NONLINEAR PROJECTIVE TECHNIQUES TO EXTRACT ARTIFACTS IN BIOMEDICAL SIGNALS

[1]*A. R.Teixeira* , [1]*A. M. Tomé* , [2]*K. Stadlthanner* , [2]*E. W. Lang*

[1] DETI/IEETA, Universidade de Aveiro, 3810-193 Aveiro, Portugal
email: ana@ieeta.pt
[2] Institute of Biophysics, University of Regensburg, D-93040 Regensburg, Germany
email: elmar.lang@biologie.uni-regensburg.de

ABSTRACT

Biomedical signals are generally contaminated with artifacts and noise. In case the artifacts dominate, the useful signal can easily be extracted with projective subspace techniques. Then, biomedical signals which often represent one dimensional time series, need to be transformed to multidimensional signal vectors for the latter techniques to be applicable. The transformation can be achieved by embedding an observed signal in its delayed coordinates. Using this embedding we propose to cluster the resulting feature vectors and apply a singular spectrum analysis (SSA) locally in each cluster to recover the undistorted signals. We also compare the reconstructed signals to results obtained with kernel-PCA. Both nonlinear subspace projection techniques are applied to artificial data to demonstrate the suppression of random noise signals as well as to an electroencephalogram (EEG) signal recorded in the frontal channel to extract its prominent electrooculogram (EOG) interference.

1. INTRODUCTION

In many biomedical signal processing applications a sensor signal is contaminated with artifactual 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 artifactual signal 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 [7]. 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 [5], [2], [12]. 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. Considering EEG signals one usually is not concerned with superimposed random noise but mainly deals with prominent artifacts like electrooculogram (EOG) and electrocardiogram (ECG) interferences, head movements, eye blinks etc. Hence we will in the following consider the artifactual 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.

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) 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.

In this work we will introduce 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. We show that both non-linear techniques, local SSA or kernel-PCA [9], turn out to be more efficient than simple SSA. Both methods will be presented following a matrix manipulation approach which is particularly suitable for dealing with the pre-image problem in kernel-PCA. Some toy data will illustrate the application of the methods to suppress noise, while later on these methods are used to extract EOG artifacts from frontal EEG recordings.

2. METHODS

2.1 Embedding

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

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

with $M < N$ being the corresponding window length or the embedding dimension. The lagged vectors then constitute the columns of the trajectory matrix $\mathbf{X} = [\mathbf{x}_1 \cdots \mathbf{x}_K]$ which represents a Toeplitz matrix. The further processing of this data matrix is performed by either applying local SSA or kernel PCA, two nonlinear projective techniques operating in a corresponding feature space.

2.2 Local SSA

Local SSA basically introduces a clustering step into the SSA technique [11] and operates in feature 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 [1]). 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

$$\mathbf{X}_c = \mathbf{X}^{(c_i)}(\mathbf{I} - \frac{1}{N_{c_i}}\mathbf{j}_{c_i}\mathbf{j}_{c_i}^T) \tag{2}$$

  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.

$$\mathbf{C} = \langle \mathbf{X}_c \mathbf{X}_c^T \rangle = \mathbf{U}\mathbf{D}\mathbf{U}^\mathbf{T} \tag{3}$$

  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 latter eigenvectors are also 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

$$\hat{\mathbf{X}}^{(\mathbf{c_i})} = \mathbf{U}\mathbf{U}^T\mathbf{X}^{(c_i)} + \frac{1}{N_{c_i}}\mathbf{X}^{(c_i)}\mathbf{j}_{c_i}\mathbf{j}_{c_i}^T \tag{4}$$

- 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.
- 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 by replacing the entries in each diagonal by their average. This procedure assures that the Frobenius norm of the difference between the original matrix and the reconstructed matrix has minimum value among all possible solutions to get a matrix with all diagonals equal.
- The noise-reduced one-dimensional signal, $\hat{x}[n]$, is then obtained by reverting the embedding.

