# SEPARATION OF POST-NONLINEAR MIXTURES USING ACE AND TEMPORAL DECORRELATION

*Andreas Ziehe*[1*], *Motoaki Kawanabe*[1], *Stefan Harmeling*[1] *and Klaus-Robert Müller*[1,2]

[1]GMD FIRST.IDA, Kekuléstr. 7, 12489 Berlin, Germany
[2]University of Potsdam, Am Neuen Palais 10, 14469 Potsdam, Germany
{ziehe,nabe,harmeli,klaus}@first.gmd.de

## ABSTRACT

We propose an efficient method based on the concept of *maximal correlation* that reduces the post-nonlinear blind source separation problem (PNL BSS) to a linear BSS problem. For this we apply the Alternating Conditional Expectation (ACE) algorithm – a powerful technique from non-parametric statistics – to approximately invert the (post-)non-linear functions. Interestingly, in the framework of the ACE method convergence can be proven and in the PNL BSS scenario the optimal transformation found by ACE will coincide with the desired inverse functions. After the non-linearities have been removed by ACE, temporal decorrelation (TD) allows us to recover the source signals. An excellent performance underlines the validity of our approach and demonstrates the ACE-TD method on realistic examples.

## 1. INTRODUCTION

Blind source separation (BSS) research has mainly been focused on variants of linear ICA and temporal decorrelation methods (see e.g. [14, 6, 5, 7, 1, 2, 13, 29, 22, 12]). Linear BSS assumes that at time $t$ each component $x_i[t]$ of the observed $n$-dimensional data vector $\mathbf{x}[t]$ is a linear combination of $m \leq n$ statistically independent signals: $x_i[t] = \sum_{j=1}^{m} A_{ij} s_j[t]$ (e.g. [12]). The source signals $s_j[t]$ are unknown, as are the coefficients $A_{ij}$ of the mixing matrix $\mathbf{A}$. The goal is therefore to estimate both unknowns from the observed signals $\mathbf{x}[t]$, i.e. a separating matrix $\mathbf{B}$ and signals $\mathbf{u}[t] = \mathbf{B}\mathbf{x}[t]$ that estimate $\mathbf{s}[t]$.

However, non-linearities that distort the mixed signals, pose a challenging problem for "conventional" BSS methods, where the mixing model is linear instantaneous or convolutive. The general nonlinear mixing model is (cf. [12])

$$\boldsymbol{x}[t] = \boldsymbol{f}(\boldsymbol{s}[t]) \tag{1}$$

[*] To whom correspondence should be addressed.

This work was partly supported by the EU under contract IST-1999-14190 – BLISS.

where $\boldsymbol{f}$ is an arbitrary nonlinear transformation (at least approximately invertible). An important special case are post-nonlinear (PNL) mixtures

$$\boldsymbol{x}[t] = \boldsymbol{f}(\mathbf{A}\boldsymbol{s}[t]), \tag{2}$$

where $\boldsymbol{f}$ is an invertible nonlinear function that operates *componentwise* and $\mathbf{A}$ is a linear mixing matrix. Because this PNL model, which has been introduced by Taleb and Jutten [25], is an important subclass with interesting properties it attracted the interest of several researchers [25, 15, 27]. Furthermore it is often an adequate modelling of real-world physical systems, where nonlinear transfer functions appear; e.g. in the fields of telecommunications or biomedical data recording sensors can have a nonlinear characteristics.

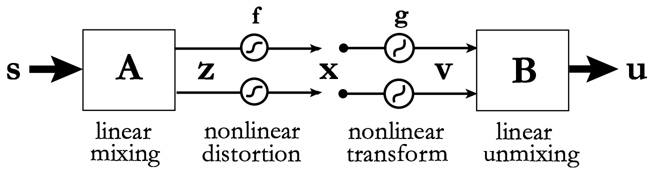



**Fig. 1**: Building blocks of the PNL mixing model and the separation system.

Algorithmic solutions of eq.(2) have used e.g. self-organizing maps [20, 18], extensions of GTM [21], neural networks [27, 19], parametric sigmoidal functions [16] or ensemble learning [26] to approximate the nonlinearity $\boldsymbol{f}$ (or its inverse $\boldsymbol{g}$). Also kernel based methods were tried on very simple toy signals [8] and more recently also on real-world data using temporal decorrelation in feature space [10]. Note, that most existing methods (except [10]) are of high computational cost and depending on the algorithm are prone to run into local minima.

