CONSIDERING TEMPORAL STRUCTURES IN INDEPENDENT COMPONENT ANALYSIS

Andreas Jung[*,1] *and Andreas Kaiser*[2]

[1]Institute for Theoretical Physics, University of Regensburg, Germany
[2]Max Planck Institute for the Physics of Complex Systems, Dresden, Germany

ABSTRACT

In this work we show how temporal structures in time series can be used in the framework of independent component analysis assuming the signals arise from Markov chains with finite order. Taking into account the past of the underlying processes, by using *time embedding* vectors, not only *instantaneous independent* but also *uncoupled* sources can be found. As a result signals which are gaussian distributed at each time can be decomposed as long as the time embedding vectors are non-gaussian. Using the model of independent time embedding vectors, we derive an algorithm (FastTeICA) which is similar to the well known FastICA algorithm introduced by Hyvärinen (1999).

A weakening of the strict assumption of independent embedding vectors which still takes into account the dynamics of the processes for signal decomposition can be achieved by assuming independent increments, i.e. the change of state of the processes. Both approaches, independent states and independent dynamics, are a special case of the independence of the time embedding vectors.

1. INTRODUCTION

Independent component analysis is a signal processing tool for decomposing observed signals into a set of stochastically independent signals. In the blind source separation problem additionally one assumes an underlying mixture model generating the observed signals. As sources independent random variables are assumed which are mixed (non-)linear.

However, in many applications one deals with time signals which are given by stochastic processes. Hence, the random variables of each source are usually correlated in time. The so far proposed algorithms [1, 2] do not consider these additional temporal structures. In particular, scrambling the time series leads to the same result. Thus, one could try to improve the recovery of the sources, using the extra structure contained in the signals. A set of algorithms dealing with time structures have been proposed in the past, using autocorrelation functions [3, 4, 5, 6], autoregressive models based on maximum likelihood approaches [7, 8], the non-stationarity of signals [9] and quasi maximum likelihood approaches [10].

In this work we extend the idea of looking for *(instantaneously) independent* signals, i.e. the states of the sources $s_1(t), \dots, s_n(t)$ at time t are independent, on independent signals and *uncoupled* signals. For this, we assume that the processes can be approximated by Markov processes of order m. We say, the Markov processes x and y, both of order m, are *(stochastically) uncoupled* from each other, if the transition probability of the process (x, y) factorizes into the transition probabilities of the processes x and y such that the transition probabilities of x does not dependent on the past of y and vice versa, i.e. $p(x(t), y(t)|x(t-1), ..., x(t-m), y(t-1), ..., y(t-m)) = p(x(t)|x(t-1), ..., x(t-m)) \cdot p(y(t)|y(t-1), ..., y(t-m))$. Based on this model of coupling, it is possible to determine the coupling direction and to quantify the information transfer between two stochastic processes [11, 12]. In particular, x and y are uncoupled if the $(m+1)$-dimensional *time embedding vectors* $(x(t), ..., x(t-m))$ and $(y(t), ..., y(t-m))$ are stochastically independent. The generalization on more than two processes is straightforward.

In the following we show, how temporal structures can be included in the ICA. In particular, we derive an algorithm analog to the FastICA algorithm, proposed in [13, 1], including temporal structures by using time embedding vectors. Although the assumption of independent time embedding vectors is very powerful, one has to deal with numerical instabilities. In order to solve this difficulty the strict requirement of independent time embedding vectors is weakened. Consider

*email: *Andreas.Jung@physik.uni-regensburg.de*

uncoupled Markov processes of order one where the dynamics of each process is independent of the other. In this case, the increments of each process, i.e. the change of state within one time step, are independent. Hence, an ICA algorithm which searches for sources with independent increments represents an alternative approach to the classical ICA. In order to apply this approach, one only has to modify the classical ICA algorithms marginally.