2.3 Kernel-PCA

2.3.1 Covariances and kernels

In kernel-PCA (KPCA) a multidimensional signal $\mathbf{x}_k$, $k = 1...K$, is considered to be mapped through a non-linear function $\phi(\mathbf{x}_k)$ into a feature space yielding the mapped data set $\mathbf{\Phi} = [\phi(\mathbf{x}_1)\phi(\mathbf{x}_2)...\phi(\mathbf{x}_K)]$. In feature space then a linear PCA is performed estimating the eigenvectors and eigenvalues of a matrix of *outer* products, called a *covariance matrix* or, for zero mean data, the non-normalized *correlation matrix* $\mathbf{C} = \mathbf{\Phi}\mathbf{\Phi}^T$. It can be shown that these eigenvectors and -values are related to those of a matrix of *inner* products, called a *kernel matrix* $\mathbf{K} = \mathbf{\Phi}^T\mathbf{\Phi}$. Using the kernel trick [9], this kernel matrix can be computed without explicitly mapping the data. The centered kernel matrix $\mathbf{K}_c$ can be computed as follows

$$\begin{aligned} \mathbf{K}_c &= (\mathbf{I} - \frac{1}{K}\mathbf{j}_K\mathbf{j}_K^T)\mathbf{\Phi}^T\mathbf{\Phi}(\mathbf{I} - \frac{1}{K}\mathbf{j}_K\mathbf{j}_K^T) \\ &= (\mathbf{I} - \frac{1}{K}\mathbf{j}_K\mathbf{j}_K^T)\mathbf{K}(\mathbf{I} - \frac{1}{K}\mathbf{j}_K\mathbf{j}_K^T) \end{aligned} \tag{5}$$

Notice that each element $k(i, j)$ of the kernel matrices depends on the inner product $\phi^T(\mathbf{x}_i)\phi(\mathbf{x}_j)$ which can be computed using only the data $\mathbf{x}_k$ in input space. For instance if an RBF kernel is used, $k(i, j)$ is calculated according to

$$k(i, j) = exp(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}) \tag{6}$$

where σ^2 is a parameter that is related to the variance of the data.

Because the eigenvalues of the non-normalized covariance matrix coincide with the eigenvalues of the kernel matrix, their eigenvectors are related by

$$\mathbf{U} = \mathbf{\Phi}(\mathbf{I} - \frac{1}{K}\mathbf{j}_K\mathbf{j}_K^T)\mathbf{V}\mathbf{D}^{-1/2} \tag{7}$$

Where $\mathbf{U}$ is the matrix with L eigenvectors of the covariance matrix, $\mathbf{V}$ is the matrix with L eigenvectors of kernel matrix and $\mathbf{D}$ is a diagonal matrix with the corresponding $L \leq K$ largest eigenvalues of both matrices. Once a data point of the input space $\mathbf{y}_j$ is mapped to the feature space to obtain its image $\phi(\mathbf{y}_j)$, the latter can be projected onto the L eigenvectors spanning the feature space to obtain

$$\mathbf{z}_j = \mathbf{D}^{-1/2}\mathbf{V}^T(\mathbf{I} - \frac{1}{K}\mathbf{j}_K\mathbf{j}_K^T)\mathbf{k}_{y_j} \tag{8}$$

where $\mathbf{k}_{y_j} = [k(\mathbf{x}_1, \mathbf{y}_j), k(\mathbf{x}_2, \mathbf{y}_j), \ldots, k(\mathbf{x}_K, \mathbf{y}_j)]^T$ is the vector of inner products between the training data $\mathbf{\Phi}$ and $\phi(\mathbf{y}_j)$, naturally computed using the kernel trick. There are many applications where the projections are the only information that is needed. However, in denoising applications it is needed to reconstruct any point in feature space using the L eigenvectors and then estimate the position of the corresponding point in the input space, i.e. compute its pre-image [8, 10]. In this application, the multidimensional signal is denoised and the one-dimensional signal can be obtained by reverting the embedding as described in the two last steps of local SSA.

