{-1} \) in solving \( C_a w = \lambda C_a w \). The method to solve it is described in [18] [19] [20].

3.1.2. Case of \( n_1 > 1 \)

The aim is now to find vectorial subspace of dimension \( n_1 \) containing less artifacts. That is to say that for any vector \( w \) on this space, \( R(w) \) must be maximal. On a formal point of view, this space can be defined by the column vectors of a full rank matrix \( A \) (\( n \times n_1 \)) that maximizing \( \min_{x \in \mathbb{R}^{n_1}} R(Ax) \).

It can be shown that this matrix \( A \) is the matrix of \( n_1 \) first eigenvectors of \( C_r^{-1} C_r \). The demonstration can be made by taking \( B = P^{-1} A \). Then the ratio \( R(Ax) \) is transformed on:

\[ R(Ax) = R'(Bx) = \frac{x^T B^T DBx}{x^T B^T Bx} \]

(15)

A matrix \( B \) maximizing \( \min_{x \in \mathbb{R}^{n_1}} R'(Bx) \) is

\[ B = \begin{pmatrix} I_{n_1} \\ 0_{n_2, n_1} \end{pmatrix} \]

So, \( A = PB = P^{1 \ldots n_1} \). Taking the notation defined in (eq. 13), the space generated by the lines of \( P^T |_{1 \ldots n_1} \) is the estimation of the one generated by \( n_1 \) first lines of \( W \). The following notation will be then adopted:

\[ W_s = P^T |_{1 \ldots n_1} \]

(16)

It is important to notice that the proportion of artifact on a component is given by the eigenvalues. Indeed, the most artifacted components have low eigenvalue whereas non artifacted have eigenvalue near to 1 or upper. It would be then possible to characterize automatically a component as artifactual or not by including a threshold for the eigenvalues.

3.2. Best distribution of sources

Once the non-artifacted source subspace is detected, the aim will be to reconstruct the channels signal by least square method using only sources of this subspace.

3.2.1. General case

Considering any \( n_1 \) column signals \( (S_1, S_2, \ldots, S_{n_1}) = S^T \) (For now, \( S \) does not represent linear combination of \( V \)), a linear combination \( x \) of these signals is expected to minimize the distance with any signal \( v_0^{T} \).

Then \( x \) aims to minimize \( d(x) \) with \( d(x) = ||v_0^T - S^T x||^2 \), that is to say \( d(x) = ||v_0^T||^2 + ||S^T x||^2 - 2(v_0^T, S^T x) \).

The minimum of \( d(x) \) is reached for \( \nabla d(x) = 0 \).

\[ \nabla d(x) = 2S^T S x - 2S^T V \]

(17)

So the minimum of \( d(x) \) is given by \( x = (SS^T)^{-1} v_0^T \).

This operation can be repeated on \( n \) signals \( (v_1, v_2, \ldots, v_n) = V^T \). So, the distance between \( V \) and \( XS \) must be minimized with \( X \) a \( n_1 \times n \) matrix. The best repartition is then:

\[ X = (SS^T)^{-1} S V^T \]

(18)

3.2.2. Case of \( S \) is a linear combination of \( V \)

Adapative FOP considers a temporal sliding window. \( V_t \) is then the matrix of channel signal during this window. \( S_t \) is a linear combination of \( V_t \) which represents interesting source signal. Considering result of Eq. (16), this expression can be written as \( S_t = P^T |_{1 \ldots n_1} V_t \). Therefore, Eq. 18 can be written as:

\[ X = (P^T |_{1 \ldots n_1} V_t V_t^T P |_{1 \ldots n_1} - 1) P^T |_{1 \ldots n_1} V_t V_t^T \]

So,

\[ X = (P^T |_{1 \ldots n_1} C_t P |_{1 \ldots n_1} - 1) P^T |_{1 \ldots n_1} C_t \]

(19)

where \( C_t \) denoted the covariance matrix of channels. \( C_t = X V_t V_t^T X \) is then an estimator or \( \hat{M}^T |_{1 \ldots n_1} \) and the estimated filtering matrix is then

\[ \hat{F} = C_t P |_{1 \ldots n_1} (P^T |_{1 \ldots n_1} C_t P |_{1 \ldots n_1} - 1) P^T |_{1 \ldots n_1} \]

