Single-channel electroencephalogram analysis using non-linear subspace techniques

A.R. Teixeira †, N. Alves†, A. M. Tomé †,M. Böhm ‡, E. W. Lang ‡, C. G.Puntonet §
†DETI- IEETA, Universidade Aveiro,3810-193 Aveiro, Portugal
‡Institute of Biophysics University of Regensburg, D-93040 Regensburg, Germany
§ESTII- University of Granada, 18071-Granada- Spain

Abstract **– In this work, we propose the correction of univariate, single channel EEGs using projective subspace techniques. The biomedical signals which often represent one dimensional time series, need to be transformed to multi-dimensional signal vectors for the latter techniques to be applicable. The transformation can be achieved by embedding an observed signal in its delayed coordinates. We propose the application of two non-linear subspace techniques to the obtained multidimensional signal. One of the techniques consists in a modified version of Singular Spectrum Analysis (SSA) and the other is kernel Principal Component Analysis (KPCA) implemented using a reduced rank approximation of the kernel matrix. Both nonlinear subspace projection techniques are applied to an electroencephalogram (EEG) signal recorded in the frontal channel to extract its prominent electrooculogram (EOG) interference.**

Keywords **– Subspace Techniques, local SSA, KPCA, EOG , EEG, Removing Artifacts**

I. Introduction

In many biomedical signal processing applications a sensor signal is contaminated with artifact signals as well as with noise signals of substantial amplitude. The former sometimes can be the most prominent signal component registered, while the latter is often assumed to be additive, white, normally distributed and non-correlated with the sensor signals. Often signal to noise ratios (SNR) are quite low. Hence to recover the signals of interest the task is to remove both the artifact components as well as the superimposed noise contributions. Projective subspace techniques can then be used favorably to get rid of most of the noise contributions to multidimensional signals. But many biomedical signals represent one dimensional time series. Clearly projective subspace techniques are not available for one dimensional time series to suppress noise contributions, hence time series analysis techniques often rely on embedding a one dimensional sensor signal in a high-dimensional space of time-delayed coordinates [1], [2], [3]. Correlations in these multidimensional signal vectors together with second order techniques can be used to decompose the signal into uncorrelated components. The multidimensional signal is then projected to the most significant directions computed using singular value decomposition (SVD) or principal component analysis (PCA). Singular spectrum analysis (SSA) [4] used in climatic, meteorologic and geophysics data analysis is the most widely used technique that follows this strategy. The general purpose of SSA is to decompose the embedded signal vectors into additive components. This can be used to separate noise contributions from a recorded signal by estimating those directions, corresponding to the L largest eigenvalues, which can be associated with the eigenvectors spanning the signal subspace. The remaining orthogonal directions then can be associated with the noise subspace. Reconstructing the signal using only those L dominant components then can result in a substantial noise reduction of the recorded signals. The time embedding of the sensor signals transforms one dimensional time series into multidimensional signal vectors. This is a necessary step to have subspace projection techniques available. However, this step unavoidably introduces a nonlinearity into the signal analysis process. Of course, there also exist generically nonlinear signal processing techniques like kernel-PCA (KPCA) [5] which are often used for denoising. So it will be of interest to explore these techniques in their ability to remove dominant artifacts and/or suppress noise. The kernel techniques are based on the mapping of the input data by a non-linear function. In feature space then a linear PCA is performed estimating the eigenvectors and eigenvalues of a matrix of dot products (kernel matrix).

In this work we will present the concept of local SSA which means that after the time embedding we cluster the resulting multidimensional signal vectors and apply the linear signal decomposition techniques only locally in each cluster. But as embedding can be regarded as a non-linear signal manipulation [6], hence a non-linear technique like KPCA should be appropriate. In order to reduce the complexity we present a variant of kernel-PCA whose parameters are computed

1-4244-0830-X/07/$20.00 ©2007 IEEE.

using the eigendecomposition of a low-rank approximation of the kernel matrix. This approach accomplished by splitting the data set into subsets.

The availability of digital EEG recordings allows the study of procedures trying to remove the artifact contributions from the brain signals. The primary goal will be to remove the artifacts without distorting the underlying brain signals. Most of the works (as an example see [7])present solutions while in the work we are proposing a single-channel approach. Furthermore,considering EEG signals one usually is not concerned with superimposed random noise but mainly deals with prominent artifacts like electrooculogram (EOG), head movements, eye blinks etc. Hence we will in the following consider the artifact related contributions to the recorded EEG signals "the signal" and the actual EEG signals as "sort of a broadband noise". Consequently we can use the projective subspace techniques referred to above to separate the artifacts from the "pure" EEG signals.

