Denoising Using Local ICA and a Generalized Eigendecomposition with Time-Delayed Signals

Peter Gruber[1], Kurt Stadlthanner[1], Ana Maria Tomé[2], Ana R. Teixeira[2], Fabian J. Theis[1], Carlos G. Puntonet[3], and Elmar W. Lang[1]

[1] Institute of Biophysics, University of Regensburg, 93040 Regensburg, Germany
elmar.lang@biologie.uni-regensburg.de
[2] Dept. de Electrónica e Telecomunicações/IEETA
Universidade de Aveiro, 3810 Aveiro, Portugal ana@ieeta.pt
[3] Dep. Arquitectura y Tecnologia de Computadores
Universidad de Granada, 18071 Granada, Spain
carlos@atc.ugr.es

Abstract. We present denoising algorithms based on either local independent component analysis (ICA) and a minimum description length (MDL) estimator or a generalized eigenvalue decomposition (GEVD) using a matrix pencil of time-delayed signals. Both methods are applied to signals embedded in delayed coordinates in a high-dim feature space Ω and denoising is achieved by projecting onto a lower dimensional signal subspace. We discuss the algorithms and provide applications to the analysis of 2D NOESY protein NMR spectra.

1 Introduction

Blind source separation (BSS) techniques have been shown to solve the problem of removing the prominent water artifact in 2D NOESY protein NMR spectra [9]. An algebraic algorithm [13], [14] based on a GEVD of a matrix pencil (GEVD-MP) has proven especially efficient in this respect. The results indicated, however, that the statistical separation process introduces unwanted additional noise into the reconstructed protein spectra (see Fig. 2). Hence denoising as a postprocessing step appeared necessary. Many denoising algorithms have been proposed [1], [4], [7] and [16] including algorithms based on local linear projective noise reduction. Noise is generally assumed to be additive Gaussian white noise whereas the signal comes from a deterministic source usually. This implies, using basic differential geometry, that the signal embedded in a high-dimensional feature space of delayed coordinates resides within a sub-manifold of the space of delayed coordinates. The task is to detect this signal manifold.

We propose denoising algorithms based either on local ICA using k-means clustering of the embedded signals and an MDL estimator of the dimension of the signal subspace or on a GEVD using a matrix pencil of time-delayed signals (GEVD-dMP). All experiments refer to the polypeptide P11 [10] to which both algorithms, local ICA and GEVD-dMP, have been applied.

C.G. Puntonet and A. Prieto (Eds.): ICA 2004, LNCS 3195, pp. 993–1000, 2004.

2 Denoising Using Local ICA

The algorithm we present is a local projective denoising algorithm. The idea is to embed the noisy signal into a high dimensional feature space of delayed signals. The denoising is then achieved by locally projecting the embedded signal onto a lower dimensional subspace which contains the characteristics of the noise free signal. The algorithm is based on local ICA using an minimum description length (MDL) criterion for parameter selection. To perform ICA we will use the popular FastICA algorithm by Hyvärinen and Oja [3], which performs ICA by maximizing the non-Gaussianity of the signal components.

Consider a signal $x(t)$ at discrete time steps $t = t_1, \ldots, t_n, \Delta t = t_n - t_{n-1}$ of which only its noise corrupted version $x_N(t) = x(t) + N(t)$ is measured. $N(t)$ are samples of a random variable with Gaussian distribution. First the noisy signal is transformed into a high-dimensional signal $\tilde{x}$ in the m-dimensional space of delayed coordinates according to $\tilde{\boldsymbol{x}}_N(t) := (x_N(t), \ldots, x_N(t + (m-1)\tau \mod n))^T, \quad \tau = c \cdot \Delta t, \quad c \in \mathcal{N}$. Then the problem is localized by selecting k clusters of the delayed time series $\{\tilde{\boldsymbol{x}_N}(t) \mid t = t_1, \ldots, t_n\}$ using a k-means cluster algorithm [5]. Now we can analyze these k m-dimensional signals using FastICA. We used an MDL criterion [6] to estimate the dimension p_{MDL} of the subspace onto which we project after using ICA:

