EIS 2004
Island of Madeira, Portugal

AUTOMATIC DENOISING USING LOCAL INDEPENDENT COMPONENT ANALYSIS

[1] P. Gruber
petergruber@gmx.net

[1] F.J. Theis
fabian@theis.name

[2] A. M. Tomé
ana@ieeta.pt

[1] E.W. Lang
elmar.lang@biologie.uni-r.de

[1] Institute of Biophysics, University of Regensburg, D-93040 Regensburg, Germany
[2] Dep. Electrónica e Telecomunicações, Universidade de Aveiro, P-3810-193 Aveiro, Portugal

ABSTRACT

We present a denoising algorithm for enhancing noisy signals based on local independent component analysis (ICA). We extend a noise reduction algorithm proposed by Vetter et al. [11] by using ICA to separate the signal from the noise. This is done by applying ICA to the signal in localized delayed coordinates. The components resembling noise are detected using estimators of kurtosis or the variance of the autocorrelation. This algorithm can also be applied to the problem of denoising multidimensional data like images or fMRI data sets. In comparison to denoising algorithms using wavelets or Wiener filters the local PCA and ICA algorithms perform considerably better especially ICA algorithms which consider the estimation of higher order statistical moments like kurtosis.

INTRODUCTION

In many fields of signal analysis the examined signals bear considerable noise which is usually assumed to be additive and non-correlated. For example in exploratory data analysis of medical data such as EEG or fMRI using statistical methods like ICA the prevalent noise greatly degrades the reliability of the algorithms and the underlying processes cannot be identified. Therefore many denoising algorithms have been proposed [2][5][9][11]. Vetter et al.[11] suggest an algorithm based on local linear projective noise reduction. The idea is to observe the data in a high-dimensional space of delayed coordinates and denoise the data locally through a projection into the lower dimensional subspace of the deterministic signals. For parameter selection an minimum description length (MDL) criterion is used to select optimal parameters. For an introduction to the concept of minimum description length see [13] and for a good description of the criterion see [3].

The noise is assumed to be stationary (at least locally stationary) Gaussian white noise. The signal usually comes from a deterministic or at least predictable source and can be described as a smooth function evaluated at discrete timesteps small enough to capture the characteristics of the function. That implies, using basic differential geometry, that the signal in delayed coordinates resides within a submanifold of the space of delayed coordinates. The task is to detect this signal manifold. In the following we call this manifold the signal+noise subspace since it contains all of the delayed signal as well as that part of the delayed noise which extends in the same directions as the signal.

PCA AND ICA

Principal component analysis (PCA)[10] is one of the most common multivariate data analysis tools. It tries to linearly transform given data $\vec{x}(t)$ following the data model

$$\vec{x}(t) = \mathbf{A}\vec{s}(t) + \vec{N}(t) \tag{1}$$

into uncorrelated data (feature space) $\vec{s}(t)$. The new orthonormal basis vectors $\vec{s}(t)$ are called principal components. PCA can be performed by eigenvalue-decomposition of the data covariance $\mathbf{C} = E[\vec{x}(t)\vec{x}^T(t)]$. The components with the largest eigenvalues contain the main signal information.

In independent component analysis (ICA), given a random vector or sensor signal $\vec{x}(t)$ again following a data model according to equn.(refeqn1), the goal is to find its statistically independent components $\vec{s}(t)$. In contrast to correlation-based transformations (PCA) independent component analysis (ICA) renders the output signals as statistically independent as possible by evaluating higher-order statistics. The idea of ICA was first expressed by Jutten and Hérault [4] while the term ICA was later coined by Comon[1]. We will use the FastICA algorithm by Hyvärinen and Oja [6], which performs ICA by maximizing the non-Gaussianity of the signal components.