2. THEORY

Consider n time series generated by n sources $\mathbf{s} = (s_1, \ldots, s_n)$ with $(\mathbf{s}(t) \in \mathbb{R}^n)$. The states of $\mathbf{s}(t)$ are recorded at time $t = 1, 2, \ldots$. The sources should be centered ($E\{\mathbf{s}\} = 0$) and should be Markov processes of order m. Furthermore, the sources should be stochastically uncoupled, $p(\mathbf{s}(t)|\mathbf{s}(t-1), \ldots, \mathbf{s}(t-m)) = \prod_{i=1}^{n} p(s_i(t)|s_i(t-1), \ldots, s_i(t-m))$, which means, that the state $s_i(t)$, $(i = 1, \ldots, n)$ of each process s_i is only affected by the process's own past $(s_i(t-1), \ldots, s_i(t-m))$ but not by the past of the other processes $(s_j(t-1), \ldots, s_j(t-m))$, $(i \neq j)$. This is fulfilled if the joint probability $p(\mathbf{s}(t), \mathbf{s}(t-1), \ldots, \mathbf{s}(t-m))$ factorizes into the probabilities of the $(m+1)$-dimensional embedding vectors $(s_i(t), s_i(t-1), \ldots, s_i(t-m))$, $(i = 1, \ldots, n)$:

$$p(\mathbf{s}(t), \mathbf{s}(t-1), \ldots, \mathbf{s}(t-m)) = \prod_{i=1}^{n} p(s_i(t), s_i(t-1), \ldots, s_i(t-m)) \,. \quad (1)$$

From the independence of these embedding vectors the instantaneous independence of the processes follows immediately, so that one looks for both uncoupled and (instantaneous) independent sources. If Eq. (1) is fulfilled, then the mutual information of all time embedding vectors is zero as well,

$$I\left(\begin{pmatrix} s_1(t) \\ \vdots \\ s_1(t-m) \end{pmatrix}, \ldots, \begin{pmatrix} s_n(t) \\ \vdots \\ s_n(t-m) \end{pmatrix}\right) = 0. \quad (2)$$

Equation (2) is obviously an extension of the classical ICA assumption.

In the following, we concentrate on linear instantaneous mixtures $\mathbf{x}(t)$ of the sources $\mathbf{s}(t)$ with a square matrix $\mathbf{A} \in \mathrm{Mat}(n \times n; \mathbb{R})$ of full rank:

$$\mathbf{x}(t) = \mathbf{A} \cdot \mathbf{s}(t). \quad (3)$$

To simplify the calculations the signals $\mathbf{x}(t)$ can be decorrelated and normalized resulting in $\mathbf{z}(t) = \mathbf{E}\mathbf{D}^{-1/2}\mathbf{E}^\top\mathbf{x}(t)$, where $\mathbf{E}$ are the eigenvectors of $\mathbf{C}_X = E\{\mathbf{x}\mathbf{x}^\top\}$ and $\mathbf{D}$ is a diagonal matrix with the corresponding eigenvalues. Then for $\mathbf{z}(t)$ holds $E\{\mathbf{z}\mathbf{z}^\top\} = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix. This step is called whitening/sphering [13]. Since we assume a linear mixing, we want to find a demixing matrix $\mathbf{W}$, so that the demixed signals

$$\mathbf{y}(t) = \mathbf{W} \cdot \mathbf{z}(t) \quad (4)$$

are equivalent to the sources or at least as independent as possible. Taking our new definition of independent and uncoupled sources (Eq. (2)) we can rewrite the mutual information by using the definitions of entropy and the negentropy J and obtain

$$I(\mathbf{y}) = -\sum_{i=1}^{n} J\left(\begin{pmatrix} y_i(0) \\ \cdots \\ y_i(m) \end{pmatrix}\right) + \text{const}\,. \quad (5)$$

Here, we dropped the time t for easy reading and introduced the denotation $y_i(k) := y_i(t-k)$, $k = 0, 1, \ldots, m$.

Analog to the classical ICA, we can minimize the mutual information of the time embedding vectors by maximizing their negentropy. This is an advantage in contrast to other approaches, e.g. AMUSE [3, 4], SOBI [5], and TDSEP [6], as well as [8], since we do not have to assume, that the autocorrelation functions of the sources are different or that the sources are different autoregressive processes. Further, we do not have to assume non-stationary sources as it was done in [9]. Merely, the *time embedding vectors* of the sources may not be gaussian distributed, since under every orthogonal transformation the independent components of a multi-dimensional Gaussian process remain independent. However, in real world problems the distribution of the time embedding vector is usually non-gaussian.