II. Application of Subspace Techniques

Time series analysis techniques often rely on embedding one dimensional sensor signals in the space of their time-delayed coordinates. Embedding can be regarded as a mapping that transforms a one-dimensional time series $x = (x[0], x[1], ..., x[N-1])$ to a multidimensional sequence of $K = N - M + 1$ lagged vectors

$$\mathbf{x}_k = [x[k-1+M-1], \ldots, x[k-1]]^T, \; k = 1, \ldots, K \quad (1)$$

The lagged vectors $\mathbf{X} = [\mathbf{x}_1 \cdots \mathbf{x}_K]$ lie in a space of dimension M, that constitute the columns of the trajectory matrix

$$\mathbf{X} = \begin{bmatrix} x[M-1] & x[M] & \ldots & x[N-1] \\ x[M-2] & x[M-1] & \ldots & x[N-2] \\ x[M-3] & x[M-2] & \ldots & x[N-3] \\ \vdots & \vdots & \ddots & \vdots \\ x[0] & x[1] & \ldots & x[N-M] \end{bmatrix} \quad (2)$$

Note that the matrix has identical entries along its diagonals.

The multidimensional signal is projected into the directions (eigenvectors) related with the largest eigenvalues of a covariance (or a scatter) matrix. The matrix can be computed in the input space (SSA or local SSA) or after transforming the data by a non-linear function (Kernel PCA). The reconstruction (and the reversion of the non-linearity for KPCA) using the same group of eigenvectors leads to $\hat{\mathbf{X}}$. Notice that in general elements along each descending diagonal of $\hat{\mathbf{X}}$ will not be identical like in case of the original trajectory matrix $\mathbf{X}$. This can be cured, however, by replacing the entries in each diagonal by their average, obtaining a Toeplitz matrix $\mathbf{X}_r$. This procedure assures that the Frobenius norm of the difference $(\mathbf{X}_r - \hat{\mathbf{X}})$ attains its minimum value among all possible solutions to get a matrix with all diagonals equal [1]. The one-dimensional signal, $\hat{x}[n]$, is then obtained by reverting the embedding, i.e. by forming the signal with the mean of the values along each descendent diagonal of $\hat{\mathbf{X}}$ [8]. Note that if $\hat{x}[n]$ corresponds to the extracted EOG, then the corrected EEG is computed as $y[n] = x[n] - \hat{x}[n]$.

A. Local SSA

Local SSA basically introduces a clustering step into the SSA technique [8] and operates in input space. It encompasses the following steps:

- After embedding, the column vectors $\mathbf{x}_k, k = 1...K$ of the trajectory matrix are clustered using any clustering algorithm (like k-means [9]). After clustering, the set of indices of the columns of $\mathbf{X}$ is subdivided into q disjoint subsets $c_1, c_2, ...c_q$. Thus sub-trajectory matrix $\mathbf{X}^{(c_i)}$ is formed with N_{c_i} columns of the matrix $\mathbf{X}$ which belong to the subset of indices c_i.
- A covariance matrix is computed in each cluster using zero mean data obtained via

$$\mathbf{X}_c = \mathbf{X}^{(c_i)}(\mathbf{I} - \frac{1}{N_{c_i}}\mathbf{j}_{c_i}\mathbf{j}_{c_i}^T) \quad (3)$$

  where $\mathbf{j}_{c_i} = [1, 1, \ldots, 1]^T$ is a vector with dimension $N_{c_i} \times 1$, and $\mathbf{I}$ is a $N_{c_i} \times N_{c_i}$ identity matrix.
- Next, the eigenvalue decomposition of the covariance matrix is computed, i.e.

$$\mathbf{C}^{c_i} = \langle \mathbf{X}_c \mathbf{X}_c^T \rangle = \mathbf{U}\mathbf{D}\mathbf{U}^{\mathbf{T}} \quad (4)$$

  Afterwards denoising can be achieved by projecting the multidimensional signal into the subspace spanned by the eigenvectors corresponding to the $L_{c_i} < M$ largest eigenvalues. The number of significant directions found using the maximum likelihood estimation of the parameter vector of the covariance matrix $\mathbf{C}^{(c_i)}$ of each cluster.
  This parameter vector, $\boldsymbol{\theta}$, is constituted by the most significant eigenvalues and the corresponding eigenvectors and the variance of the noise which is estimated by the average of the discarded eigenvalues. The number of directions $k = L_{c_i}$ will be the value that minimizes the following expression [10].