(20)

This matrix is well an idempotent matrix \( (\hat{F} \hat{F} = \hat{F}) \).
3.2.3. Interesting properties on application to learning moments

The principle of standard FOP method resides in executing AFOP on the learning instants and then to consider the filtering matrix as constant for the remaining signal. Considering $V_t = V_r$: $X = (P_{\mu_1}^T C_{\mu_1} P_{\mu_1}^{-1})^{-1} P_{\mu_1}^T C_{\mu_1}$. So $X = D_{n_1}^{-1} P_{\mu_1}^{-1} D P_{\mu_1}^{-1}$. The threshold is then fixed in such a way that 90% of component variance ratios are lower than the threshold. If an eigenvalue is smaller than this threshold, this means that the corresponding component will be optimal at the same time if it is considered as artifactual or not.

Therefore, all eigenvectors of matrix $C_a^{-1} C_{\mu_1}$ defined in Eq. 13 can be consider as components of the filtering matrix. So the estimated filtering matrix used with FOP for the entire signal is:

$$\hat{F} = P^{-1} D_{n_1} P^T \quad (22)$$

$D_{n_1}$ is a diagonal matrix with 1 on the $n_1$ first diagonal elements and 0 on the others.

- It can be noticed that these components ($P^T$) do not depend of the value $n_1$. That is to say that a component will be optimal at the same time if it is considered as artifactual or not.
- It can also be notified that considering $V_t = V_a$, the same results can be obtained. This proves that presence of artifacts does not affect the repartition of interesting sources.
- Finally, It is possible to notice that, if matrix $C_a = I$, then the FOP method is equivalent to ACP.

4. Methodology

4.1. Algorithm

The FOP and AFOP algorithms can be summed up in Fig. 2. For real time using, FOP method consists in directly applying the filtering matrix on the signal once it is determined. AFOP consists in applying the process of source distribution on the last seconds of the recorded signal.

In order to determine the number of sources which are artifactual, the eigenvalues defined in Section 3.1 are compared to a threshold. If an eigenvalue is smaller than this threshold, this means that the corresponding component has an increased variance during the artifact moment. This threshold can be fixed in comparing two rest moments and calculating Eq. (13). The threshold is then fixed in such a way that 90% of component variance ratios are lower than the threshold.

FOP and AFOP can also be used when the interesting signal has an increased variance between the first and the second instant whereas the noise has a constant variance. The principle is then to consider $V_a$ as a rest instant and $V_r$ as the interesting instant and to carry out the same process excepting that the threshold has a value $> 1$.

With AFOP, it is also possible to consider that the interesting source distribution ($\mathbf{M}|_{1,...,n_1}$) is constant instead of artifactual sources ($\mathbf{M}|_{n_1+1,...,n}$). Using the same process, the estimated filtering matrix will be:

$$\hat{F} = \hat{I}_n - C_a \mathbf{P}|_{n_1+1,...,n} \times (\mathbf{P}^T|_{n_1+1,...,n} C_a \mathbf{P}|_{n_1+1,...,n})^{-1} \mathbf{P}^T|_{n_1,1,...,n} \quad (23)$$

These models can be better for some problems. There are many applications. For example, this can be useful in order to isolate electrooculogram or a P300 response of a stimulation from EEG. With the AFOP method, these signals will be isolated even if unexpected signals occur on a subspace not defined on the learning moments. First tests have been made on BCI P300 speller [21] [22] and proved that this method can emphasize the P300 response signal.

4.2. Computation time

Computation time is very short. Experiments have been carried out on Pentium IV 3.0GHz with the matlab pro-

**Figure 2. Algorithm of FOP and Adapative FOP**
gram using the EEGLab interface [23]. Calling \( n \) the number of channels and \( m \) the number of time samples, the calculation of covariance matrix is in \( O(n^2m) \) and the diagonalization in \( O(n^3) \). Taking \( n = 1200 \) and \( m = 20000 \) which are values much higher than those used in common problems, the calculation of the covariance matrix lasts about 30 s as does the diagonalization. This method can then be applied to huge dimension problem.