$$p_{MDL} = \underset{p=1,\ldots,m}{\operatorname{argmin}} \left\{ -\ln \left(\frac{\Pi_{j=p+1}^{m} \lambda_j^{\frac{1}{m-p}}}{\frac{1}{m-p}\sum_{j=p+1}^{m} \lambda_j} \right)^{(m-p)n} + \left(pm - \frac{p^2}{2} + \frac{p}{2} + 1 \right) \cdot \left(\frac{1}{2} + \ln \gamma \right) - \frac{pm - \frac{p^2}{2} + \frac{p}{2} + 1}{p} \sum_{j=1}^{p} \ln \left(\lambda_j \sqrt{\frac{2}{n}} \right) \right\} \tag{1}$$

Here λ_j represents the ordered eigenvalues of the covariance matrix of the signal and γ represents a parameter of the MDL estimator. The MDL criterion is a maximum likelihood estimator of the number of signal components for data with additional white Gaussian noise. Using ICA we extract $p_{MDL} + 1$ independent components (ICs) of the signal (one additional component for the noise). Like in all MDL based algorithms noise reduction is achieved by projection of the signal onto a p_{MDL}-dimensional subspace. For PCA one applicable method is to select the largest components in terms of signal variance. For ICA applied to data with a non-Gaussian distribution we select the noise component as the component with the smallest value of the kurtosis. For non-stationary data with stationary noise we identify the noise by the least variance of its autocorrelation.

To reconstruct the noise reduced signal we reverse the clustering process to obtain a signal $\tilde{x_e} : \{1, \ldots, n\} \to \mathbb{R}^m$ and then average over the candidates in the delayed data.

$$x_e(t) := \frac{1}{m} \sum_{i=0}^{m-1} [\tilde{\boldsymbol{x}}_e(t - i \cdot \tau \mod n)]_i \tag{2}$$

The selection of optimal parameters m and k can again be based on an MDL criterion for the detected noise $\boldsymbol{e} := \boldsymbol{x} - \boldsymbol{x}_e$. Accordingly we project these signals $\boldsymbol{e}$ for different m and k in a high dimensional space of delayed coordinates and choose the parameters m and k such that the MDL criterion with respect to the eigenvalues of the correlation matrix of $\boldsymbol{e}$ is minimal.

3 Denoising Using GEVD-dMP

We present an algorithm similar to the recently proposed algorithm dAMUSE [12], [11] which can be used to solve BSS problems and simultaneously denoise the estimated source signals. Consider sensor signals x_i embedded in a high-dim feature space Ω of delayed signals. The trajectory matrix [8] of the sensor signals $x_i(t_0)$ and their M delayed versions $x_i(t_0 + m\Delta t),\ \ m = 0, ..., M-1$ computed for a set of L samples is given by ($t_0 = 0$ for simplicity)

$$\mathbf{X}_i = \begin{bmatrix} x_i((M-1)\Delta t) & x_i(\tau + (M-1)\Delta t) & \cdots & x_i(L\tau) \\ x_i((M-2)\Delta t) & x_i(\tau + (M-2)\Delta t) & \cdots & x_i(L\tau - \Delta t) \\ \vdots & \vdots & \cdots & \vdots \\ x_i(0) & x_i(\tau) & \cdots & x_i(L\tau - (M-1)\Delta t) \end{bmatrix} \tag{3}$$

where τ^{-1} is the sampling rate. Considering a group of N L-dim sensor signals, $\boldsymbol{x}_i,\ i = 1 \ldots N$, the trajectory matrix of the set will be a concatenation of the component trajectory matrices computed for each sensor. Assuming that each sensor signal is a linear combination of N underlying but unknown source signals ($\boldsymbol{s}_i$), a matrix $\mathbf{S}$ can be written in analogy to eqn(3). Then the sensor signals can be expressed as $\mathbf{X} = \mathbf{AS}$, where the mixing matrix $\mathbf{A} = \mathbf{a} \otimes \mathbf{I}$ is a block matrix with a diagonal matrix $a_{ij}\mathbf{I}_{M \times M}$ in each block. The matrix $\mathbf{I}_{MxM}$ is the identity matrix and the mixing coefficient a_{ij} relates the sensor signal i with the source signal j.