$$MDL(k) = -\ln f(\mathbf{X}^{c_i}|\widehat{\boldsymbol{\theta}}) + \frac{1}{2}P \ln N, \quad k = 0, \ldots, M-1 \quad (5)$$

  where $N = N_{c_i}$ is the number of observations available to estimate the covariance matrix and $f(\mathbf{X}^{c_i}|\widehat{\boldsymbol{\theta}})$ denotes the conditional probability density parameterized by $\widehat{\boldsymbol{\theta}}$. This log likelihood function $L(\widehat{\boldsymbol{\theta}}) = \ln f(\mathbf{X}^{c_i}|\widehat{\boldsymbol{\theta}})$ represents the accuracy of the representation of the data with the parameter vector and depends on the discarded eigenvalues

$$L(\widehat{\boldsymbol{\theta}}) = N(M-k) \ln \left[\frac{\prod_{i=k+1}^{M} \lambda_i^{1/(M-k)}}{\frac{1}{M-k}\sum_{i=k+1}^{M} \lambda_i} \right] \quad (6)$$

The negative log-likelihood $-L(\widehat{\boldsymbol{\theta}})$ is recognized to be a standard measure of the training error. However it has been reported that the simple maximization of this term tends to result in the phenomenon of over-fitting. Thus the second term in eqn. (5) was added as a regularization term to penalize complexity. The value of P is related with the number of parameters in $\boldsymbol{\theta}$ and the complexity of its estimation. Considering real valued signals, the value of P is computed according to

$$P = k+1+Mk-k^2/2-k/2 = -k^2/2+k(M+1/2)+1 \tag{7}$$

- The eigenvectors related with the largest are used in the reconstruction process. Considering the matrix $\mathbf{U}$ with L_{c_i} eigenvectors in its columns, the reconstructed vectors in each cluster are obtained as

$$\hat{\mathbf{X}}^{\mathbf{c_i}} = \mathbf{U}\mathbf{U}^T\mathbf{X}^{c_i} + \frac{1}{N_{c_i}}\mathbf{X}^{c_i}\mathbf{j}_{c_i}\mathbf{j}_{c_i}^T \tag{8}$$

- The clustering is reverted by forming an estimate $\hat{\mathbf{X}}$ of the reconstructed, noise-free trajectory matrix using the columns of the extracted sub-trajectory matrices, $\hat{\mathbf{X}}^{c_i}$, $i = 1, ..., q$ according to the contents of subsets c_i.

It has to be noticed that the SSA is obtained by skipping the clustering step, i.e , $q = 1$. The figure 1 illustrates the application of the method to decompose a noisy sinusoid (fig. 1-b) into two components. The sinusoid, embedded with $M = 2$, has an elliptic trajectory in 2-D space as shown in fig. 1-a). Then using local SSA with $q = 10$ clusters and projecting (and reconstructing) using the direction related with the largest eigenvector the 2-D trajectory of $\hat{\mathbf{X}}$ is a piece-wise approximation of the original (fig. 1-a). It has to be noticed that the projective subspace denoising (SSA) will result in a straight line corresponding to the direction of maximum variance of the data. Further note that unlike linear PCA, kernel-PCA allows to extract a number of principal components that exceeds the dimensionality of the input data. Notice that having $K \geqslant M$ examples of data with dimension M, working in input space, the maximum number of nonzero eigenvalues will also be M as can be seen by either computing the covariance matrix or the matrix of dot products. In kernel-PCA instead, the kernel matrix in feature space will have size $K \times K$ and the number of nonzero eigenvalues can often be higher than M.

B. KPCA using greedy approach

Kernel Principal Component Analysis (KPCA) relies on a non-linear mapping of given data to a higher dimensional space, called feature space. Without loosing generality, let's assume that the data set is centered and split into two parts yielding the mapped data set

$$\begin{aligned} \boldsymbol{\Phi} &= [\phi(\mathbf{x}_1)\phi(\mathbf{x}_2)\dots\phi(\mathbf{x}_r),\phi(\mathbf{x}_{r+1})\dots\phi(\mathbf{x}_K)] \\ &= [\ \boldsymbol{\Phi}_R \quad \boldsymbol{\Phi}_S] \end{aligned} \tag{9}$$