4.3. Theoretical signal

4.3.1. FOP testing

In order to prove FOP method efficiency, theoretical aspects are tested with the model described as follows. A five channel signal \( \mathbf{V} \) is a mixture of five sources, the first two being artifactual. The authors takes for example the mixing matrix:

\[
\mathbf{M} = \begin{pmatrix}
1 & 0 & 1 & 0 & 3 \\
1 & 1 & -1 & 0 & 2 \\
1 & 0 & 0 & 1 & 1 \\
1 & 1 & 0 & 1 & 1 \\
1 & 0 & 0 & -1 & 0
\end{pmatrix}
\] (24)

Sources \( \mathbf{S} \) are broken down into two moments of 10,000 time samples: \( \mathbf{S}_r \), corresponding to rest period and \( \mathbf{S}_a \), corresponding to artifact period.

During the first one, \( \mathbf{S}_{r,i,j} \) follows a normal law of means 0 and variance 1 for all \( i,j \). During the second one (\( \mathbf{S}_a \)), the variance of the first two lines is multiplied by 2 and the others keep a variance of 1.

The two covariance matrices \( \mathbf{C}_a \) and \( \mathbf{C}_r \) are computed and the matrix \( \mathbf{C}_a^{-1} \mathbf{C}_r \) is diagonalized. Three eigenvalues are close to 1 1.01 0.98 0.97). The corresponding eigenvectors \( (\mathbf{P}|^{1,\ldots,3}) \) represent an estimation of space of interesting source reconstruction and so this is the orthogonal to artifact spaces (Section 2.3). The two other eigenvalues are close to 0.5 (0.49 0.52). The corresponding eigenvectors represent an estimation of artifactual source reconstruction space.

It is important to notice that the basis \( \mathbf{P}|^{1,\ldots,3} \) is very different from the theoretical one but they represent approximatively the two same spaces. Indeed the angle between \( E_1 \) and its estimation is 5.14° and between \( E_2 \) and its estimation 2.37°. It can be noticed also that the smaller angle between \( E_1 \) and \( E_2 \), the more inaccurate the estimation.

Under the hypothesis described above, the theoretical filtering matrix of \( \mathbf{V} \) is

\[
\mathbf{F} = \mathbf{MDM}^{-1}
\] (25)

where \( \mathbf{D} \) is a 5 x 5 diagonal matrix with 0 for the first two components and 1 on the others.

Using only \( \mathbf{V}_r \) and \( \mathbf{V}_a \) (respectively, channel signals during rest and artifact), FOP can approximate the filtering matrix (\( \hat{\mathbf{F}} \)) for those two instants (Eq. (22). Fig. 3 illustrates these various steps and shows that the estimated filtered signal is very close to the theoretical one.

4.3.2. AFOP testing

In order to test AFOP a third period \( \mathbf{S}_t \) are randomly computed. \( \mathbf{S}_t \) is constructed in the same way as \( \mathbf{S}_r \). A sec-

![Fig. 3. Signal samples at various step. (Time 0-10 : rest period, Time 10-20 : artifact period)](image-url)
Fig. 4. Signal samples at various step during non learned moment (rest (10 s) and artifact (10 s))

ond mixing matrix $M_t$ is taken in such a way that the first two columns are the same as $M$ (distribution of artifactual sources is constant). So, the authors take:

$$M_t = \begin{pmatrix}
1 & 0 & 1 & 2 & -1 \\
1 & 1 & 2 & -1 & 0 \\
1 & 0 & 0 & 1 & 1 \\
1 & 1 & 0 & -1 & 1 \\
1 & 0 & -1 & 2 & 2
\end{pmatrix}$$

(26)

AFOP method gives an approximation of new filtering matrix $F_t = M_tD^{-1}M_t^*$ using only $V_r$, $V_a$ and $V_t$ ($V_t = M_tS_t$) (Eq. 20). The estimated filtering matrix is very closed to the theoretical one.

$$F = \begin{pmatrix}
0.75 & 0.75 & -2 & -0.75 & 1.25 \\
-0.75 & 1.25 & 1 & -1.25 & -0.25 \\
-0.25 & 0.75 & -1 & -0.75 & 1.25 \\
-0.75 & 0.25 & 1 & -0.25 & -0.25 \\
-0.25 & 0.75 & -2 & -0.75 & 2.25
\end{pmatrix}$$

$$\hat{F} = \begin{pmatrix}
0.7227 & 0.7605 & -1.9515 & -0.7918 & 1.2348 \\
-0.7356 & 1.2516 & 0.9660 & -1.2213 & -0.2474 \\
-0.2753 & 0.7476 & -0.9113 & -0.7825 & 1.2109 \\
-0.7444 & 0.2419 & 1.0218 & -0.2296 & -0.2758 \\
-0.2972 & 0.7365 & -1.8181 & -0.8095 & 2.1666
\end{pmatrix}$$

Fig. 4 illustrates various steps and shows that the estimated filtered signal is very close to the theoretical one.