In our approach to the PNL BSS problem we first approximately invert the post-nonlinearity using the ACE algorithm (estimating $\boldsymbol{g}$) and then apply a standard BSS technique [3, 29] that relies on temporal decorrelation (estimating the unmixing matrix $\mathbf{B}$) (cf. Fig.1). By virtue of the ACE framework, which is briefly introduced in subsection 2.2, we prove that the algorithm converges to the correct inverse nonlinearities – provided that they exist. Some implementation issues are discussed and numerical simulations illustrating the method are described in section 3. Finally a conclusion is given in section 4.

## 2. METHODS

For the sake of simplicity we introduce our method for the $2 \times 2$ case. The extension to the general case is easily possible, but omitted for better readability.

### 2.1. Problem statement

Let us consider the 2 dimensional post-nonlinear mixing model:

$$x_1 = f_1(a_{11}s_1 + a_{12}s_2)$$
$$x_2 = f_2(a_{21}s_1 + a_{22}s_2)$$

where $s_1$ and $s_2$ are independent source signals, that are temporally correlated, $x_1$ and $x_2$ are the observed signals, $\mathbf{A} = (a_{ij})$ is the mixing matrix and $f_1$ and $f_2$ are the componentwise nonlinear transformations which are invertible.

Obviously, any attempt to separate such a mixture by a linear BSS algorithm will fail, unless one could invert the functions $f_1$ and $f_2$ at least approximately. In this work we propose that this can be achieved by maximizing the correlation

$$\mathit{corr}(g_1(x_1), g_2(x_2)) \tag{3}$$

with respect to nonlinear functions $g_1$ and $g_2$. This means, we want to find transformations $g_1$ and $g_2$ of the observed signals such that the relationship between the transformed variables becomes linear. Intuitively speaking, the relationship is linear, if the signals are aligned in a scatterplot, i.e. if they are *maximally* correlated. Under certain conditions that we will state in detail later, this problem is solved by the ACE method that finds so called optimal transformations $g_1^*$ and $g_2^*$ which maximize eq.(3). One can prove existence and uniqueness of those optimal transformations and it can be shown that the ACE algorithm, which is described in the following, converges to these solutions (cf. [4]).

### 2.2. ACE algorithm

The ACE algorithm is an iterative procedure for finding the optimal nonlinear functions $g_1^*$ and $g_2^*$. The starting point is the observation that for fixed $g_1$ the optimal $g_2$ is given by

$$g_2(x_2) = E\{g_1(x_1)|x_2\},$$

and conversely, for fixed $g_2$ the optimal $g_1$ is

$$g_1(x_1) = E\{g_2(x_2)|x_1\}.$$

The key idea of the ACE algorithm is therefore to compute alternately the respective conditional expectations. To avoid trivial solutions one normalizes $g_1(x_1)$ in each step by using the function norm $||\cdot|| := \left(E\{\cdot\}^2\right)^{1/2}$. The algorithm for two variables is summarized below. It is also possible to extend the procedure to the multivariate case, however, for further details we refer to [11, 4].

**Algorithm 1** The ACE algorithm for two variables

$\{$initialize$\}$
$g_1^{(0)}(x_1) \leftarrow x_1/||x_1||$
**repeat**
$g_2^{(k+1)}(x_2) \leftarrow E\{g_1^{(k)}(x_1) \mid x_2\}$
$\tilde{g}_1^{(k+1)}(x_1) \leftarrow E\{g_2^{(k+1)}(x_2) \mid x_1\}$
$g_1^{(k+1)}(x_1) \leftarrow \tilde{g}_1^{(k+1)}(x_1)/||\tilde{g}_1^{(k+1)}(x_1)||$
**until** $E\{g_1(x_1) - g_2(x_2)\}^2$ fails to decrease

An important point in the implementation of this algorithm is the estimation of the conditional expectations from the data. Usually, the conditional expectations are computed by data smoothing for which numerous techniques exist (cf. [4, 9]). Care has to be taken to balance the trade-off between the fidelity to the data against the smoothness of the estimated curve. Our implementation utilizes a nearest neighbor smoothing that applies a simple moving average filter to appropriately sorted data.