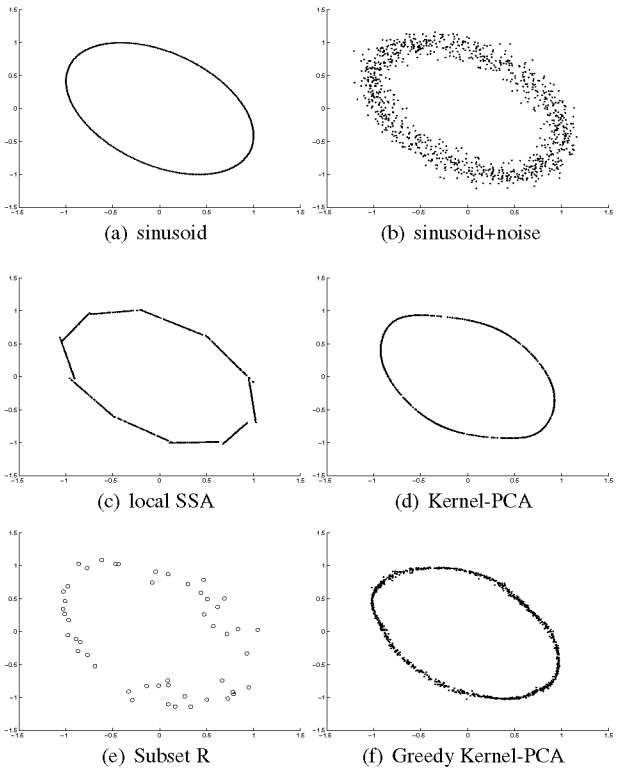


Fig. 1. SIGNALS EMBEDDED IN TIME-DELAYED COORDINATES $M = 2$

In denoising applications, the first step of KPCA is to compute the projections of a mapped data set onto a feature subspace. Considering L eigenvectors (columns of $\mathbf{U}$) of a covariance matrix (a correlation matrix if the data is centered) corresponding to the L largest eigenvalues, the projections of the mapped data set $\boldsymbol{\Phi}$ are

$$\mathbf{Z} = \mathbf{U}^T\boldsymbol{\Phi} \tag{10}$$

The columns of the matrix $\mathbf{U}$ form the basis in feature space onto which to project the data set. This basis can be written as a linear combination of the mapped data

$$\mathbf{U} = \boldsymbol{\Phi}_B\mathbf{A} \tag{11}$$

The matrix $\mathbf{A}$ is a matrix of coefficients and either $\boldsymbol{\Phi}_B = \boldsymbol{\Phi}$ (KPCA) or $\boldsymbol{\Phi}_B = \boldsymbol{\Phi}_R$ (greedy KPCA), representing a subset of the data set only. Note that the column j of $\mathbf{Z}$ depends on the dot products $\boldsymbol{\Phi}_B^T\phi(\mathbf{x}_j)$. However to avoid an explicit mapping into feature space, all data manipulations are achieved by dot products [5] and the kernel trick is applied. For instance, using an RBF kernel, the dot product between a vector i, belonging to B subset, and $\phi(\mathbf{x}_j)$ is computed using a kernel function that only depends on the input data

$$k(\mathbf{x}_i, \mathbf{x}_j) = exp(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}) \tag{12}$$

Finally, to recover the noise-reduced signal after denoising in feature space, the non-linear mapping must be reverted, i.e.

the pre-image in input space of every signal, denoised and reconstructed in feature space, must be estimated. Denoising using KPCA thus comprises two steps after the computation of the projections in the feature space: a) the reconstruction in feature space and b) the estimation of the pre-image of the reconstructed point $\hat{\phi}(\mathbf{x}_j) = \mathbf{U}\mathbf{z}_j$, where $\mathbf{z}_j$ represents the projections of a noisy point $\mathbf{x}_j$. These two steps can be joined together by minimizing the Euclidian distance of the image $\phi(\mathbf{p})$ of a yet unknown point $\mathbf{p}$ from $\hat{\phi}(\mathbf{x}_j)$

$$\begin{aligned} \tilde{d}^{(2)} &= \|\phi(\mathbf{p}) - \hat{\phi}(\mathbf{x}_j)\|^2 \\ &= (\phi(\mathbf{p}) - \hat{\phi}(\mathbf{x}_j))^T(\phi(\mathbf{p}) - \hat{\phi}(\mathbf{x}_j)) \end{aligned} \quad (13)$$