5. Application to EEG filtering

5.1. EEG description

The statistical analysis of electrical recordings of brain activity by an electroencephalogram (EEG) is a major problem in neuroscience. One of the major applications is for the diagnostic of epilepsy [24]. Therefore, noise removal is a prime necessity to make easier data interpretation and representation, and to recover the signal that perfectly matches brain activity.

The authors have been working on a standard examination with a 19 electrodes 10/20 system. This examination lasts about 20 min. During this time, the patient is exposed to flashing lights at various frequencies and also carries out two hyperventilations which consist in breathing quickly and deeply. These tests can reveal pathological trouble or cause epilepsy seizures.

5.2. Problem definition

During the standard examination, some elements must be kept due to the fact that they correspond to cerebral activity. It can be noticed especially:

- Paroxysms (epilepsy indicators). They are graphical elements that can be seen under various forms. The most common one is called spike wave. The paroxysms belongs to a very large frequency band (130 Hz) which makes the numerical filtering difficult to use.

- Reactivity to eye closure. This is characterized by an alpha rhythm apparition (813 Hz) mainly located in the occipital region.

- Slow waves ($< 4$ Hz). These waves can be seen as pathological for the adult. However, this activity is normal for people under 25 years old during hypopnoea.

The main purpose of the filtering process will be to keep these signals as well as possible and erase the ones with
artifactual origins. The artifact types can be broken down as follows:

- Eye blink.\cite{25} It is represented by a low frequency signal (<4 Hz) that can be important in amplitude. It is a symmetrical activity mainly located on the front electrodes (FP1, FP2) with a low propagation.
- Eye movement.\cite{25} It is also represented by a low frequency signal (<4 Hz) but with a higher propagation. It is caused by the fact that eyes represent dipoles and their movements lead to an alteration of the electrical field. It is characterized by a dissymmetry between the two hemispheres.
- Forehead movement. It is mainly a high frequency activity (>13 Hz) due to its muscular origin. However, slight electrode displacement can be observed on low frequencies (<4 Hz).
- Jaw clenching. It is also a high frequency (>13 Hz) and muscular activity and may also cause some low frequencies.

It is obvious that standard digital filter cannot separate artifacts of cerebral activity due to the fact that they can have the same frequency. However, it is important to conserve this information.

5.3. **Segmentation in frequencies bands**

The methods described here allow to spatially characterize the artifact. However, artifacts can be defined as well as by their frequencies. This is the reason why the method is applied on a segmented signal in frequency bands. Filtering matrix is then computed for each frequency band.

The aim of this study is to filter according to information of both distribution and frequency. Another advantage of this method is that it enables the number of sources to be increased.

Here is a description for using FOP on frequency band decomposition:

(i) **Learning process**:
- Cut each channel in frequency bands.
- Make the learning process described in Section 4.1 of either FOP or AFOP on each frequency band and obtain matrix \( \hat{F} \) or \( \hat{W} \).

(ii) **Filtering process**:
- Cut each channel in frequency bands.
- Make the filtering process described in Section 4.1 on each frequency band.
- Rebuild the signal by adding up the signals of each frequency band.

Various tests were carried out to determine the best choice for these frequency bands. The common segmentation used by neurologists was chosen namely:

(i) The band \( \Delta \) (0–4 Hz): Corresponds to a slow rhythm, often pathological, and corresponds as well as to the speed of a human motor movement (blinks or eye movements or electrode displacement).
(ii) The band \( \theta \) (4–8 Hz): Generally in low quantity, with a bitemporal localization. It may reveal some abnormalities.
(iii) The band \( \alpha \) (8–13 Hz): Generally poor in artifacts. The signal is located mainly in the occipital regions when eyes are closed.
(iv) The band \( \beta \) (>13 Hz): Generally characterized by muscular activity. Taking account of its width, a more precise segmentation is carried out for the followed frequency bands (13–20 Hz), (20–30 Hz) and (>30 Hz).