By applying $g_1^*$ and $g_2^*$ to the mixed signals $x_1$ and $x_2$ we remove the effect of the nonlinear functions $f_1$ and $f_2$. In the following we will substantiate this claim more formally. We show for $z_1 = a_{11}s_1 + a_{12}s_2$ and $z_2 = a_{21}s_1 + a_{22}s_2$ that $g_1^*$ and $g_2^*$ obtained from the ACE procedure are the desired inverse functions for the case that $z_1$ and $z_2$ are jointly normal distributed, with other words we prove the following relationship:

$$\begin{aligned} h_1^*(z_1) &:= g_1^*(f_1(z_1)) \propto z_1 \\ h_2^*(z_2) &:= g_2^*(f_2(z_2)) \propto z_2. \end{aligned} \tag{4}$$

Almost all work for the proof has already been done in *Proposition 5.4.* and *Theorem 5.3.* of [4] which – by noticing that the correlation of two signals does not change, if we scale one or both signals – implies:

$$g_1^*(x_1) \propto E\{g_2^*(x_2)|x_1\}$$
$$g_2^*(x_2) \propto E\{g_1^*(x_1)|x_2\}.$$

Note that the conditional expectation $E\{g_2^*(x_2)|x_1\}$ is a function of $x_1$ and the expectation is taken with respect to $x_2$, analogously for the second expression.
Since $g_1^*(x_1) = h_1^*(z_1)$ and $g_2^*(x_2) = h_2^*(z_2)$, furthermore $x_1 = f_1(z_1)$ and $x_2 = f_2(z_2)$ we get:

$$h_1^*(z_1) \propto E\{h_2^*(z_2)|f_1(z_1)\}$$
$$h_2^*(z_2) \propto E\{h_1^*(z_1)|f_2(z_2)\}.$$

Because $f_1$ and $f_2$ are invertible functions they can be omitted in the condition of the conditional expectation, leading us to:

$$\begin{aligned} h_1^*(z_1) &\propto E\{h_2^*(z_2)|z_1\} \\ h_2^*(z_2) &\propto E\{h_1^*(z_1)|z_2\}. \end{aligned} \tag{5}$$

Assuming that the vector $(z_1, z_2)^\top$ is normally distributed and the correlation *corr*$(z_1, z_2)$ does not vanish, a straightforward calculation shows

$$\begin{aligned} E\{z_2|z_1\} &\propto z_1 \\ E\{z_1|z_2\} &\propto z_2. \end{aligned}$$

This means that $z_1$ and $z_2$ satisfy eq. (5), which then immediately implies our claim eq. (4). Fortunately, in our application the above assumptions are usually fulfilled because mixed signals are more Gaussian and more correlated than unmixed signals. On the other hand, even if the assumptions are not perfectly met, experiments show that the ACE algorithm still equalizes the nonlinearities well.

Summarizing the key idea, by searching for nonlinear transformations, that *maximize* the *linear* correlations between the *non-linearly* transformed observed variables, we can approximate the inverses of the post-nonlinearities.

### 2.3. Source separation

For a separation of the signals one could in principle apply any BSS technique, capable of solving the now approximately linear problem. However, experiments show that only second-order methods which use temporal information are sufficiently robust to reliably recover the sources. Therefore we use TDSEP, an implementation based on the simultaneous diagonalization of several time-delayed correlation matrices for the blind identification of the unmixing matrix **B** (cf. [3, 29, 28]).

## 3. NUMERICAL SIMULATIONS

To demonstrate the performance of the proposed method we apply our algorithm to several post-nonlinear mixtures, both instantaneous and convolutive.

The first data set consists of Gaussian AR-processes of the form:

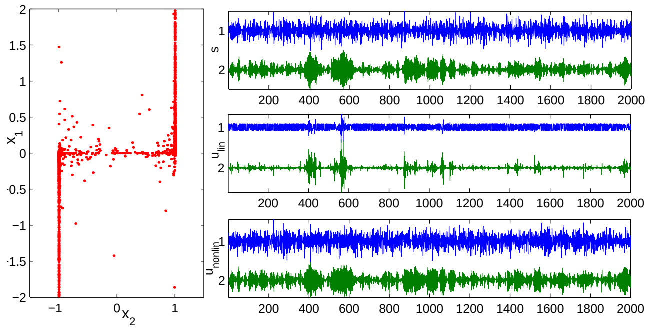
$$s_i = \sum_{m=1}^{M} a_m s_{i-m} + \xi_i, \qquad i = 1, \ldots, n, \tag{6}$$

where $\xi_i$ is white Gaussian noise with mean zero and variance $\sigma^2$. For the experiment we choose $\sigma^2 = 1$, $M = 3$, $n = 2$ and generate 2000 data points.