The central idea of the fixed-point method [5]consists in computing the unknown pre-image of a reconstructed point in the projected feature subspace by finding a $\mathbf{p}$ which minimizes that distance (eq. 13). If an RBF kernel is considered, the iterative procedure is described by the following equation

$$\mathbf{p}_{t+1} = \frac{\mathbf{X}_B(\mathbf{g} \diamondsuit \mathbf{k}_{p_t})}{\mathbf{g}^T \mathbf{k}_{\mathbf{p_t}}} \quad (14)$$

where $\diamondsuit$ represents a Hadamard product, $\mathbf{g} = \mathbf{A}\mathbf{z}_j$. The components of the vector $\mathbf{k}_{\mathbf{p_t}} = \mathbf{k}(\mathbf{X}_B, \mathbf{p}_t)$ are given by the dot products between $\phi(\mathbf{p}_t)$ and the images $\mathbf{\Phi}_B$ of the training subset $\mathbf{X}_B$. The algorithm must be initialized and $\mathbf{p}_0 \equiv \mathbf{x}_i$ is a valid choice [6]. The points $\mathbf{p}_k$ then form the columns of $\hat{\mathbf{X}}$, the noise-free multidimensional signal in input space. The application of the method is illustrated in fig. 1-d) and f) where we can see that the denoised trajectory is smoother than the one obtained with local SSA. In the feature space the data was projected (and reconstructed) using $L = 4$ or $L = 7$ directions, respectively using kernel-PCA or greedy kernel-PCA.

Computing the Basis

The two approaches, KPCA and greedy KPCA, respectively, arise from two distinct strategies to deal with the eigendecomposition of the kernel matrix ($\mathbf{K}$) of the data set. In KPCA the matrix $\mathbf{A}$ is computed using the largest eigenvalues ($\mathbf{D}$) and corresponding eigenvectors ($\mathbf{V}$) of $\mathbf{K}$ [6]. Then the basis vector is

$$\mathbf{U} = \mathbf{\Phi}\mathbf{V}\mathbf{D}^{-1/2} \quad (15)$$

In greedy KPCA a low-rank approximation of the kernel matrix is considered. This leads to the eigendecomposition of matrices with reduced size. Considering the kernel matrix, written in block notation [11],[12]

$$\mathbf{K} = \begin{bmatrix} \mathbf{K}_r & \mathbf{K}_{rs} \\ \mathbf{K}_{rs}^T & \mathbf{K}_s \end{bmatrix} \quad (16)$$

This manipulation takes into account that the original training set is divided into two subsets. Where the $\mathbf{K}_r$ is the kernel matrix within subset $\mathbf{\Phi}_R$, $\mathbf{K}_{rs}$ is the kernel matrix between subset $\mathbf{\Phi}_R$ and $\mathbf{\Phi}_S$ and $\mathbf{K}_s$ is the kernel matrix within the subset $\mathbf{\Phi}_S$. The approximation is written using the upper blocks of the original matrix [11], [12]

$$\tilde{\mathbf{K}} = \begin{bmatrix} \mathbf{K}_r \\ \mathbf{K}_{rs}^T \end{bmatrix} \mathbf{K}_r^{-1} \begin{bmatrix} \mathbf{K}_r & \mathbf{K}_{rs} \end{bmatrix} \quad (17)$$

It can be verified that the lower block is approximated by $\mathbf{K}_s \approx \mathbf{K}_{rs}^T \mathbf{K}_r^{-1} \mathbf{K}_{rs}$. There are different suggestions to perform this eigendecomposition [11],[13], one of them is based on an incomplete Cholesky decomposition $\tilde{\mathbf{K}} = \mathbf{C}^T\mathbf{C}$, with

$$\mathbf{C} = \begin{bmatrix} \mathbf{L} & \mathbf{L}^{-T}\mathbf{K}_{rs} \end{bmatrix} \quad (18)$$

where $\mathbf{L}$ is a triangular matrix corresponding to the decomposition of $\mathbf{K}_r = \mathbf{L}^T\mathbf{L}$. The $R \times R$ matrix $\mathbf{Q} = \mathbf{C}\mathbf{C}^T$ and its eigendecomposition $\mathbf{V}_q\mathbf{D}\mathbf{V}_q^T$ in conjunction with $\mathbf{L}$ is used to form the basis

$$\mathbf{U} = \mathbf{\Phi}_R\mathbf{L}^{-1}\mathbf{V}_q \quad (19)$$

Note that the pivoting index within the incomplete Cholesky decomposition [13] leads to the selection of $\mathbf{\Phi}_R$ within the training set.

Choosing training and testing subsets

In the last section it was discussed that with an incomplete Cholesky decomposition it is possible to compute the parameters of the model using a training data set divided into two subsets. But the parameters depend on both subsets. However, there are approaches, where the parameters depend on only one of the subsets [11] whose elements are chosen randomly. In this application, we consider an hybrid approach which leads to the choice of three subsets of data. We start by splitting the data (with J vectors) into two data sets: the training set with K vectors and the testing set which contains the remaining data to be processed. In this application, we consider to form the training set with two strategies

- choosing the K vectors randomly.
- choosing the K vectors that correspond to a subsegment (with artifact) of the segment to be processed.