It can be noticed that only a few cerebral sources are saved in the last frequency band (>30 Hz). This is due to the fact that noise appears in a non-synchronized way in this band, thus showing the limitation of the method.

The results of this method are shown in Figs. 5c,e and g.

5.4. **Protocol**

The FOP and AFOP methods require a learning step to learn distribution of artifactual activity. During 2 min at the beginning of the recording, patients have then to carry out four artifact types several times. This period is then compared to a rest period (without artifacts) for the learning step. The four artifact types are:

- eye blinks
- eye movements
- jaw clenching
- forehead movements

This list is sufficient to filter most artifacts of EEG except perhaps for some very uncommon artifacts like blinks of one eye. However, as noticed in Section 5.2, some artifacts cause electrode displacement or dipole movement that is difficult to filter because they are too unpredictable.

Various parameters and settings have been fixed empirically:

- The threshold defined in Section 4.1 can be fixed to a value of 0.3. It has been fixed by comparing two resting periods (without artifacts) in the same conditions. Carrying out this experience several times, this threshold is set so that it is not reached in about 90% of the cases. That is to say that, in 90% of the cases, if there are no artifacts, no components will be considered as artifactual.
- The training step must contain several repetitions of each artifact especially for the AFOP because it requires a more precise learning. A minimum of 20 eye blinks, 15 eye movements, eight jaw clenchings and eight forehead movements.
- This method can operate even if there are a few artifacts during the resting time used in learning.
- These artifacts can be learnt with closed or opened eyes but the resting time should contain eyes closed EEG.
- The rest moment must last at least 30 s, and each artifact type must last at least 15 s.
- For the AFOP, the sliding window of 30 s is convenient but 15 s or 60 s gives approximately the same results.
5.5. Hypotheses verifications

In reality, the source number is much higher than the electrode number due to the fact that every neurone represents sources. However, when making PCA on resting EEG, it can be proved that 96% of explanatory variance resides on the first seven components. This proves that seven sources can be considered in order to reconstruct almost all the cerebral signals. Experiments show that artifacts can also be considered as \( m_a \) sources with \( m_s + m_a \leq m \).

The fact that the artifact distribution is constant can be easily proved by the fact that the orthogonal sources do not contain any artifacts. However, it may be different for the cerebral distribution. Indeed, for example, slow waves can appear on a space which has not been taken into account during the learning step. The AFOP method then shows its interest.

5.6. Results

Twenty records of different patients have been tested with the protocol described in Section 5.4, five of them contain pathological slow waves and two contain paroxysms. Each record has been automatically filtered using ICA, FOP and AFOP, with and without frequency band decomposition. Examples of results have been shown in Fig. 5. Since the standard method of ICA is not automated, the authors work on an adaptation developed in [16].

With visual inspection by a neurologist, it can be noticed that the frequency band decomposition seems to greatly improve the quantity of removed artifacts for all methods. Both FOP and ICA (with frequency band decomposition) filter nearly completely all artifacts but at the same time, they reduce a little cerebral activity during slow waves and paroxysms especially on channels FP1 and FP2 that are usually the most artifacted channels. On the other hand, perfectly AFOP restores all cerebral activity and when the artifacts are well learned the filtering is very efficient. However, this method seems to be less stable than standard FOP concerning artifact removal. Indeed, electrode displacements are usually kept because they are not located on precise enough vectorial space. It is the same for high frequency muscular activity that is usually less reduced than the standard FOP. However, this instability can be controlled by measuring the angle between the two vectorial subspaces. Indeed when the angle is small (< 8°), this means that the first space is not well defined enough and therefore the AFOP method must not be used.

In order to complete visual analysis of EEG, two ratios are computed on the filtered EEG with frequency band decomposition. These ratios aim to quantify filtering quality on various frequency bands \( (\Delta, \theta, \alpha, \beta) \).