PSfrag replacements
SNR
MDL
Delay
dimension
k

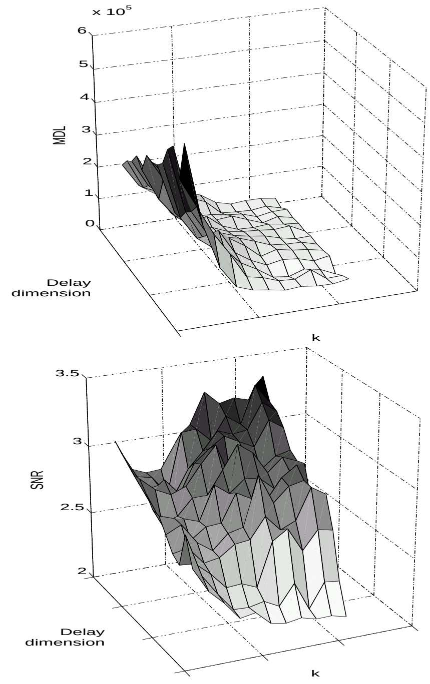

Fig. 1. Comparison of MDL selected parameters and SNRs of denoised signals for different m and k. Delay dimension ranges from 40 to 160 and k from 5 to 35. Selected values by MDL m=60,k=35 SNR=2.7, but best SNR=3.5.

LOCAL PROJECTIVE DENOISING ALGORITHMS

The algorithm we will present represents a local projective denoising algorithm. The idea of these algorithms is to embed the noisy signal into a high dimensional feature space by a method which adds the temporal information to the signal. The denoising is then achieved by locally projecting the embedded noisy signal vectors onto a lower dimensional subspace which contains the characteristics of the noise free signal. Finally the signal has to be reconstructed using the various candidates generated by the embedding.

Implementation of the algorithm

We now present a denoising algorithm based on local ICA or PCA using an MDL criterion for parameter selection. Consider the situation, where we have a signal $x(t)$ at discrete timesteps $t = 1, \dots, n$. But only a distorted signal is measured

$$x_N(t) = x(t) + N(t) \tag{2}$$

where $N(t)$ are samples of a random variable with Gaussian distribution, i.e. $x_N(t)$ equals $x(t)$ up to additive stationary white noise.

At first the noisy scalar signal $x_N(t)$is transformed into a sensor signal vector $\vec{x}_N(t)$ in the m-dimensional space of delayed coordinates

$$\vec{x}_N(t) := (x_N(t), \dots, x_N(t+m-1 \mod n))$$

For computational simplicity we use the samples in round robin manner. This leads to the problem of compatibility of the beginning and end of the signal and is solved with some preprocessing as explained in the last section.

Then we localize the problem by selecting k clusters of the delayed time series $\{\vec{x_N}(t) \mid t = 1, \dots, n\}$. This can for example be done by a k-means cluster algorithm[7], which seems to be appropriate for noise selection schemes based on the strength or the kurtosis of the signal since the statistical properties do not depend on the signal structure. Using other methods like considering the variance of autocorrelations it is usually better to find an appropriate partitioning of $\{1, \dots, n\}$ into k *successive* parts since this preserves the time structure of the signal.

Now we can analyze these k m-dimensional signals $\vec{x}^{(}k)(t)$ using PCA or ICA. We used two different criteria to estimate the number of signal+noise components, i.e. the dimension of the subspace onto which we project after using PCA or ICA. One criterion is the MDL estimator for the data model in eqn.(2) pro-