The subset R of the training set is chosen using Cholesky decomposition performed using the symmetric pivoting algorithm [13]. The methodology is based on the minimization of the trace $tr(\mathbf{K}_s - \mathbf{K}_{rs}^T\mathbf{K}_r^{-1}\mathbf{K}_{rs})$ applied to iteratively update the subset R. So by identifying the maximum value of the trace operator (the pivot), an element of subset S is moved to the subset R and the matrix $\mathbf{K}_r$ increases its size while the others decrease. The process stops when the trace of the matrix corresponding to the actual approximation is less than a threshold [13] and/or the matrix $\mathbf{K}_r$ is not well conditioned [14] or up to a maximum number of pivots[15].

III. Results

The two suggested projective techniques will be illustrated using a frontal channel (Fp1-CZ) contaminated with a very prominent electrooculogram (EOG) artifact. The segment, with

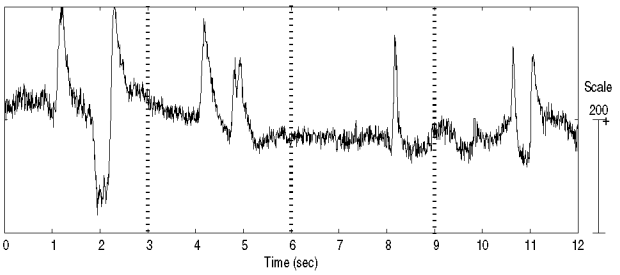
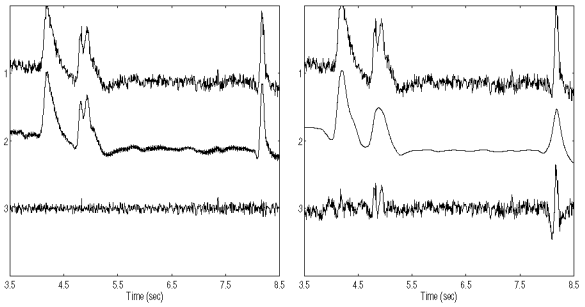


Fig. 2. FP1 CHANNEL (WITH REFERENCE TO CZ): ALL SUBSEGMENTS CONTAIN AN EOG ARTIFACT



(a) L_{c_i} selected by MDL (b) $L_{c_i} = 1$

Fig. 3. EXTRACTED ARTEFACT AND CORRECTED EEG SIGNAL OBTAINED WITH LOCAL SSA USING 6 CLUSTERS: *top*: ORIGINAL EEG, *middle*: EXTRACTED EOG AND *bottom*: CORRECTED EEG.

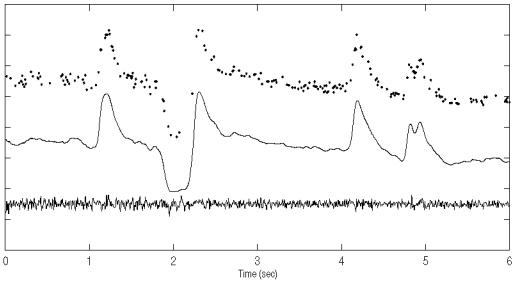


Fig. 4. ILLUSTRATION OF GREEDY KPCA: THE SIGNALS IN THE DIFFERENT STEPS ARE: $1strow$ - FIRST ROW OF RANDOM SELECTED POINTS, $2ndrow$ - EXTRACTED EOG SIGNAL AND $3throw$ - CORRECTED EEG

$12s$, of the signal containing high-amplitude EOG artifacts is shown in fig. 2. The embedding dimension M was experimentally chosen and for local SSA there is a minimum value related with the frequency of the signal to be extracted [1], [6]. But with KPCA the embedding value can be smaller but we can deduce any proper rule.

A. Local SSA

Local SSA was applied following the steps described before assigning $M = 41$ and varying the number of clusters. A visual inspection of the extracted EOG using the algorithm and the corrected EEG (the difference between original EEG and the extracted EOG) by a clinical expert revealed no significant difference between the different versions computed for different numbers of clusters. It was also observed that the $50Hz$ interference is consistently removed together with the EOG signal in all cases. However quantifying the frequency contents in different bands in the corrected EEG component the distortion in theta band $(4-7Hz)$ is less significant when the q has higher values. However in order to have a correct estimate of the covariance matrix in each cluster we should have enough data in each cluster. This constitute a practical bound to the number of clusters. In each cluster the MDL criterion was used to select the subspace dimension or using only one direction corresponding to the largest eigenvalue to reconstruct the EOG signal. Fig.3 illustrates the difference between the two approaches. It has to be notice that with MDL the number of directions range from $5-10$ in a space of dimension $M = 41$. Using only one direction in each cluster, the extracted signal corresponds to the EOG artefact but the corrected EEG still contains some remnants of the original EOG signal. This demonstrates that with EOG artefacts the dominant principal component might be not sufficient to represent the large amplitude component of the signal. The complexity of this approach depends on the clustering step (the algorithm k-means) and the eigendecomposition of q MxM covariance matrices in each cluster.