Tables 13 represent the proportion of conserved cerebral signals. They can be seen as the ratio of standard deviation between filtered signals and unfiltered signals. These ratios are calculated at various times with grapho-elements. A ra-

<table>
<thead>
<tr>
<th>Table 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICA : ( \sigma_{\text{Filtered instant}} / \sigma_{\text{Unfiltered instant}} )</td>
</tr>
<tr>
<td>Frequency band &amp; ( \Delta ) &amp; ( \theta ) &amp; ( \alpha ) &amp; ( \beta )</td>
</tr>
<tr>
<td>Rest (opened eyes) &amp; 0.81 &amp; 0.95 &amp; 0.99 &amp; 0.88</td>
</tr>
<tr>
<td>Rest (closed eyes) &amp; 0.84 &amp; 0.96 &amp; 1.00 &amp; 0.89</td>
</tr>
<tr>
<td>Cerebral slow waves &amp; 0.82 &amp; 0.93 &amp; 1.00 &amp; 0.90</td>
</tr>
<tr>
<td>Paroxysms &amp; 0.74 &amp; 0.96 &amp; 1.02 &amp; 0.94</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>FOP : ( \sigma_{\text{Filtered instant}} / \sigma_{\text{Unfiltered instant}} )</td>
</tr>
<tr>
<td>Frequency band &amp; ( \Delta ) &amp; ( \theta ) &amp; ( \alpha ) &amp; ( \beta )</td>
</tr>
<tr>
<td>Rest (opened eyes) &amp; 0.83 &amp; 0.95 &amp; 0.99 &amp; 0.89</td>
</tr>
<tr>
<td>Rest (closed eyes) &amp; 0.83 &amp; 0.97 &amp; 0.99 &amp; 0.90</td>
</tr>
<tr>
<td>Cerebral slow waves &amp; 0.82 &amp; 0.98 &amp; 1.01 &amp; 0.92</td>
</tr>
<tr>
<td>Paroxysms &amp; 0.72 &amp; 0.91 &amp; 1.02 &amp; 0.92</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adaptive FOP : ( \sigma_{\text{Filtered instant}} / \sigma_{\text{Unfiltered instant}} )</td>
</tr>
<tr>
<td>Frequency band &amp; ( \Delta ) &amp; ( \theta ) &amp; ( \alpha ) &amp; ( \beta )</td>
</tr>
<tr>
<td>Rest (opened eyes) &amp; 0.88 &amp; 0.96 &amp; 0.99 &amp; 0.89</td>
</tr>
<tr>
<td>Rest (closed eyes) &amp; 0.87 &amp; 0.98 &amp; 0.99</td>
</tr>
<tr>
<td>Cerebral slow waves &amp; 0.90 &amp; 0.99 &amp; 1.00 &amp; 0.92</td>
</tr>
<tr>
<td>Paroxysms &amp; 0.93 &amp; 0.97 &amp; 1.00 &amp; 0.93</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICA : ( \sigma_{\text{Artefact}} / \sigma_{\text{Rest (eyes closed)}} ) (After filtering / Before filtering)</td>
</tr>
<tr>
<td>Frequency band &amp; ( \Delta ) &amp; ( \theta ) &amp; ( \alpha ) &amp; ( \beta )</td>
</tr>
<tr>
<td>Eye blink &amp; 1.08/3.42 &amp; 0.83/1.90 &amp; 0.69/0.78 &amp; 0.93/1.13</td>
</tr>
<tr>
<td>Eye movement &amp; 1.13/5.05 &amp; 0.90/1.55 &amp; 0.65/0.69 &amp; 0.95/1.23</td>
</tr>
<tr>
<td>Fronthead &amp; 1.20/2.63 &amp; 0.94/1.15 &amp; 0.87/1.08 &amp; 0.91/2.53</td>
</tr>
<tr>
<td>Jaw &amp; 1.11/1.74 &amp; 0.92/1.00 &amp; 0.99/1.02 &amp; 1.08/3.92</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>FOP : ( \sigma_{\text{Artefact}} / \sigma_{\text{Rest (eyes closed)}} ) (After filtering / Before filtering)</td>
</tr>
<tr>
<td>Frequency band &amp; ( \Delta ) &amp; ( \theta ) &amp; ( \alpha ) &amp; ( \beta )</td>
</tr>
<tr>
<td>%hline Eye blink &amp; 1.04/3.42 &amp; 0.81/1.90 &amp; 0.69/0.78 &amp; 0.93/1.13</td>
</tr>
<tr>
<td>Eye movement &amp; 1.10/5.05 &amp; 0.89/1.55 &amp; 0.63/0.69 &amp; 0.95/1.23</td>
</tr>
<tr>
<td>Fronthead &amp; 1.20/2.63 &amp; 0.95/1.15 &amp; 0.85/1.08 &amp; 0.91/2.53</td>
</tr>
<tr>
<td>Jaw &amp; 1.09/1.74 &amp; 0.95/1.00 &amp; 0.98/1.02 &amp; 1.09/3.92</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adaptive FOP : ( \sigma_{\text{Artefact}} / \sigma_{\text{Rest (eyes closed)}} ) (After filtering / Before filtering)</td>
</tr>
<tr>
<td>Frequency band &amp; ( \Delta ) &amp; ( \theta ) &amp; ( \alpha ) &amp; ( \beta )</td>
</tr>
<tr>
<td>Eye blink &amp; 1.85/3.42 &amp; 0.99/1.90 &amp; 0.69/0.78 &amp; 0.95/1.13</td>
</tr>
<tr>
<td>Eye movement &amp; 1.66/5.05 &amp; 0.91/1.55 &amp; 0.63/0.69 &amp; 0.95/1.23</td>
</tr>
<tr>
<td>Fronthead &amp; 1.87/2.63 &amp; 1.02/1.15 &amp; 0.83/1.08 &amp; 1.18/2.53</td>
</tr>
<tr>
<td>Jaw &amp; 1.50/1.74 &amp; 0.98/1.00 &amp; 0.98/1.02 &amp; 1.32/3.92</td>
</tr>
</tbody>
</table>
(a) Original signals