3. ALGORITHM

Using Eq. (5) we can derive an algorithm which uses temporal structures but is not limited to a special set of sources – except of sources where the $(m+1)$-dimensional time embedding vectors are gaussian distributed, since these sources cannot be separated as mentioned above. In the following we will derive an algorithm analog to the FastICA algorithm [1] including a deflation scheme [14] using the Gram-Schmidt orthogonalization method.

As an estimate for the negentropy we use the function $J(\mathbf{w})$ from the (classical) FastICA algorithm

$$J(\mathbf{w}) = \Big[E\left\{G\left(\mathbf{w}^\top\mathbf{z}(0), \ldots, \mathbf{w}^\top\mathbf{z}(m)\right)\right\} - E\left\{G\left(\nu\right)\right\}\Big]^2 \quad (6)$$

but with a different function $G(u_0, \dots, u_m)$, in particular, $G : \mathbb{R}^{m+1} \to \mathbb{R}$. Thus, by considering the time embedding vectors the time structures are taken into account. ν is a $(m+1)$-dimensional Gaussian random variable with the same variance as the sources.

In order to maximize Eq. (6) one has to find the extremas of $E\left\{G\left(\mathbf{w}^\top \mathbf{z}(0), \dots, \mathbf{w}^\top \mathbf{z}(m)\right)\right\}$. Furthermore the demixed signals $\mathbf{y}$ shall be normalized, thus, we obtain the constraint $E\{(\mathbf{w}^\top \mathbf{z})^2\} = ||\mathbf{w}||^2 = 1$, which acts as a regulization. Using the method of the Lagrangian multipliers, by which constrained extremas can be found, we get the function

$$L(\mathbf{w}) = E\left\{G\left(\mathbf{w}^\top \mathbf{z}(0), \dots, \mathbf{w}^\top \mathbf{z}(m)\right)\right\} - \lambda(||\mathbf{w}||^2 - 1)\,. \quad (7)$$

λ is the Lagrangian multiplier. The constrained extremas are given by the zeros of the first derivative of $L(\mathbf{w})$. To solve this numerically Newton's method can be applied. For this, the first $\frac{\partial L}{\partial \mathbf{w}}$ and second derivatives $\frac{\partial^2 L}{\partial \mathbf{w}^2}$ of $L(\mathbf{w})$ are needed:

$$\frac{\partial L}{\partial \mathbf{w}} = E\left\{\sum_{k=0}^{m} \mathbf{z}(k) \frac{\partial}{\partial u_k} G\left(\mathbf{w}^\top \mathbf{z}(0), \dots, \mathbf{w}^\top \mathbf{z}(m)\right)\right\} - 2\lambda \mathbf{w} \stackrel{!}{=} 0 \quad (8)$$

$$\text{with } \lambda = \frac{1}{2} E\left\{\sum_{k=0}^{m} \mathbf{w}^\top \mathbf{z}(k) \frac{\partial}{\partial u_k} G\left(\mathbf{w}^\top \mathbf{z}(0), \dots, \mathbf{w}^\top \mathbf{z}(m)\right)\right\} \quad (9)$$

$$\begin{aligned}\frac{\partial^2 L}{\partial \mathbf{w}^2} &= E\left\{\sum_{k,l=0}^{m} \mathbf{z}(k)\mathbf{z}^\top(l) \frac{\partial^2}{\partial u_k \partial u_l} G\left(\mathbf{w}^\top \mathbf{z}(0), \dots, \mathbf{w}^\top \mathbf{z}(m)\right)\right\} - 2\lambda \mathbf{I} \\ &\approx \sum_{k,l=0}^{m} E\left\{\mathbf{z}(k)\mathbf{z}^\top(l)\right\} E\left\{\frac{\partial^2}{\partial u_k \partial u_l} G\left(\mathbf{w}^\top \mathbf{z}(0), \dots, \mathbf{w}^\top \mathbf{z}(m)\right)\right\} - 2\lambda \mathbf{I} \\ &\approx \sum_{k,l=0}^{m} E\left\{\frac{\partial^2}{\partial u_k \partial u_l} G\left(\mathbf{w}^\top \mathbf{z}(0), \dots, \mathbf{w}^\top \mathbf{z}(m)\right)\right\} \mathbf{I} - 2\lambda \mathbf{I}\end{aligned} \quad (10)$$