Considering NMR spectra it seems natural to deal with data in the frequency domain. Hence a data matrix $\hat{\mathbf{X}}$ is constructed by Fourier transforming every row of $\mathbf{X}$ to the frequency domain. Additionally a filtered version $\hat{\mathbf{X}}_f$ of $\hat{\mathbf{X}}$ is generated by computing the Hadamard product between the rows of $\hat{\mathbf{X}}$ and the frequency response function of an appropriate filter. Then a matrix pencil $(\mathbf{R}_{x,f}, \mathbf{R}_x)$ is formed where $\mathbf{R}_x$ is the correlation matrix of the unfiltered and $\mathbf{R}_{x,f}$ is the correlation matrix of the filtered signals. According to the linear mixing model the correlation matrix $\mathbf{R}_x$ can then be related to a corresponding matrix in the source signal domain via:

$$\mathbf{R}_x = \hat{\mathbf{X}}\hat{\mathbf{X}}^H = \mathbf{A}\mathbf{R}_s\mathbf{A}^H = \mathbf{A}\hat{\mathbf{S}}\hat{\mathbf{S}}^H\mathbf{A}^H \tag{4}$$

Analogously the correlation matrix of the filtered signals $\mathbf{R}_{x,f}$ is related to the correlation matrix $\mathbf{R}_{s,f}$ of the filtered sources.

Then the two pairs of matrices $(\mathbf{R}_{x,f}, \mathbf{R}_x)$ and $(\mathbf{R}_{s,f}, \mathbf{R}_s)$ represent a congruent pencil [15] with identical eigenvalues, i.e. $\mathbf{D}_x = \mathbf{D}_s$ and corresponding

eigenvectors which are related by $\mathbf{E}_s = \mathbf{A}^H\mathbf{E}_x$ in case of non-degenerate eigenvalues.

Assuming that all sources are uncorrelated, the matrices $\mathbf{R}_s$ and $\mathbf{R}_{s,f}$ are block diagonal with block matrices along the diagonal given by $\mathbf{R}_{mm}=\langle\hat{\mathbf{s}}_m(\hat{\mathbf{s}}_m)^H\rangle$ and $\mathbf{R}_{mm,f} = \langle\hat{\mathbf{s}}_{m,f}(\hat{\mathbf{s}}_{m,f})^H\rangle$. The eigenvector matrix of the GEVD of the pencil $(\mathbf{R}_{s,f}, \mathbf{R}_s)$ is also block-diagonal with the block matrix $\mathbf{E}_{mm}$ on the diagonal being the $M \times M$ eigenvector matrix of the GEVD of the pencil $(\mathbf{R}_{mm}, \mathbf{R}_{mm,f})$. The independent components can be estimated from linearly transformed sensor signals via

$$\mathbf{Y} = \mathbf{E}_x^H\mathbf{X} = \mathbf{E}_x^H\mathbf{A}\mathbf{S} = \mathbf{E}_s^H\mathbf{S} \tag{5}$$

and turn out to be filtered versions of the underlying source signals.

To simultaneously perform BSS and denoising the GEVD of $(\mathbf{R}_{x,f}, \mathbf{R}_x)$ is determined in a two step procedure. First, the EVD of $\mathbf{R}_x = \mathbf{U}\mathbf{V}\mathbf{U}^H$ is calculated but only its l largest eigenvalues and corresponding eigenvectors are considered assuming that small eigenvalues are related to noise only. Following this dimension reduction the $l \times NM$ matrix $\mathbf{Q} = \mathbf{V}^{-1/2}\mathbf{U}^H$ is defined and the EVD problem of the matrix $\mathbf{C} = \mathbf{Q}\mathbf{R}_{x,f}\mathbf{Q}^H$ is solved. Then the eigenvector matrix $\mathbf{E}_x$ of the matrix pencil is

$$\mathbf{E}_x = \mathbf{Q}^H\mathbf{E}_C \tag{6}$$

where $\mathbf{E}_C$ is the eigenvector matrix of $\mathbf{C}$.