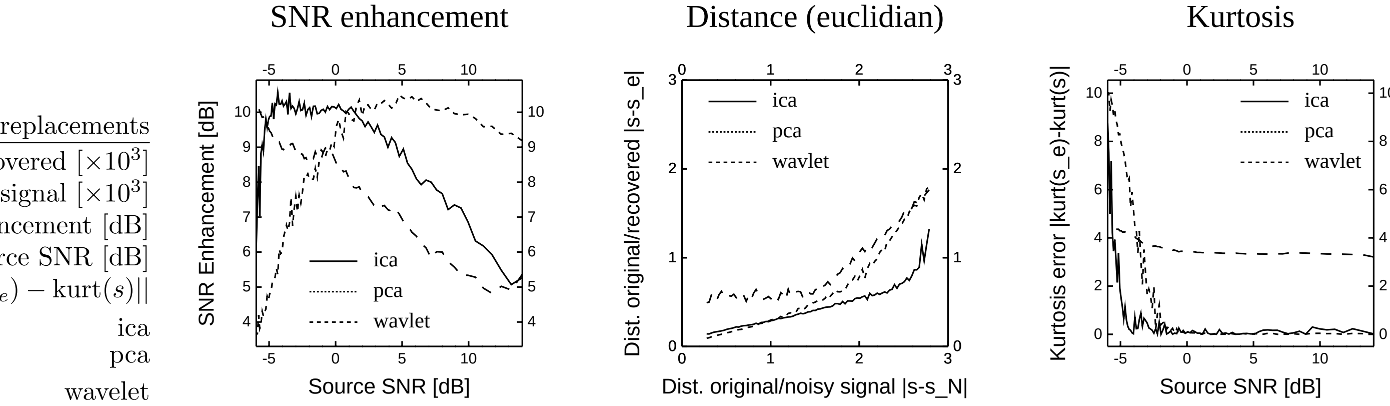

Fig. 2. Comparison between local PCA, ICA and wavelet based denoising. Here the mean square error of two signals s, t with n samples is $\frac{1}{n}\sum_i ||s_i - t_i||^2$.

posed by Vetter [12]

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

where δ_j are the ordered singular values of the covariance matrix of the signal and γ a parameter of the MDL estimator and hence of the final denoising algorithm. The MDL criterion is a maximum likelihood estimator for the number of signal components for data with additional white Gaussian noise.

Based on [8] and our observations we also used another approach: We clustered the singular values of the covariance matrix into two clusters using k-means and defined p_{cl} as the number of elements in the cluster which contains the largest eigenvalue. This gives a good estimation of the number of signal components if the noise variances are not clustered well enough together but nevertheless are separated from the signal by a large gap.

In the ICA case we apply ICA to extract $p_{MDL}+1$ or $p_{cl}+1$ independent components of the signal (one additional component for the noise). Like in all MDL based algorithms the noise reduction is achieved by projection of the signal onto a p_{MDL} or p_{cl}-dimensional subspace. For PCA one applicable method is to select the largest components in terms of signal variance. For ICA we can apply several methods depending on the nature of the data. For signals 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 $x_e(t)$ representing an approximation to the noise-free signal $x(t)$ we only have to reverse the clustering of the data to get a signal

$$\vec{x_e}(t) : \{1, \ldots, n\} \to \mathbb{R}^m$$

and then average over the candidates in the delayed data.

$$x_e(t) := \frac{1}{m}\sum_{i=0}^{m-1} \left[\vec{x}_e(t - i \mod n)\right]_i \tag{3}$$

In experiments it has proven to be effective to weight this sum such that the center has the strongest influence on the resulting signal.

Now we still have to find suitable values for the global parameters m, k and γ. Vetter proposed to base the selection of m and k also on a MDL criterion for the reconstruction error corresponding closely to the detected noise $e(t) := x_N(t) - x_e(t) \approx .N(t)$. Again we represent these signals $e(t)$ for the 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 singular values of the correlation matrix

EXPERIMENTAL RESULTS

We will present some sample experimental results using generated signals and artificial noise. In the following

replacements
$\times 10^3$

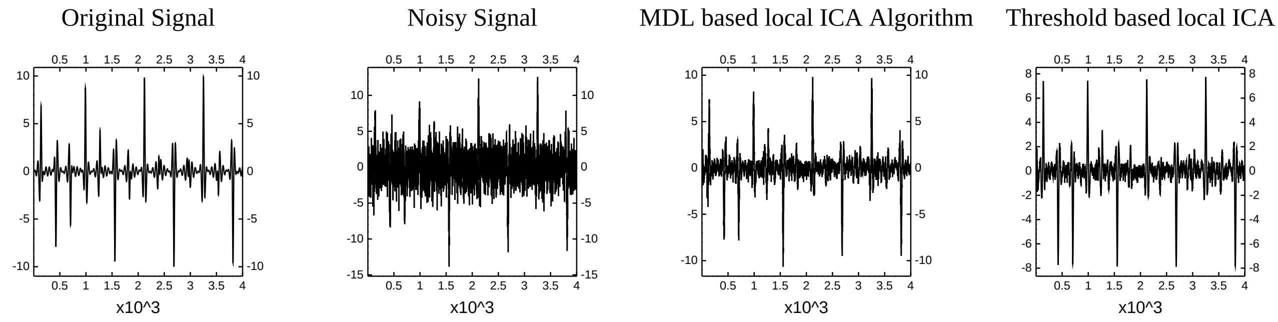

Fig. 3. Comparison between MDL and threshold denoising. The comparison was done with an artificial signal and a known SNR of 0, $m = 40$ and $k = 35$. (MDL-SNR enhancement 8.9, Threshold-SNR enhancement 10.5)