The stabilized Newton's method to find the zeros of a function $f(\mathbf{w})$ reads

$$\mathbf{w} \leftarrow \mathbf{w} - \mu \cdot J_f^{-1} \cdot f(\mathbf{w}) \quad (11)$$

where J_f is the Jacobian matrix of f and $\mu \in (0, 1]$ is the stabilization factor. Inserting $f = \frac{\partial L}{\partial \mathbf{w}}$ and $J_f = \frac{\partial^2 L}{\partial \mathbf{w}^2}$, we obtain the following iteration scheme for finding the unmixing vectors $\mathbf{w}_i$ using the deflation described in [14]:

$$\mathbf{w}_i \leftarrow \mathbf{w}_i - \mu \cdot \left(\frac{\partial^2 L}{\partial \mathbf{w}^2}\right)^{-1} \cdot \frac{\partial L}{\partial \mathbf{w}} \quad (12)$$

$$\mathbf{w}_i \leftarrow \mathbf{w}_i - \sum_{j=1}^{i-1} (\mathbf{w}_i^\top \mathbf{w}_j)\mathbf{w}_j \quad (13)$$

$$\mathbf{w}_i \leftarrow \frac{\mathbf{w}_i}{||\mathbf{w}_i||}. \quad (14)$$

To ensure the constraint $||\mathbf{w}||^2 = 1$, unmixing vector $\mathbf{w}_i$ is normalized at the end of the iteration step again.

Analog to the classical FastICA algorithm, in the second derivative of $L(\mathbf{w})$, Eq. (10), we made the approximation that the $(\mathbf{z}(k)\mathbf{z}^\top(l))$ are independent of the other part of the addend. Furthermore, we assumed that the expected value $E\{\mathbf{z}(k)\,\mathbf{z}^\top(l)\}$ reduces to the identity matrix if $k \neq l$ as well. If $k = l$ then $E\{\mathbf{z}(k)\,\mathbf{z}^\top(l)\} = \mathbf{I}$ holds due to whitening. Using this approximation for $\frac{\partial^2 L}{\partial \mathbf{w}^2}$ one gets good convergence while using the exact second derivative, the Newton's method has difficulties to converge. This was also observed for the embedding dimension 1 (m=0) when our algorithm is identical to the classical one. We do not have any satisfying explanation yet.

Following the proof for the FastICA algorithm [13], one can also show, that if $\mathbf{w}_j$ is a sought demixing vector for s_j, ($s_j = \mathbf{w}_j \cdot \mathbf{z}$), then $E\left\{G\left(\mathbf{w}_j^\top \mathbf{z}(0), \dots, \mathbf{w}_j^\top \mathbf{z}(m)\right)\right\}$ has a maximum/minimum under the constraint $E\{(\mathbf{w}_j^\top \mathbf{z})^2\} = ||\mathbf{w}_j||^2 = 1$ if the condition

$$\sum_{k,l=0}^{m} E\left\{s_i(k)s_i(l)\right\} \cdot E\left\{\frac{\partial^2}{\partial u_k \partial u_l} G(s_j(0), \dots, s_j(m))\right\} - \sum_{k=0}^{m} E\left\{s_j(k) \frac{\partial}{\partial u_k} G(s_j(0), \dots, s_j(m))\right\} \lessgtr 0 \quad (15)$$

holds for all $i = 1, \ldots, n$, $i \neq j$. Furthermore, each sought demixing vector $\mathbf{w}_j$ is a zero of $\frac{\partial L(\mathbf{w}_j)}{\partial \mathbf{w}} = 0$. Equation (15) represents a guideline for choosing the function G, so that one obtains a converging algorithm.

Since we want to approximate the negentropy of a $(m+1)$-dimensional random variable, we have to take the higher dimensional structure of the probability distribution into account. Using a function like $G(u_0, \ldots, u_m) = u_0^4 + \ldots + u_m^4$ would not yield any approvement because it does not consider any correlation between the time steps. As a starting point we use

$$G(u_0, \ldots, u_m) = F(u_0) + \sum_{k=1}^{m} F(u_0 - u_k) \qquad (16)$$

with $F(u) = \frac{1}{4}u^4$.