2.3.2 Reconstruction and Pre-Image

The reconstruction of the point in the feature space is combined with the estimate of the pre-image in the input space. Combining these two steps the kernel trick can be used and the reconstruction never needs to be computed explicitly [8] avoiding to work in the feature space. The most recent work [8] to find the pre-image of a point of the feature space is based on the fact that it is possible to compute the coordinates of a new point if we know its distance to a set of known points [6]. The proposed method consists of the following steps:

- The vector of squared distances $\tilde{\mathbf{d}}^{(2)}$ between the reconstructed images $\hat{\phi}(\mathbf{y}_j)$ and the training data set $\mathbf{\Phi}$ are computed. Considering the reconstructed point $\hat{\phi}(\mathbf{y}_j) = \mathbf{U}\mathbf{z}_j$ and substituting eqn. 7, we can re-write $\hat{\phi}(\mathbf{y}_j) = \mathbf{\Phi}\mathbf{g}$, where $\mathbf{g}$ is a $K \times 1$ vector. Then using the kernel trick we obtain the row vector of distances

$$\tilde{\mathbf{d}}^{(2)} = \mathbf{g}^T\mathbf{K}\mathbf{g} + diag(\mathbf{K}) - 2\mathbf{g}^T\mathbf{K} \tag{9}$$

- If an RBF is considered, there is a relation between the input space distance $\mathbf{d}^{(2)}$ and the corresponding feature space distance. Once the distance in feature space can also be computed as

$$\tilde{\mathbf{d}}^{(2)} = 2(\mathbf{j}_K - exp(-\frac{\mathbf{d}^{(2)}}{2\sigma^2})) \tag{10}$$

  the vector of distances in input space is then given by

$$\mathbf{d}^{(2)} = -2\sigma^2 \ln(\mathbf{j}_K - 0.5\tilde{\mathbf{d}}^{(2)}) \tag{11}$$

- Considering a subset of neighbors of the reconstructed point $\hat{\phi}(\mathbf{y}_j)$ (i.e choosing the S points with smallest distance $\tilde{\mathbf{d}}^{(2)}$), and given the corresponding points $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, ...\mathbf{p}_S]$ in input space, the coordinate system for the subset may be defined as the columns of the eigenvector matrix $\mathbf{E}$ of their covariance matrix. After centering

$$\mathbf{P}_c = \mathbf{P}(\mathbf{I} - \frac{1}{S}\mathbf{j}_S^T\mathbf{j}_S) \tag{12}$$

  the standard eigendecomposition is given by $\langle \mathbf{P}_c\mathbf{P}_c^T \rangle = \mathbf{E}\mathbf{D}\mathbf{E}^T$.
- Then $\mathbf{W} = \mathbf{E}^T\mathbf{P}_c$ represent the new coordinates of the points $\mathbf{P}_c$. Their distance to the origin is obtained as $\mathbf{d}_0^{(2)} = [\| \mathbf{w}_1 \|^2, \| \mathbf{w}_2 \|^2 ... \| \mathbf{w}_S \|^2]$. The new coordinates of the projected point in the input space are then given by

$$\tilde{\mathbf{c}} = (-1/2)(\mathbf{W}\mathbf{W}^T)^{-1}\mathbf{W}(\mathbf{d}^{(2)} - \mathbf{d}_0^{(2)}) \tag{13}$$

  The pre-image $\mathbf{c}$ of $\hat{\phi}(\mathbf{y}_j)$ is finally obtained as

$$\mathbf{c} = \mathbf{E}\tilde{\mathbf{c}} + \frac{1}{S}\mathbf{P}\mathbf{j}_S \tag{14}$$

  Note that not every point in feature space may have a corresponding pre-image in input space [9].

3. RESULTS AND DISCUSSION

The methods were implemented in MATLAB using the toolbox provided by Franc [3], where basic pattern recognition tools and kernel methods can be found. In the following the methods discussed above will be applied to toy examples as well as EEG signals.