4 Denoising of Reconstructed NMR Spectra

Both algorithms have been applied recently to artificially generated signals and random noise to test their performance and evaluate their properties [2], [11]. These results suggest that a local ICA approach is more effective when the signal is infested with a large amount of noise whereas local PCA seems to be better suited for signals with high SNRs. Comparable investigations have been performed with the dAMUSE algorithm also with similar results [11].

In the following the local ICA based denoising algorithm uses the component kurtosis for noise selection and the MDL criterion to determine the number of independent signal components. The MDL criterion is used a second time to optimize the dimension m of the embedding space and the number k of neighborhoods to be used in local ICA. We apply the local ICA denoising algorithm as well as a local PCA and a Kernel-PCA based denoising algorithm to 2D NOESY NMR spectra of the polypeptide P11 [10].

The local ICA algorithm has been applied only to those independent components (IC), obtained with a GEVD-MP algorithm [10], related with the water peak. These ICs have been considered the signal now and were embedded in an m-dim feature space of delayed coordinates. There a k-means cluster algorithm was used to detect clusters of similar feature vectors. An MDL criterion has been used to estimate optimal values for the feature space dimension m and the number k of nearest neighbors. On these clusters an ICA has been performed locally

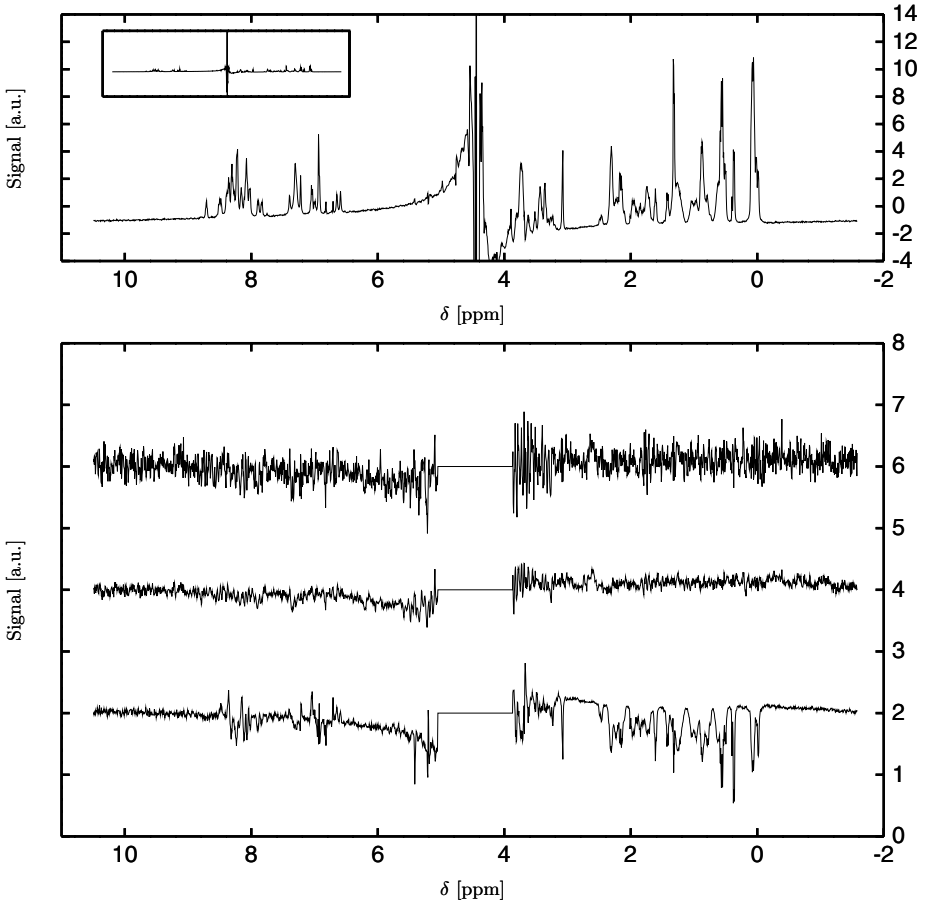

Fig. 1. The graph uncovers the differences of local ICA and Kernel-PCA denoising. The original spectrum is displayed on top with an insert showing the full water peak. The three curves represent the difference of the original and reconstructed spectra using *top* GEVD-MP, *middle* local ICA denoising, and *bottom* Kernel-PCA denoising. Note that the graphs are vertically translated by 2, 4 and 6 *a.u.*, respectively. Also note that both graphs have an identical scale.

using the fastICA algorithm. Again an MDL criterion was used to estimate an optimal number of signal components to be used to reconstruct the noise-reduced signals. These now represent "noise-free" versions of the ICs obtained with the GEVD-MP algorithm. Calculating the difference of the ICs obtained from the GEVD-MP algorithm directly and those reconstructed after the local ICA denoising has been applied, yields the noise contribution to these components. Now noise-reduced artifact-free protein spectra can be reconstructed using those ICs from the GEVD-MP algorithm not assigned to the water peak and the noise components obtained after removing the noise-free water components. Adding

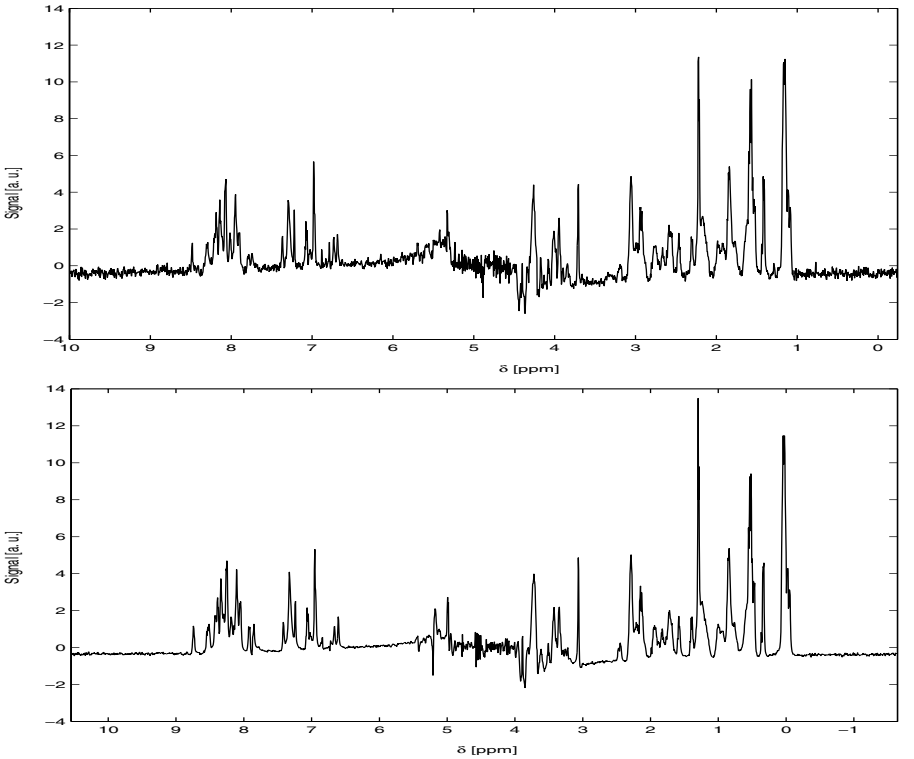

Fig. 2. Reconstructed spectrum of the polypeptide P11 using top: GEVD-MP [10] bottom: GEVD-dMP.