We use a $2 \times 2$ mixing matrix to get linearly mixed signals **z** and apply strong nonlinear distortions

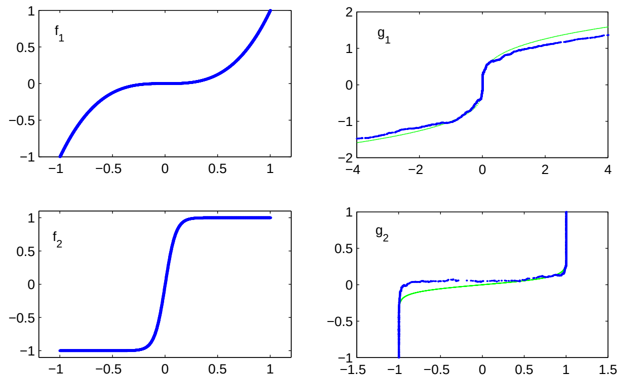
$$\begin{aligned} x_1[t] &= f_1(z_1[t]) = z_1^3[t], \\ x_2[t] &= f_2(z_2[t]) = \tanh(10 z_2[t]), \end{aligned} \tag{7}$$

which were also used by Taleb and Jutten in [24]. The distribution of these mixed signals has a highly nonlinear structure as visible in the scatter plot in Fig. 2.




(a) (b)

**Fig. 2**: (a) Scatter plot of the mixed AR-processes ($x_1[t]$ vs $x_2[t]$) and (b) waveforms of the original sources (top), the linearly unmixed signals (middle) and recovered sources (bottom).




(a) (b)

**Fig. 3**: (a) Nonlinear functions $f_1$ and $f_2$. (b) True (thin line) and estimated (bold line) inverse functions $g_1$ and $g_2$.

The application of the ACE algorithm – using a local nearest neighbor smoother (window length 31) for the conditional expectation – yields the estimated nonlinear functions $g_1$ and $g_2$ shown in Fig. 3. We see that the true inverses of the nonlinearities $f_1$ and $f_2$ are approximated well. Although the match is not perfect (could be optimized by better smoothers) it is now possible to separate the signals using the TDSEP algorithm, where 20 time-delayed correlation matrices are simultaneously diagonalized (time lags $\tau = 0..20$). Figure 2 (b) shows that the waveforms of the recovered sources closely resemble the original ones, while the result of the linear unmixing of the PNL mixture can clearly not recover the sources. This is also confirmed by comparing the output distributions that are shown in Fig. 4 as a scatter plot.

One favorable property of our method is its nice scaling behavior. To show this, we will now test the algorithm with

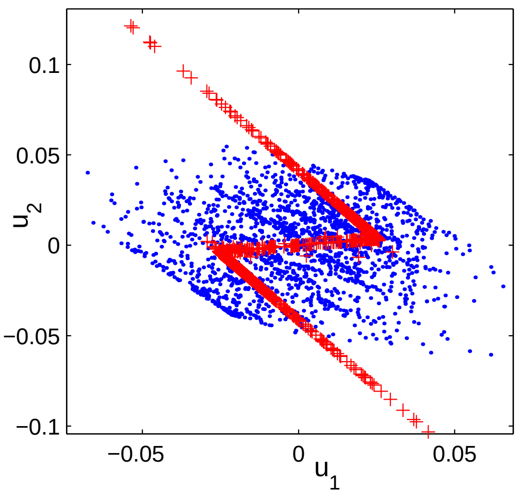



**Fig. 4**: Scatter plot of the output distribution of a linear ('+') and the proposed nonlinear ACE-TD algorithm ('.').

natural audio sources, where the input data set consists of 4 sound signals with $20,000$ data points each. For this case we apply the multivariate version of the ACE algorithm, which computes the optimal functions by maximizing the generalized correlation criterion *corr* $(g_1(x_1), \sum_{k=2}^{n} g_k(x_k))$. For details of the implementation we refer to [4, 11, 9]. As in the first experiment, these source signals were mixed by a linear model $\mathbf{z}[t] = \mathbf{A}\mathbf{s}[t]$, with a random $(4 \times 4)$ matrix $\mathbf{A}$. After the linear mixing the following nonlinearities were applied:

$$\begin{aligned} f_1(z_1) &= z_1 + 0.1 * z_1^3 \\ f_2(z_2) &= 0.3z_2 + tanh(3z_2) \\ f_3(z_3) &= tanh(2z_3) \\ f_4(z_4) &= z_4^3. \end{aligned} \quad (8)$$