Even when the existence of an extremum for the original sources for this algorithm can be proofen, we like to mention the critical point: Due to approximations and the estimation of high dimensional functions, which is necessary for including the information from the temporal structure, the numerical stability of this algorithm is more critical than for the classical FastICA algorithm. However, the proposed algorithm works with any mixture of sources, as long as the time embedding vectors are non-gaussian and independent.

Note: One can also derive an algorithm analog to the Bell and Sejnowski algorithm [2] by optimizing the mutual information of the time embedding vectors. However, more knowledge about the $(m+1)$-dimensional source distributions is necessary. Even for simple distributions the algorithm works unstable and convergence is often not achieved.

4. APPLICATIONS

In an example, we want to demonstrate the application of our algorithm. We have mixed 4 sources with gaussian distributed states at each time step according to our mixing model (3), using a random 4x4 matrix. In particular, the source data set consists of two synthetical data sets (a sinusoidal and triangle signal), one autoregressive process of order one and a voice signal. All sources were made gaussian and each time series had a length of 10,000 samples. The first few hundred samples of the time series of the four sources are shown in Fig. 1.

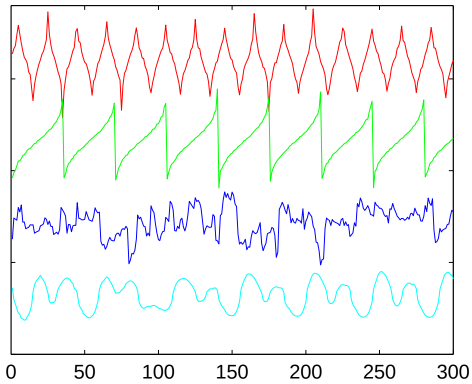

Figure 1: Time series with the first 300 samples of the four sources with gaussian distributed states at each time step.

To visualize the temporal structure of the signals, in Fig. 2 the scatter-plots of the sources ($s_i(t-1)$ over $s_i(t)$), $i = 1, \ldots, 4$ are given. All four plots show a strong deviation from a 2-dimensional gaussian distribution. In particular, the sinusoidal and the triangle signals, which is in addition asymmetric, can be clearly identified (upper two plots in Fig. 2).

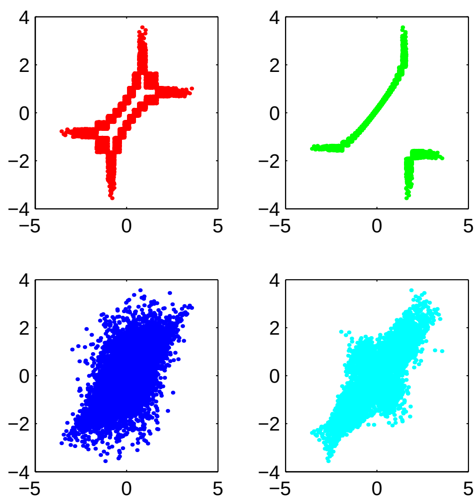

Figure 2: Scatter-plot $s_i(t-1)$ over $s_i(t)$ of the sources s_i shown in Fig. 1

When plotting the histograms of the sources $s_i(t)$ one obtains perfect gaussian distributions (see left column of Fig. 3). In contrast, the histograms of $s_i(t) - s_i(t-1)$ (right column of Fig. 3) deviate strongly from a gaussian distribution. One can interpret the histograms as a projection of the 2-dimensional distribution (Fig. 2) along the diagonal. These deviations are used in the time embedding algorithm to separate the sources.

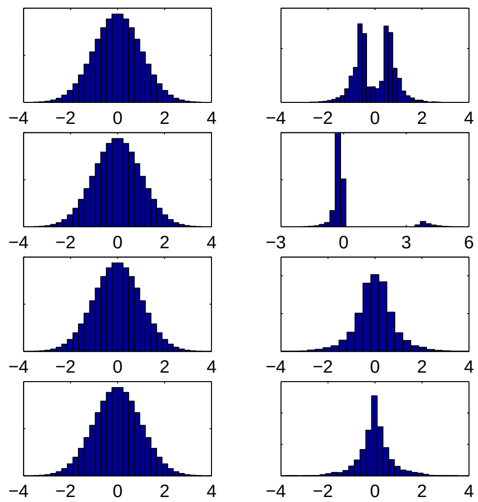

Figure 3: Histograms of the sources $s_i(t)$ (left column) and of the increments $s_i(t) - s_i(t-1)$ (right column).