Greedy KPCA

The multidimensional signal is obtained with an embedding of $M = 11$. The RBF kernel with $\sigma = max_i(\|\mathbf{x}_i - \mathbf{x}_{mean}\|), i = 1, \ldots, J$, where $\mathbf{x}_{mean}$ is the mean of the data set. After embedding the data set with $J = 1526$ vectors was split into testing and training sets. The training set is formed with roughly 25% of the data considering the following strategies

1. a random choice of $K = 381$ data vectors. Fig.4 (plot on top) illustrate the random choice by plotting the first row of the training data vectors according to its time reference.
2. Using all the vectors that correspond either to first subsegment of signal $(0-3s)$ or to the second subsegment $(3-6s)$.

The incomplete Cholesky decomposition algorithm is used with $R = 20$. For the three cases data was projected onto $L = 6$ basis vectors ($\mathbf{U}$) which correspond to the leveling off of the eigenspectrum of matrix $\mathbf{Q}$. The Fig. 4 shows the result of the application of the algorithm when the training set is selected randomly. The fig. 5 shows the first $6s$ of corrected EEG when the training set is formed in a subsegment of signal. And it easily visible that when the training set do not have "examples" of the features to extract it does not work well. The high-negative features present on the first subsegment are not represented in the second subsegment. Then when the training is formed with vectors belonging to 2nd segment the artifact is not removed (see fig. 5- plot on bottom). It has to be noticed that the complexity of the algorithm is depend on the size of subset R. Experimentally we conclude that a random choice of the training set followed by the Cholesky decomposition up to $R = 20$ the algorithm presents good performance. In particular for this segment the results are comparable with the application of KPCA to each subsegment

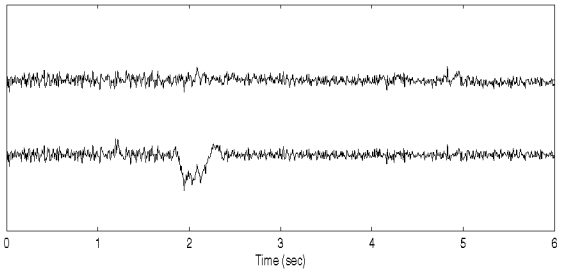


Fig. 5. CORRECTED EEG USING AS TRAINING SET: *top* FIRST SUBSEGMENT OF FIG. 2, *bottom* THE SECOND SEGMENT OF FIG. 2

(of $3s$) [16]. The correlation coefficients between the extracted signals are around 0.90.

IV. CONCLUDING REMARKS

The artifact reduction in EEG recordings [17] is a very important problem that needs to be addressed in a systematic way. We have been studying the feasibility of projective subspace techniques to address this problem using a single-channel approach. The application of projective subspace techniques to one-dimensional time series relies on an embedding step which introduces a nonlinearity in the processing chain. This fact leads to the proposal of local SSA, which is the application of SSA to the clusters formed with the multidimensional signals resulting from the embedding step. This piecewise linear approximation is then compared to a generically non-linear subspace projection technique like KPCA. The toy example illustrates that noise reduction of multidimensional signals cannot be achieved using plain SSA. Rather local SSA is needed which results in a piecewise linear approximation of the original trajectory matrix of the data. While with the kernel-PCA the denoised trajectory is smoother than with local SSA. In what concerns the KPCA the greedy implementation using the strategy of splitting the data set decreases the complexity by avoiding the eigendecomposition of large kernel matrices ($\mathbf{K}$). The size of $\mathbf{K}$ depends on the amount of data. Considering the EEG denosing application, we are interested on segments that correspond to the normal window size (around $10s$) in a visualization task. Both algorithms (local SSA and greedy KPCA) are being incorporated in the EEGLAB [18] environment. This open-software tool based on MATLAB offers visualization facilities that will allow to accomplish clinical evaluation task. Our goal is to use this facility in the visualization of critical segments of signals from a database of epileptic patients recorded in long-term monitoring sessions and study the impact of the application of the algorithms. In this scenario the algorithm can be applied in parallel to the channels that suffer from high-amplitude artifacts. This could be useful to detect the onset of a focal seizure. Furthermore, the proposed methods need to be evaluated in a more quantitative and systematic approach, concerning for instance the spectral distortion in the important frequency ranges of EEG. Some preliminary results [6] show that KPCA distorts less the low-frequency contents of the signals.