the ICA based denoising algorithm uses the component kurtosis for noise selection.

Discussion of MDL based selection of principal or independent components

In our experiments we have noticed that the MDL criterion often overestimates the number of components of the signal+noise subspace. Therefore we also give a comparison to a threshold based algorithm which requires prior knowledge of the signal to noise ratio, see also the next sections.

The MDL criterion is used a second time to select the best dimension m for the delayed signal and k the number of neighborhoods to use. This method for selecting the best algorithm seems to be problematic for some situations, see for example figure 1. Experiments indicate that this is especially the case if the Signal to Noise Ratio (SNR) defined by

$$SNR(x, x_N) := 20 \log_{10} \frac{||x||}{||x - x_N||} \tag{4}$$

is very low i.e. in situations with strong noise.

As explained in [8] there are some situations where using the MDL criterion can lead to a massive overmodeling of the signal, i.e. underestimating the number of noise components. This is especially true if the noise is not completely white. Since the overmodeling mostly happens if the singular values of the covariance matrix which describe noise components are not sufficiently close together and are divided from the signal components by a gap, the clustering criterion can yield better results. For an actual example of this situation see the last section.

Comparisons between local ICA and local PCA for selection of the noise subspace

We use the artificial signal shown in figure 3 with varying additive Gaussian white noise. We apply the denoising algorithm which is described at the beginning of this text as well as a wavelet based denoising algorithm. The results are depicted in figure 2.

The first and second diagram compares the performance, here the enhancement of SNR (see equation4) and mean error, of the three different algorithms depending on input SNR. Here a source SNR of 0 describes the case where the signal and the noise have the same strength. The third graph shows the difference in kurtosis of the original signal and the source signal again depending on the input SNR. All three diagrams correspond to the same data set, i.e. the same signal and, for a given input SNR, the same additive noise.

Our examples suggest that a local ICA approach is more effective when the signal is infested with a large amount of noise whereas the local PCA seems to be better suited for signals with high SNRs. This might be due to the nature of the our selection of subspaces based on kurtosis or variance of the autocorrelation. Also the comparison of higher statistical moments of the restored data, for example the kurtosis, indicate that the noise reduction can be enhanced if we are using a local ICA approach.

Comparison between the MDL criterion and the threshold criterion

Using a threshold instead of the adaptive MDL criterion to select the dimension of the signal+noise subspace has proven to be more effective in some examples, although it requires knowledge about the strength of the additive noise. In this case we can discard the weakest components up to cumulative strength given

Fig. 4. Image denoising. Comparison of a wavelet based denoising filter, a wiener filter and a local PCA with cluster criterion on a image infested with Gaussian noise.

by the known noise level.

The result of a simulation is displayed in figure 3. For this simulation we used the same signal as in the last section and additional noise with a SNR=0. Therefore we chose for the threshold criterion the weakest components whose cumulative strength was at least half as strong as the complete noisy signal.

MULTIDIMENSIONAL APPLICATIONS

A direct generalization of the algorithm for multidimensional data by simply looking at delayed coordinates of vectors instead of scalars seems to be unpractical due to the computational effort. More importantly this approach significantly lessens the number of samples yielding far less accurate estimators of important aspects like MDL or the kurtosis in the ICA case.