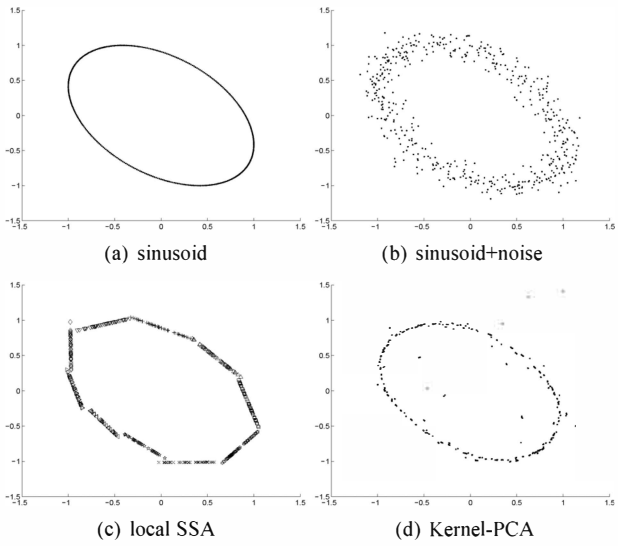


Figure 1: Signals embedded in time-delayed coordinates $M = 2$

3.1 Toy examples

The first toy example to be discussed comprises an artificially generated sinusoid, the other uses a chaotic time series given by Hénon's equation. To both time series $x[n]$ Gaussian white noise will be added $\tilde{x}[n] = x[n] + r[n]$ corresponding to a signal-to-noise ratio $SNR = 20dB$. Afterwards the data is embedded in delayed coordinates using $M = 2$.

3.1.1 Sinusoidal time series

Figure 1 shows the 2-D representation of the corresponding vector signals. To achieve denoising, at first local SSA is applied. The embedded data are grouped into $q = 10$ clusters using *k-means* clustering. In each cluster the covariance matrix of the data is calculated. As its size is only 2×2, we project the data on the eigenvector which corresponds to the largest eigenvalue. The other direction is considered to be related with the noise. Figure 1 shows that in each cluster the reconstructed data, following the direction of maximum variance locally, represents a good approximation to the underlying trajectory in phase space.

KPCA was implemented using an RBF kernel with width parameter $\sigma^2 = 1$ and $L = 3$ principal components have been used for the PCA in feature space. To estimate the pre-images of the data using the distance method referred to above, $S = 10$ nearest neighbors were considered. Figure 1 shows that KPCA results in a smoother approximation to the underlying phase space trajectory. But some outliers remain which probably result from bad estimates of the corresponding pre-images.

3.1.2 *Hénon time series*

The second example considers a nonlinear time series resulting from the following dynamics $x[n+1] = 1 - a \cdot x^2[n] + b \cdot x[n-1]$ where $a = 1.4, b = 0.3$ has been used. The resulting data was embedded in delayed coordinates using $M = 2$. Figure 2-a) shows the phase space trajectories generated by this dynamical model known as the Hénon map. Figure 2-b) shows the corresponding data set with added Gausssian noise. Applying local SSA with $q = 15$ clusters results in the "denoised" trajectories shown in Figure 2-c). It can be seen that the local approximations to the underlying dynamics reflect the general trend of the data very well. But it is also obvious that the mapping is not always smooth. This results from the structure of the local clusters which possess principal directions deviating from the underlying dynamics due to noise. Also it can be seen that the fine structure of the Hénon map cannot be captured where the spacing of segments of the trajectory are too closely spaced compared to the spread of the noise. The latter also holds in case of kernel-PCA using an RBF kernel with a width parameter $\sigma = 1$ and $L = 4$ principal components in feature space. The pre-images of the data reconstructed in feature space have been estimated using $S = 10$. Though the resulting trajectories are much smoother than in case of local SSA they also are much more noisy still.

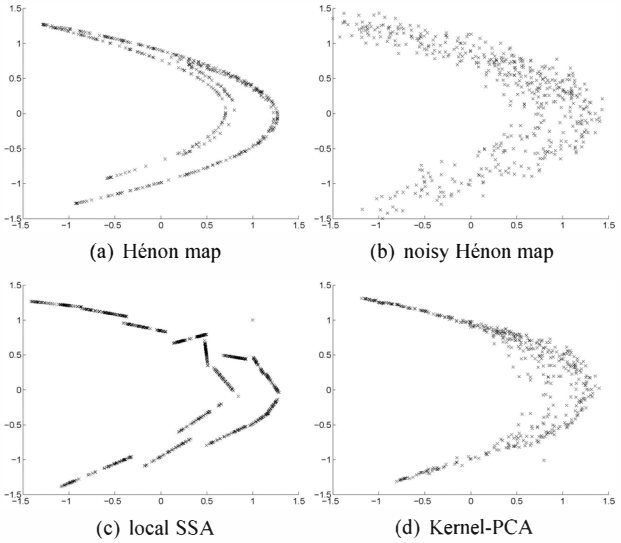


Figure 2: Signals embedded in time-delayed coordinates $M = 2$

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 $K \geq M$ 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.