Using the function given in (16) and a time embedding vector of dimension 2 (m=1), we can separate the mixed signals applying the algorithm described in the previous section. The best convergence in our example was achieved for a time delay τ of 10 samples, $(s_i(0), s_i(1)) \to (s_i(t), s_i(t-\tau))$.

To determine the correct time embedding dimension, one calculates the conditional entropy $H(x(t)|x(t-1), ..., x(t-m))$ and varies m until a plateau is reached. A misestimation of m results in ignoring longer time correlation. This problem is subject of current research.

To verify how close the estimated signals are to the original sources, we calculate the matrix $E\{\mathbf{s}\mathbf{y}^\top\}$, where $\mathbf{s}$ are the sources and $\mathbf{y}$ the estimated sources by the algorithm. Both signals are normalized to a variance of one so that this matrix has only entries in the range from -1 to 1.

A good measure for the accuracy of the estimate is the mean off-diagonal element of this matrix. Due to a finite number of samples, we obtain already a value of 0.036 when calculating $E\{\mathbf{s}\mathbf{s}^\top\}$ for the sources. For the estimated sources of our algorithm, we obtain a value of 0.049 (= mean off-diagonal element of $E\{\mathbf{s}\mathbf{y}^\top\}$), which can be seen as a good reconstruction of the sources. The signal to noise (S/N) or more accurate the signal to error ratio for our estimated sources is in the range of 200 to 1,000, which indicates also a good reconstruction.

For comparison we made calculations with the TDSEP algorithm where we obtain values of 0.021 for the mean off-diagonal element. This indicates a better performance of the algorithm, but one should realize, that TDSEP, SOBI etc. can only be used in cases with temporal correlations and different autocorrelation functions of the sources. In contrast, our algorithm can in principal deal with any independent sources. Still this algorithm should only be seen as a proof of concept to test the presented approach.

Furthermore it is clear, that classical ICA algorithms will give no results for these mixture of sources. For real world applications, e.g EEGs with 21 signals, the proposed algorithm converges and finds sources as expected by physicians.

5. A NEW CONCEPT FOR ICA – INDEPENDENT INCREMENTS

As seen in the previous sections, the independence of the time embedding vectors is very strict but also powerful. Asking for independent sources in real world application, one does not only want instantaneous independent sources, but those for which the current state does not depend on the past of the other sources. In principal, one has to take the complete past of the time series into account in order to guarantee full independence. However, if the sources are Markov processes of order m then it is sufficient to look for sources whose (m+1)-dimensional time embedding vectors are independent.

Assuming that the sources are only random variables (Markov processes of order zero), we can reduce the embedding dimension to 1. Hence, we end up at the assumption of the classical ICA ($\mathbf{x}(t)$ is independent of $\mathbf{y}(t)$). Therefore the classical ICA is a special case of the time embedding ICA.

Due to numerical difficulties when trying to find sources with independent embedding vectors of dimension $m+1$, one would like to weaken the assumption of the time embedding ICA. We want to propose a new assumption for independent sources: Assume the sources are Markov processes of first order and their time embedding vector with dimension two are independent. Then the increments $(s_i(t) - s_i(t-\tau))$ of the processes s_i, $i = 1, \ldots, n$ are also independent. On the other hand, the increments display the dynamics of the sources. This assumption weakens the time embedding ICA, but even for Markov processes with order $m > 1$ it takes the dynamics of the system in a certain degree into account.

In order to perform an ICA with the aim to seek sources with independent increments the classical FastICA algorithm can be applied on the increments $z_i(t) - z_i(t-\tau)$. In contrast to the time embedding ICA the classical FastICA algorithm is more stable. Using this algorithm to look for independent increments, it gives not only the correct results for Markov processes of order 1 but also for random variables (Markov processes of order 0). For the example above we obtain a mean off-diagonal element of 0.039 and a signal to error ratio of 1,000 to 5,000 using increments.