(b) Filtered signals with ICA

(c) Filtered signals with ICA on frequency band decomposition

(d) Filtered signals with FOP
tio close to one means that the signal is well conserved, and a ratio close to zero means the signal is removed. These tables show that the α activity is kept at 99% for all methods, but slow waves and paroxysms are slightly reduced with the FOP and ICA methods (about 82% in low frequencies). On the other hand, AFOP keeps this activity much more (90% and 93%).

Tables 46 represent the proportion of remaining artifacts after and before filtering. They can be interpreted as the ratio of standard deviation between various artifact time and rest time. The ratio value which is calculated before and after filtering proves that all artifacts are greatly reduced with FOP and ICA. However, AFOP can have worse results which are mainly due to the fact that the learning is not accurate enough.

In conclusion, the standard FOP can guarantee that all artifacts are filtered and AFOP guarantees that all cerebral activity is kept.

6. Conclusion and future works

FOP and AFOP are automatic methods for global artifact filtering. Indeed, once the rest period and artifact period are identified, these methods do not require any human intervention contrary to ICA that requires a manual identification of component. In addition, the method is independent of artifact types (e.g. muscular and ocular activity interfering EEG). These methods also have the advantage of not carrying out any artifact detection [26] and thus avoiding errors in recognition of artifact or interesting signals. However, it may be interesting to use it in order to adapt the best filter to the instant.
Using these methods on band frequency decomposed signal, they provide excellent results. Besides, they can be used on other temporal decomposition like wavelet decomposition, thus providing better results, depending on the problem.

Generally, AFOP provides better guarantees that interesting signals are not filtered. However, some instability may occur if learning is not precise enough and then artifacts will not be well filtered. This instability can be controlled measuring the angle between two subspaces $E_1$ and $E_2$. However, in this case, FOP can provide better results.

The authors are working on a solution to improve this stability.

Concerning theoretical aspects, it is very close to PCA. Consequently, there are numerous properties justifying the modeling. Moreover, the formal resolution of the problem enables computation time to greatly reduce. This method can then be used on huge dimensions problem and on the other way the model can be extended without having to consider the time calculation problem.

A limitation of these methods would be that they do not take into account dephasing of sources. As shown in Section 5.6, this can cause trouble on high frequencies. This is the reason why the authors are working on an adaptation taking into account temporal gaps between channels.

Using various adaptations of this method described in Section 4.1, the applications can be numerous. First, tests have been made to prove that this method can emphasize various kinds of signal.

7. ACKNOWLEDGMENTS

The authors gratefully acknowledge the medical personnel of Hospital St Vincent and St Philibert of Lille. They are also obliged for the financial support of AINF association and FNS and for the ERASM HEI laboratory.

References