Experiments seem to indicate a solution by arranging the data in the following way: Let $S(x, y)$ for $x, y = 1, \ldots, n$ be a 2-dimensional signal. Define the 1-dimensional signal with $2n^2$-samples by

$$\begin{aligned} s := \big(&S(1,1), \ldots, S(n,1), \ldots \\ &\ldots, S(n,1), S(n-1,1), \ldots, S(1,1), \ldots \\ &\ldots, S(1,2), \ldots, S(n,2), S(n,2), \ldots \\ &\ldots, S(n,n), \ldots, S(1,n)\big) \end{aligned}$$

This can be easily extended to higher dimensional data. Then apply any ordinary 1-dimensional denoising algorithm. This signal arrangement mirrors the row structure of the image. Applying the algorithm once more to the transposed denoised image also uses the column structure.

Depending on the nature of the signal another approach can be effective. Instead of converting the multidimensional data into 2-dimensional data prior to applying the algorithm, we can use translated versions of the signal wrapped around at the corners and then transform each of these to 1-dimensional data and using them as the delayed coordinates. For an image P represented by a matrix $P = (a_{ij})_{i,j=1\ldots10}$ that means

$\ldots, P_{-1-1}, P_{-10}, \ldots, P_{10}, P_{11} \ldots$ with

$$P_{-1-1} = \begin{pmatrix} a_{22} & \ldots & a_{210} & a_{21} \\ \vdots & & \vdots & \vdots \\ a_{102} & \ldots & a_{1010} & a_{101} \\ a_{12} & \ldots & a_{110} & a_{11} \end{pmatrix}$$

represent the delayed data for the local subspace algorithm.

In figure 4 we see that this approach using the clustering criterion to select the number of components and PCA to identify the noise components compares favourably to a wavelet denoising algorithm and a Wiener filter.

CONCLUSIONS

We presented extensions to the local PCA noise reduction algorithm using local ICA and a kurtosis-based selection of the noise subspace. We pointed out that this can further enhance the signal reconstruction in comparison to the original algorithm. Further we saw that in some situations the MDL based subspace selection does not yield optimal results and we provided a cluster criterion to replace the MDL criterion. In the future we will investigate other methods of measuring the noisiness of the independent components since the kurtosis estimator seems to be inappropriate for low noise situations. Possible extensions could by given by a combination of kurtosis and variance based selection. Also we want to investigate if we could use a MDL based criterion to not only estimate the optimal parameters of the algorithm but also to select the best alternative approach (i.e. PCA or ICA, MDL or cluster criterion) for the given situation.

ACKNOWLEDGMENTS

This research was supported by the BMBF[1].

REFERENCES

[1] P. Comon. Independent component analysis - a new concept? *Signal Processing*, 36:287–314, 1994.

[2] 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.

[3] E. Fishler and H. Messer. On the use of order statistics for improved detection of signals by the mdl criterion. *IEEE Transactions on Signal Processing*, 48:2242–2247, 2000.

[4] J. Hérault and C. Jutten. Space or time adaptive signal processing by neural network models. In J.S. Denker, editor, *Neural Networks for Computing. Proceedings of the AIP Conference*, pages 206–211, New York, 1986. American Institute of Physics.

[5] Aapo Hyvärinen, Patrik Hoyer, and Erkki Oja. *Intelligent Signal Processing*, chapter Image Denoising by Sparse Code Shrinkage. IEEE Press, 2001.

[6] A. Hyvärinen and E. Oja. A fast fixed-point algorithm for independent component analysis. *Neural Computation*, 9:1483–1492, 1997.

[7] A.K. Jain and R.C. Dubes. *Algorithms for Clustering Data*. Prentice Hall: New Jersey, 1988.

[8] 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.

[9] Chor Tin Ma, Zhi Ding, and Sze Fong Yau. A two-stage algorithm for mimo blind deconvolution of nonstationary colored noise. *IEEE Transactions on Signal Processing*, 48:1187–1192, 2000.

[10] K. Pearson. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2:559–572, 1901.

[11] 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.

[12] Rolf Vetter. *Extraction of efficient and characteristic features of multidimensional time series*. PhD thesis, EPFL, Lausanne, 1999.

[13] P. Vitányi and M. Li. Miminum description length induction, bayesianism, and kolmogorov complexity. *IEEE Transactions on Information Theory*, 46:446–464, 2000.

[1] project 'ModKog'