3.2 EEG data

Biomedical signals are often contaminated with artifactual signals which severely distort the signals to be investigated. As an example we will study the removal of the prominent EOG artefact from EEG recordings. Hence all remaining signals in the EEG recording will be considered "noise" for the sake of the argument. Figure 3 gives an illustrative example of the results obtained. A local SSA analysis has been performed using $N = 1536$ samples recorded with a sampling rate of $128Hz$. The data have been embedded in delayed coordinates using $M = 41$. The resulting columns of the trajectory matrix have been clustered choosing 6 clusters. The dimension of the signal subspace in each cluster has been estimated applying an MDL (minimum description length) criterion. The dimension of the signal subspace is different in each cluster and takes a value in the range $[5-10]$ [11]. Figure 4-a) shows the power spectral densities (psd) of the recorded EEG signal, the extracted EOG signal and the residual signal (corrected EEG). It can be seen that local SSA allows to remove both the EOG artefact as well as the 50 Hz line noise without distorting the remaining psd. But it seems to suppress the psd in the low frequency band too strongly, hence may remove other low frequency components as well.

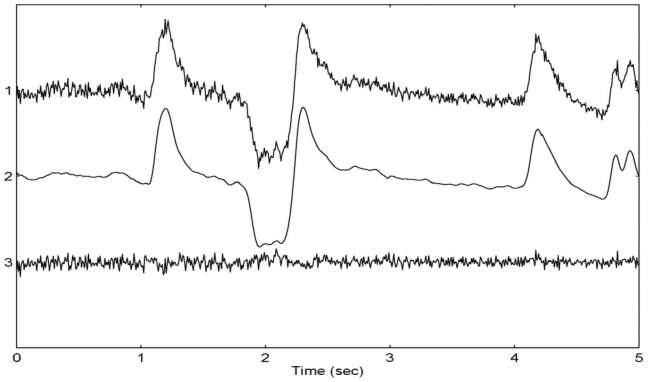
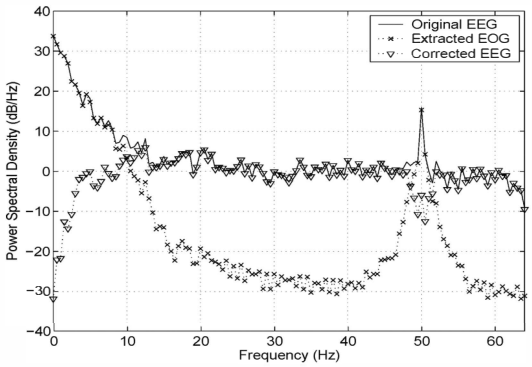
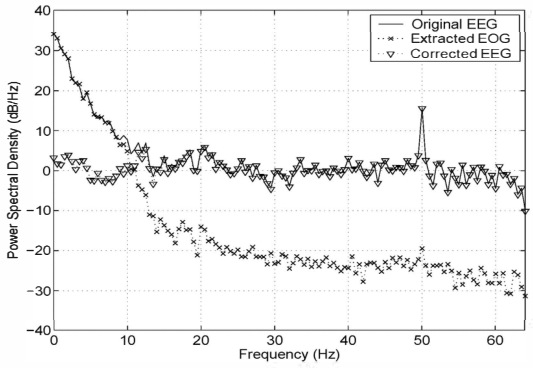


Figure 3: A segment of *top*: a recorded EEG signal, *middle*: an EOG signal extracted with kernel-PCA and *bottom*: the residual signal (corrected EEG)

kernel-PCA could not be applied to the whole segment because of a prohibitive computational load, hence only subsegments with $N = 384$ samples could be analyzed. With kernel-PCA an embedding dimension of $M = 41$ was used again; the width parameter of the RBF kernel was chosen as a fixed percentage of the variance of the data set, i.e. $\sigma^2_{RBF} = 0.5\sigma^2_{data}$. For the reconstruction an 8-dim feature space corresponding to the eight largest eigenvalues has been chosen in accordance with the eigenvalue spectrum obtained. Figure 3 illustrates the extracted artifact as well as the corrected EEG signal within two subsegments. Figure 4-b) shows the resulting power spectral densities for the whole data segment. It can clearly be seen that kernel-PCA removes the EOG signal completely and does not suppress all low frequency contributions to the psd. However, it does not remove the 50 Hz line noise. It has to be noted that the correlation coefficients between the EOGs and the corrected EEGs extracted by both methods amount to 0.99 and 0.81, respec-