Figure 5 shows the results of the separation using ACE-TD (smoothing window length 51) and TDSEP ($\tau = 0..30$). We observe again a very good separation performance that is quantified by calculating the correlation coefficients (shown in table 1) between the source signals and the extracted components. This is also confirmed by listening to the separated audio signals, were we perceive almost no crosstalk, although the noise level is slightly increased (cf. the silent parts of signal 2 in Fig. 5).

The third experiment gives an example for the application of our method to convolutive mixtures with a PNL distortion. We deliberately distorted real-room recordings[1] of speech and background music made by Lee [17] with nonlinear transfer functions as in our first example (cf. eq.(7)). For the separation we apply a convolutive BSS algorithm of Parra et al. that requires only second-order statistics by

[1] Available on the internet via
`http://sloan.salk.edu/~tewon/Blind/blind_audio.html`

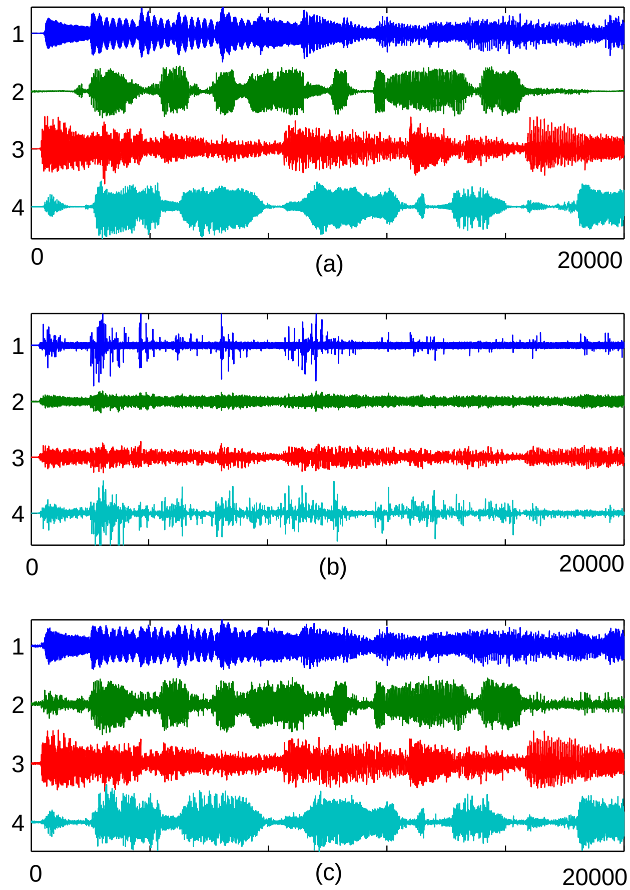



**Fig. 5**: Four channel audio dataset: (a) waveforms of the original sources, (b) linearly unmixed signals with TDSEP and (c) recovered sources using ACE-TD.

exploiting the non-stationarity of the signals [23]. While an unmixing of the distorted recordings obviously fails, we could achieve a good separation after the unsupervised linearization by the ACE procedure (cf. Fig. 6).

## 4. DISCUSSION AND CONCLUSION

In this work we proposed a simple technique for the blind separation of linear mixtures with a post-nonlinear distortion. The main ingredients of our algorithm, which we call ACE-TD, are: first, a search for nonlinear transformations that *maximize* the linear correlations between transformed variables and which approximate the inverses of the PNLs. This search can be done highly efficient by the ACE technique [4] from non-parametric statistics, that performs an alternating estimation of conditional expectations by smoothing of scatter plots. Effectively, this nonlinear modeling procedure solves the PNL mixture problem by transforming it back into a linear one. Therefore, second, a temporal decorrelation BSS algorithm (e.g. [3, 29]) can be applied.