the noise components during the reconstruction process is essential as the noise which the ICs of the GEVD-MP algorithm convey only results from the statistical nature of the separation process. Hence all noise contributions to all ICs must be added up for them to compensate and result in the low noise content of the experimental spectra as can clearly be seen from figure 1. On the part of the spectrum away from the water artifact, we could estimate the increase of the SNR defined by

$$SNR(\boldsymbol{x}, \boldsymbol{x}_N)[dB] := 20 \log_{10} \frac{||\boldsymbol{x}||}{||\boldsymbol{x} - \boldsymbol{x}_N||} \tag{7}$$

with the original spectrum as reference. We calculated a SNR of 17.3 dB of the noisy spectrum and a SNR of 21.6 dB after applying local ICA denoising.

We compare the reconstructed artifact-free protein spectrum of the local ICA denoising algorithm to the result of a Kernel-PCA based denoising algorithm [7] using a gaussian kernel in figure 1. The figure depicts the differences between the denoised reconstructed spectra and the original spectrum in regions away from the water peak. Local ICA denoising reduces the noise without changing

the intensity of the protein signals, whereas Kernel-PCA denoising [7] clearly distorts the peak amplitudes of the protein resonances as well. This is detrimental to any spatial structure determination of the protein and is not acceptable to NMR spectroscopists.

The GEVD-dMP algorithm can be used elegantly to *simultaneously* separate the water artifact from the protein spectrum and perform denoising of the reconstructed spectrum. The starting point for the application of GEVD-dMP to P11 were 128 time domain signals taken from a 2D NOESY NMR experiment where each of the 128 signals consisted of 2048 data points. For every signal the component trajectory matrix $\mathbf{X}_i$, $i = 1, ..., 128$, (cf. eqn(3)) was formed by using $M = 1$ time delays of size $\Delta t = 2 \cdot \tau$. Thus the resulting trajectory matrix $\mathbf{X}$ was of size 256×2046. The matrix $\hat{\mathbf{X}}$ was determined by Fourier transforming each row of $\mathbf{X}$ to the frequency domain. After calculating the correlation matrix $\mathbf{R}_x = \hat{\mathbf{X}}\hat{\mathbf{X}}^H$ and a gaussian shaped filter of width $\sigma = 1$ was applied to every row of $\hat{\mathbf{X}}$ leading to the matrix $\hat{\mathbf{X}}_f$. Its correlation matrix was determined by $\mathbf{R}_{x,f} = \hat{\mathbf{X}}_f\hat{\mathbf{X}}_f^H$.

In the two step GEVD procedure, after the first EVD only the $l = 95$ largest eigenvalues of the 256×256 correlation matrix $\mathbf{R}$ were considered in order to reduce noise. Then the 95×2046 matrix $\mathbf{Q}$ was computed and the EVD of the matrix $\mathbf{C} = \mathbf{Q}\mathbf{R}_{x,f}\mathbf{Q}^H$ was performed which eventually lead to the eigenvector matrix $\mathbf{E}_x$ of the matrix pencil (eqn 6). Finally, those estimated components of $\mathbf{Y}$ (eqn. 5) which showed a high spectral density at the resonance frequency of the water protons were set to zero to reconstruct the artifact-free protein spectra. Note that with less than the 95 largest eigenvalues the separation of the water and the protein signals failed whereas considering more than 100 of the largest eigenvalues lead to a drastic increase in noise. Fig. 2 compares the results obtained by the standard GEVD-MP and the GEVD-dMP algorithms corresponding to SNRs of 17.3 dB and 22.43 dB, respectively.

5 Conclusions

Water artifact separation from 2D NOESY NMR protein spectra with statistical techniques like ICA introduce unwanted noise into the independent components obtained. We present noise reduction techniques using local ICA and an MDL based selection of the signal subspace as well as a GEVD-dMP algorithm. Both algorithms are based on local projective methods imposed on signals embedded in a high-dim feature space of delayed coordinates. The proposed methods are very effective in reducing the noise and show better results than a Kernel-based PCA method. Whereas local ICA denoising needs another GEVD or ICA preprocessing step to effect the artifact separation, GEVD-dMP provides both the artifact removal and the denoising in one stroke and is computationally very efficient with comparable results.