Performing ICA using the states of the processes or the increments of the processes depends on the question one wants to be answered. When searching for sources which show little *similarities* (synchronization), i.e. the sources shall be instantaneous independent, then the classical ICA has to be applied while ICA performed on the increments is more appropriate to seek sources with *independent dynamics*, i.e. which are uncoupled.

6. CONCLUSIONS

We have shown that in the framework of ICA temporal structure in the time series can be taken into account by using time embedding vectors. This is a generalization of the classical ICA. Instead of considering only instantaneous independent sources the sources are assumed to be uncoupled, i.e. the dynamics of the processes are independent of each other. This extension of the ICA model is very powerful and we have shown how to implement it analogously to the FastICA algorithm resulting in an algorithm which we call Fast Time Embedding ICA (FastTeICA). As long as the $(m+1)$-dimensional time embedding vectors are non-gaussian the sources can be separated with the FastTeICA algorithm.

A weakening of the strict requirement of independent time embedding vectors which still takes into account the dynamics of the processes can be achieved by assuming that the increments of the processes are independent. This is an alternative approach to the decomposition into instantaneous independent sources. For this, standard ICA algorithms can be used just by applying them on the increments of the signals. Thus, the advantage of the better stability in convergence of the standard ICA algorithms compared with the FastTeICA algorithm can be combined with the aim of seaching for dynamical independent sources.

7. ACKNOWLEDGMENTS

This project was supported by the Deutsche Forschungsgemeinschaft (DFG) in the Graduiertenkolleg "Nonlinearity and Nonequilibrium in Condensed Matter". Furthermore, we would like to thank Klaus Richter and Thomas Schreiber for fruitful discussions.

8. REFERENCES

[1] A. Hyvärinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans. on Neural Networks*, vol. 10, no. 3, pp. 626–634, 1999.

[2] A. J. Bell and T. J. Sejnowski, "An information-maximisation approach to blind separation and blind deconvolution," *Neural Computation*, vol. 7, pp. 1129–1159, 1995.

[3] L. Tong, R.-W. Liu, V.C. Soon, and Y.F. Huang, "Indeterminacy and identifiability of blind identification," *IEEE Trans. on Circuits and Systems*, vol. 38, pp. 499–509, 1991.

[4] L. Molgedey and H.G. Schuster, "Separation of a mixture of independent signals using time delayed correlations," *Physical Review Letters*, vol. 72, pp. 3634–3636, 1994.

[5] A. Belouchrani, K. Abed Meraim, J.-F. Cardoso, and E. Moulines, "A blind source separation technique based on second order statistics," *IEEE Trans. on Signal Processing*, vol. 45, no. 2, pp. 434–444, 1997.

[6] A. Ziehe and K.-R. Müller, "TDSEP – an efficient algorithm for blind separation using time structure," *Proc. Int. Conf. on Artificial Neural Networks (ICANN'98)*, pp. 675–680, 1998, Skövde, Sweden.

[7] B.A. Pearlmutter and L.C. Parra, "Maximum likelihood blind source separation: A context-sensitive generalization of ICA," *Advances in Neural Information Processing Systems*, 1997.

[8] S.-I. Amari, "Estimating functions of independent component analysis for temporally correlated signals," *Neural Computation*, vol. 12, no. 9, pp. 2083–2107, 2000.

[9] K. Matsuoka, M. Ohya, and M. Kawamoto, "A neural net for blind separation of nonstationary signals," *Neural Networks*, vol. 8, no. 3, pp. 411–419, 1995.

[10] S. Hosseini and C. Jutten, "Blind seperation of temporally correlated sources using a quasi maximum likelihood approach," *ICA2001 - Proceedings, San Diego*, 2001.

[11] T. Schreiber, "Measuring information transfer," *Phys. Rev. Lett.*, vol. 85, pp. 461, 2000.

[12] A. Kaiser and T. Schreiber, "Information transfer in continuous processes," *Physica D*, vol. 166, pp. 43–62, 2002.

[13] A. Hyvärinen and E. Oja, "Independent component analysis by general nonlinear hebbian-like learning rules," *Signal Processing*, vol. 64, no. 3, pp. 301–313, 1997.

[14] N. Delfosse and P. Loubaton, "Adaptive blind separation of independent sources: A deflation approach," *Signal Processing*, vol. 45, pp. 59–83, 1995.