| | TDSEP | | | |
|---|---|---|---|---|
| | $u_1$ | $u_2$ | $u_3$ | $u_4$ |
| $s_1$ | 0.10 | 0.56 | 0.31 | -0.13 |
| $s_2$ | -0.01 | 0.26 | 0.02 | 0.47 |
| $s_3$ | 0.06 | 0.12 | 0.76 | -0.05 |
| $s_4$ | -0.07 | 0.66 | -0.21 | 0.11 |
| | ACE-TD | | | |
| $s_1$ | 0.97 | -.01 | -.005 | 0.03 |
| $s_2$ | 0.03 | 0.94 | -0.02 | -0.005 |
| $s_3$ | 0.01 | 0.07 | 0.95 | -0.007 |
| $s_4$ | 0.04 | 0.002 | 0.001 | 0.96 |

**Table 1**: Correlation coefficients for the signals shown in Fig. 5

Clearly, ACE is not limited to the $2 \times 2$ case but it scales naturally to the $n \times n$ case for which an algorithmic description can be found in [4, 9]. Moreover, the algorithm can make beneficial use of additional sensors in the overdetermined BSS case as then the joint distribution of $\mathbf{z}[t]$ becomes more and more Gaussian, which is beneficial for ACE. Furthermore, our method works also for convolutive mixtures, which is attractive for real-room BSS, where nonlinear transfer functions of the sensors (microphones) or amplifiers would impede a proper separation. Concluding, the proposed framework gives a simple algorithm of *high* efficiency with a solid theoretical background for signal separation in applications with a PNL distortion, that are of importance e.g. in real-world sensor technology.

Future research will be concerned with a better tuning of the smoothers which are essential in the ACE algorithm to the PNL blind source separation scenario.

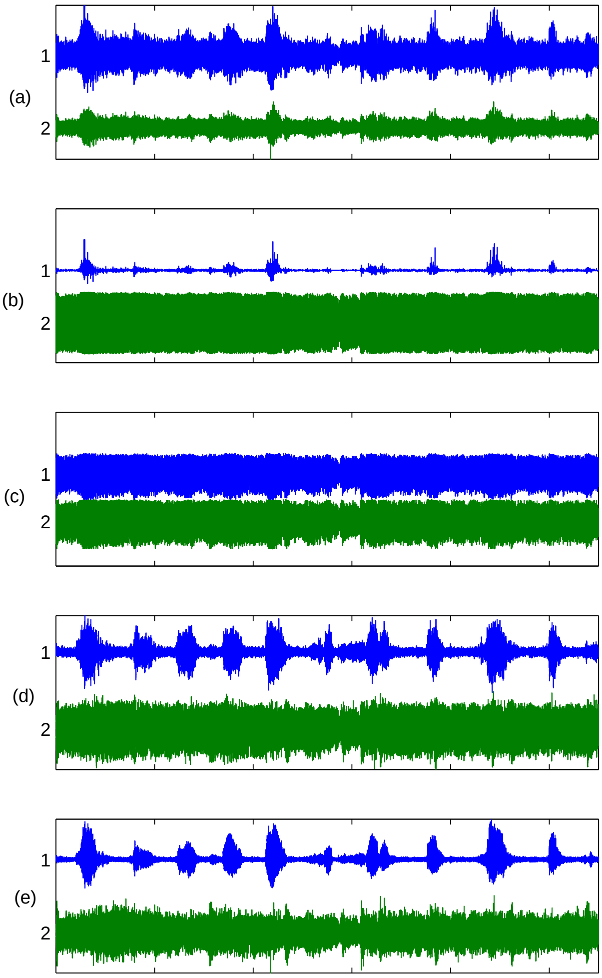



**Fig. 6**: Two channel auido dataset: (a) waveforms of the recorded (undistorted) microphone signals, (b) observed PNL distorted signals, (c) result of ACE, (d) recovered sources using ACE and a convolutive BBS algorithm and (e) for comparison convolutive BSS separation result for undistorted signals from (a).

## 5. REFERENCES

[1] S.-I. Amari, A. Cichocki, and H.H. Yang. A new learning algorithm for blind source separation. In *Advances in Neural Information Processing Systems 8*, pages 757–763. MIT Press, 1996.

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

[3] 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*, 45(2):434–444, 1997.

[4] L. Breiman and J. H. Friedman. Estimating optimal transformations for multiple regression and correlation. *Journal of the American Statistical Association*, 80(391):580–598, September 1985.

[5] J.-F. Cardoso and A. Souloumiac. Blind beamforming for non Gaussian signals. *IEE Proceedings-F*, 140(6):362–370, 1993.