References

1. A. Effern, K. Lehnertz, T. Schreiber, P. David T. Grunwald, and C.E. Elger. Nonlinear denoising of transient signals with application to event-related potentials. *Physica D*, 140:257–266, 2000.
2. Peter Gruber, Fabian J. Theis, Ana Maria Tomé, and Elmar W. Lang. Automatic denoising using local independent component analysis. In *Proc. International ICSC Conference on Engineering of Intelligent Systems, ESI'2004*, 2004.
3. A. Hyvärinen and E. Oja. A fast fixed-point algorithm for independent component analysis. *Neural Computation*, 9:1483–1492, 1997.
4. Aapo Hyvärinen, Patrik Hoyer, and Erkki Oja. *Intelligent Signal Processing*, chapter Image Denoising by Sparse Code Shrinkage. IEEE Press, 2001.
5. A.K. Jain and R.C. Dubes. *Algorithms for Clustering Data*. Prentice Hall: New Jersey, 1988.
6. A.P. Liavas and P.A. Regalia. On the behavior of information theoretic criteria for model order selection. *IEEE Transactions on Signal Processing*, 49:1689–1695, 2001.
7. S. Mika, B. Schölkopf, A. Smola, K. Müller, M. Scholz, and G. Rätsch. Kernel PCA and denoising in feature spaces. *Adv. Neural Information Processing Systems, NIPS11*, 11, 1998.
8. V. Moskvina and K. M. Schmidt. Approximate projectors in singular spectrum analysis. *SIAM Journal Mat. Anal. Appl.*, 24(4):932–942, 2003.
9. K. Stadlthanner, A. M. Tomé, F. J. Theis, W. Gronwald, K. R. Kalbitzer, and E. W. Lang. Blind source separation of water artifacts in NMR spectra using a matrix pencil. In *Fourth International Symposium On Independent Component Analysis and Blind Source Separation, ICA'2003*, pages 167–172, Nara, Japan, 2003.
10. K. Stadlthanner, A. M. Tomé, F. J. Theis, W. Gronwald, K. R. Kalbitzer, and E. W. Lang. On the use of independent component analysis to remove water artifacts of 2D NMR protein spectra. In *7th Portuguese Conference on Biomedical Engineering, BIOENG'2003*, Lisbon, Portugal, 2003.
11. A. R. Teixeira, A. M. Tomé, E. W. Lang, and K. Stadlthanner. dAMUSE - A Tool for BSS and Denoising. *ICA2004, submitted*, 2004.
12. Ana R. Teixeira, A. P. Rocha, R. Almeida, and A. M. Tomé. The analysis of heart rate variability using independent component signals. In *2nd Intern. Conf. on Biomedical Engineering, BIOMED'2004*, pages 240–243, Innsbruck, Austria, 2004. IASTED.
13. Ana Maria Tomé. Blind source separation using a matrix pencil. In *Int. Joint Conf. on Neural Networks, IJCNN'2000*, Como, Italy, 2000.
14. Ana Maria Tomé. An iterative eigendecomposition approach to blind source separation. In *3rd Intern. Conf. on Independent Component Analysis and Signal Separation, ICA'2003*, pages 424–428, San Diego, USA, 2001.
15. Ana Maria Tomé and Nuno Ferreira. On-line source separation of temporally correlated signals. In *European Signal Processing Conference, EUSIPCO2002*, Toulouse, France, 2002.
16. Rolf Vetter, J.M. Vesin, Patrick Celka, and Jens Krauss Philippe Renevey. Automatic nonlinear noise reduction using local principal component analysis and MDL parameter selection. *Proceedings of the IASTED International Conference on Signal Processing Pattern Recognition and Applications (SPPRA 02) Crete*, pages 290–294, 2002.