(a) local SSA



(b) Kernel-PCA

Figure 4: Power spectral densities (psd) resulting from (a)local SSA (b) kernel-PCA

tively. This demonstrates a rather good correspondence between both methods concerning the extracted artifacts. The less good agreement concerning the corrected EEGs is due to some low frequency distortions remaining after the application of local SSA. The results shown here were achieved using the same embedding dimension for both methods but with Kernel-PCA that value can be decreased without affecting the performance of the algorithm. However, in the biomedical example the dimension of the reconstruction space in kernel feature space never exceeds the dimension of the input space contrary to what is observed with toy examples.

4. CONCLUSIONS

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 linear SSA to the clusters formed with the multidimensional signals resulting from the embedding step. This piecewise linear approximation is then compared to a generically nonlinear subspace projection technique like KPCA. The toy examples serve to illustrate 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. The denoising performance turns out to be very effective. However, those examples also seem to indicate that KPCA can be superior in complex cases like the Hénon map where the local linear projection results in a too strong contraction of the data points to the low-dimensional submanifold. The superior performance of KPCA in this case results because the number of components to reconstruct the multidimensional signal can be larger than the input space dimension. In the extraction of EOG artifacts from EEG recordings, however, the number of components never exceeded the dimension of the input space and the signal extracted with KPCA is very similar to the one extracted with local SSA. However, KPCA seems to result in less distortions in the low frequency regime of the EEG spectrum. In summary, though local SSA is less complex hence much easier to implement, kernel-PCA results in less distortions or over-fitting in the low frequency regime where the EOG artifact dominates. Further it has to be mentioned that with the prominent EOG artifact, local SSA is able to extract also the 50 Hz line noise artifact simultaneously which is not the case with KPCA.

REFERENCES

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

[2] Y. Ephraim and H. L. V. Trees. A signal subspace approach for speech enhancement. *IEEE Transactions on Acoustic, Speech and Signal Processing,*, 3(4):251–266, 1995.

[3] V. Franc and V. Hlaváč. Stastical pattern recognition toolbox for matlab, 2004.

[4] M. Ghil, M. Allen, M. D. Dettinger, K. Ide, and e. al. Advanced spectral methods for climatic time series. *Reviews of Geophysics*, 40(1):3.1–3.41, 2002.

[5] N. Golyandina, V. Nekrutkin, and A. Zhigljavsky. *Analysis of Time Series Structure: SSA and Related Techniques*. Chapman & HALL/CRC, 2001.

[6] J. C. Gower. Adding a point to vector diagram in multivariate analysis. *Biometrika*, 55:582–585, 1968.

[7] P. Gruber, K. Stadlthanner, A. M. Tomé, A. R. Teixeira, F. J. Theis, C. G. Puntonet, and E. W. Lang. Denoising using local ICA and a generalized eigendecomposition with time-delayed signals. In *LNCS 3195, Proc. ICA' 2004*, pages 993–1000, Granada, 2004. Springer.

[8] J. T. Kwok and I. W. Tsang. The pre-image problem in kernel methods. *IEEE Transactions on Neural Networks*, 15(6):1517–1525, 2004.

[9] K.-R. Müller, S. Mika, G. Rätsch, K. Tsuda, and B. Schölkopf. An introduction to kernel-based algorithms. *IEEE Transactions on Neural Networks*, 12(2):181–202, 2001.

[10] B. Schölkopf, S. Mika, C. J. Barges, P. Knirsch, K.-R. Müller, G. Ratsch, and A. J.Smola. Input space versus feature space in kernel-based methods. *IEEE Transactions on Neural Networks*, 10(5):1000–1016, 1999.

[11] A. Teixeira, A. M. Tomé, E.W.Lang, P. Gruber, and A. M. d. Silva. On the use of clustering and local singular spectrum analysis to remove ocular artifacts from electroencephalograms. In *IJCNN2005*, pages 2514–2519, Montréal, Canada, 2005. IEEE.

[12] C. H. You, S. N. Koh, and S. Rahardja. Signal subspace speech enhancement for audible noise reduction. In *ICASPP 2005*, volume I, pages 145–148, Philadelphia, USA, 2005. IEEE.