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

[7] G. Deco and D. Obradovic. Linear redundancy reduction learning. *Neural Networks*, 8(5):751–755, 1995.

[8] C. Fyfe and P. L. Lai. ICA using kernel canonical correlation analysis. In *Proc. Int. Workshop on Independent Component Analysis and Blind Signal Separation (ICA2000)*, pages 279–284, Helsinki, Finland, 2000.

[9] W. Härdle. *Applied Nonparametric Regression*. Cambridge University Press, Cambridge, 1990.

[10] S. Harmeling, A. Ziehe, M. Kawanabe, B. Blankertz, and K.-R. Müller. Nonlinear blind source separation using kernel feature spaces. submitted to ICA 2001.

[11] T.J. Hastie and R.J. Tibshirani. *Generalized Additive Models*, volume 43 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London, 1990.

[12] A. Hyvärinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. Wiley, 2001.

[13] A. Hyvärinen and E. Oja. A fast fixed-point algorithm for independent component analysis. *Neural Computation*, 9(7):1483–1492, 1997.

[14] C. Jutten and J. Hérault. Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture. *Signal Processing*, 24:1–10, 1991.

[15] T.-W. Lee, B.U. Koehler, and R. Orglmeister. Blind source separation of nonlinear mixing models. In *Neural Networks for Signal Processing VII*, pages 406–415. IEEE Press, 1997.

[16] T-W. Lee, B.U. Koehler, and R. Orglmeister. Blind source separation of nonlinear mixing models. *IEEE International Workshop on Neural Networks for Signal Processing*, pages 406–415, September 1997.

[17] T-W. Lee, A. Ziehe, R. Orglmeister, and T. J. Sejnowski. Combining time-delayed decorrelation and ICA: Towards solving the cocktail party problem. In *Proc. ICASSP98*, volume 2, pages 1249–1252, Seattle, 1998.

[18] J. K. Lin, D. G. Grier, and J. D. Cowan. Faithful representation of separable distributions. *Neural Computation*, 9(6):1305–1320, 1997.

[19] G. Marques and L. Almeida. Separation of nonlinear mixtures using pattern repulsion. In *Proc. Int. Workshop on Independent Component Analysis and Signal Separation (ICA'99)*, pages 277–282, Aussois, France, 1999.

[20] P. Pajunen, A. Hyvärinen, and J. Karhunen. Nonlinear blind source separation by self-organizing maps. In *Proc. Int. Conf. on Neural Information Processing*, pages 1207–1210, Hong Kong, 1996.

[21] P. Pajunen and J. Karhunen. A maximum likelihood approach to nonlinear blind source separation. In *Proceedings of the 1997 Int. Conf. on Artificial Neural Networks (ICANN'97)*, pages 541–546, Lausanne, Switzerland, 1997.

[22] P. Pajunen and J. Karhunen, editors. *Proc. of the 2nd Int. Workshop on Independent Component Analysis and Blind Signal Separation, Helsinki, Finland, June 19-22, 2000*. Otamedia, 2000.

[23] L. Parra and C. Spence. Convolutive blind source separation of non-stationary sources. *IEEE Trans. on Speech and Audio Processing*, 8(3):320–327, May 2000. US Patent US6167417.

[24] A. Taleb and C. Jutten. Batch algorithm for source separation in post-nonlinear mixtures. In *Proc. First Int. Workshop on Independent Component Analysis and Signal Separation (ICA'99)*, pages 155–160, Aussois, France, 1999.

[25] A. Taleb and C. Jutten. Source separation in post-nonlinear mixtures. *IEEE Trans. on Signal Processing*, 47(10):2807–2820, 1999.

[26] H. Valpola, X. Giannakopoulos, A. Honkela, and J. Karhunen. Nonlinear independent component analysis using ensemble learning: Experiments and discussion. In *Proc. Int. Workshop on Independent Component Analysis and Blind Signal Separation (ICA2000)*, pages 351–356, Helsinki, Finland, 2000.

[27] H. H. Yang, S. Amari, and A. Cichocki. Information-theoretic approach to blind separation of sources in non-linear mixture. *Signal Processing*, 64(3):291–300, 1998.

[28] A. Yeredor. Blind separation of gaussian sources via second-order statistics with asymptotically optimal weighting. *IEEE Signal Processing Letters*, 7(7):197–200, 2000.

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