ACKNOWLEDGMENT

A.R. Teixeira received a PhD Scholarship (SFRH/BD/28404/2006) supported by the Portuguese Foundation for Science and Technology (FCT). This work was also supported by grants from DAAD and CRUP which are gratefully acknowledged. We are grateful to A. Martins da Silva and Hospital Geral de Santo António for providing the signals and the helpful discussions.

REFERENCES

[1] Nina Golyandina, Vladimir Nekrutkin, and Anatoly Zhigljavsky, *Analysis of Time Series Structure: SSA and Related Techniques*, Chapman & HALL/CRC, 2001.

[2] Yariv Ephraim and Harry L. Van Trees, "A signal subspace approach for speech enhancement," *IEEE Transactions on Acoustic, Speech and Signal Processing.*, vol. 3, no. 4, pp. 251–266, 1995.

[3] Chang Huai You, Soo Ngee Koh, and Susanto Rahardja, "Signal subspace speech enhancement for audible noise reduction," in *ICASPP 2005*, Philadelphia, USA, 2005, vol. I, pp. 145–148, IEEE.

[4] M. Ghil, M.R. Allen, M. D. Dettinger, K. Ide, and et al, "Advanced spectral methods for climatic time series," *Reviews of Geophysics*, vol. 40, no. 1, pp. 3.1–3.41, 2002.

[5] Klaus-Robert Müller, Sebastian Mika, Gunnar Rätsch, Koji Tsuda, and Bernhard Schölkopf, "An introduction to kernel-based algorithms," *IEEE Transactions on Neural Networks*, vol. 12, no. 2, pp. 181–202, 2001.

[6] A. R. Teixeira, A. M. Tomé, E. W. Lang, and K.Stadlthanner, "Nonlinear projective techniques to extract artifacts in biomedical signals," in *EUSIPCO2006*, Florence, Italy, 2006.

[7] Tzyy-Ping Jung, Scott Makeig, Colin Humphries, Te-Won Lee, Martin J. Mckeown, Vicente Iragui, and Terrence j. Sejnowski, "Removing electroencephalographic artifacts by blind source separation," *Psychophysiology*, vol. 37, pp. 163–178, 2000.

[8] A. R. Teixeira, A. M. Tomé, E.W.Lang, P. Gruber, and A. Martins da Silva, "On the use of clustering and local singular spectrum analysis to remove ocular artifacts from electroencephalograms," in *IJCNN2005*, Montréal, Canada, 2005, pp. 2514–2519, IEEE.

[9] Christopher M. Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press, Oxford, 1995.

[10] Athanasios P. Liavas and Philip A. Regalia, "On the behavior of information theoretic criteria for model order selection," *IEEE Transactions on Signal Processing*, vol. 49, no. 8, pp. 1689–1695, 2001.

[11] Christopher K.I. Williams and Mathias Seeger, "Using the nyström method to speed up kernel machines," in *Advances in Neural Information Processing Systems*. 2000, pp. 682–688, MIT Press.

[12] Charless Fowlkes, Sergie Belongie, Fan Chung, and Jitendra Malik, "Spectral grouping using the nyström method," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 2, pp. 214–225, 2004.

[13] Francis R. Bach and Michael I. Jordan, "Kernel independent component analysis," *Journal of Machine Learning Research*, vol. 3, pp. 1–48, 2002.

[14] Gavin C. Cawley and Nicola L. C. Talbot, "Efficient formation of a basis in a kernel induced feature space," in *European Symposium on Artificial Neural Networks*, Michel Verleysen, Ed., Bruges, Belgium, 2002, pp. 1–6, d-side.

[15] Vojtěch Franc and Václav Hlaváč, "Stastical pattern recognition toolbox for MATLAB," 2004.

[16] A. R. Teixeira, A.M.Tomé, E.W.Lang, R. Schachtner, and K.Stadlthanner, "On the use of KPCA to extract artifacts in one-dimensional biomedical signals," in *Machine Learning for Signal Porcessing, MLSP 2006*, Seán McLoone, Jan Larsen, Marc Van Hulle, Alan Rogers, and Scott C. Douglas, Eds., Dublin, 2006, pp. 385–390, IEEE.

[17] R. J. Croft and R. J. Barry, "Removal of ocular artifact from the EEG: a review," *Neurophysiol. Clin.*, vol. 30, pp. 5–19, 2000.

[18] A. Delorme and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics," *Journal of Neuroscience Methods*, vol. 134, pp. 9–21, 2004.