Independent Component Analysis and Blind Signal Separation ICA 2000, June 19–22, 2000, Helsinki, Finland, pp. 221–226.

BLIND SEPARATION OF MIXED-KURTOSIS SIGNALS USING AN ADAPTIVE THRESHOLD NONLINEARITY

Heinz Mathis, Thomas P. von Hoff, and Marcel Joho

Signal and Information Processing Laboratory,
Swiss Federal Institute of Technology Zurich, Switzerland
mathis@isi.ee.ethz.ch, *vonhoff@isi.ee.ethz.ch*, *joho@isi.ee.ethz.ch*

ABSTRACT

A parameterized threshold nonlinearity, which separates a mixture of signals with any distribution (except for Gaussian), is introduced. This nonlinearity is particularly simple to implement, since it neither uses hyperbolic nor polynomial functions, unlike most nonlinearities used for blind separation. For some specific distributions, the stable region of the threshold parameter is derived, and optimal values for best separation performance are given. If the threshold parameter is made adaptive during the separation process, the successful separation of signals whose distribution is unknown is demonstrated and compared against other known methods.

1. INTRODUCTION

Blind signal separation using higher-order statistics either explicitly or implicitly has attracted many researchers whose main goal is to separate a set of mixed signals as fast as possible with the smallest residual mixing. Most approaches require complete or at least some knowledge of the source distributions. If sources of different distributions are mixed, such techniques may fail to work.

Throughout this paper we assume a linear mixing and separation process, where the measured signals $\mathbf{x} = [x_1, \ldots, x_{M_\mathrm{s}}]^T$ to be processed are linear combinations of the original source signals $\mathbf{s} = [s_1, \ldots, s_{M_\mathrm{s}}]^T$, weighted by scalars, which are the elements of the mixing matrix $\mathbf{A}$. M_s denotes the number of sources as well as the number of sensors. Recovery of the signals is carried out by a blind adaptive algorithm adjusting the coefficients of the separation matrix $\mathbf{W}$. The output of the algorithm is therefore

$$\mathbf{u} = [u_1, \ldots, u_{M_\mathrm{s}}]^T = \mathbf{W}\mathbf{x} = \mathbf{W}\mathbf{A}\mathbf{s} = \mathbf{P}\mathbf{s}. \tag{1}$$

In order to successfully separate the signals, $\mathbf{P} = \mathbf{W}\mathbf{A}$ should approximate as closely as possible a scaled permutation matrix. A possible update equation for the separation matrix $\mathbf{W}$ results from either the minimization of the mutual information of the output signals [1], the output entropy maximization [2], the ML estimator [3], the maximum negentropy [4], and applying the natural gradient [1] to these methods, or exchanging a non-blind criterion of an RLS-like algorithm with a blind criterion [5],

$$\mathbf{W}_{t+1} = \mathbf{W}_t + \mu \left(\mathbf{I} - \mathbf{g}(\mathbf{u})\mathbf{u}^T\right) \mathbf{W}_t \tag{2}$$

where μ is the step size, $\mathbf{I}$ the identity matrix, and $\mathbf{g}(\mathbf{u})$ a *Bussgang nonlinearity* [6].

2. THE THRESHOLD NONLINEARITY

The nonlinearity plays a central role in blind signal separation. Its nature is defined by the objective or contrast function, which is often some kind of information-theoretic measure, such as entropy or mutual information. Very frequently, the nonlinearities derived by different methods are similar in nature for a given probability distribution of the signals to separate. In fact, the exact curve of the nonlinearity might not matter [7]. Whereas the minimization of the mutual information leads to a pdf-independent polynomial with several terms [1], both the Infomax and the Maximum-Likelihood approach [8] lead to

$$g(u_i) = -\frac{\partial \log p_S(u_i)}{\partial u_i} = -\frac{p_S'(u_i)}{p_S(u_i)}, \qquad i = 1, \ldots, M_\mathrm{s} \tag{3}$$

where $p_S(u_i)$ and $p_S'(u_i)$ are the pdf and its derivative, respectively, of the source signals. In the following, the range of i will be assumed as that given in (3) if not indicated otherwise. (3) is referred to as the score function of a certain pdf $p_S(.)$. From (3) it can be seen, that super-Gaussian signals typically have sigmoidal nonlinearities such as $\mathrm{sign}(.)$ or $\tanh(.)$, whereas sub-Gaussian signals can be separated using a nonlinearity of the form

$$g(u_i) = a \cdot u_i^p, \quad p \text{ odd}, \, p \geq 3. \tag{4}$$

p being odd ensures the validity of the sign after the nonlinearity. If (4) is rewritten as

$$g(u_i) = a \cdot |u_i|^{p-1} u_i, \quad p > 1 \tag{5}$$

p is no longer restricted to odd integers, but can be any rational number greater than one. (5) also has the advantage that it is directly applicable to complex signals. A very simple nonlinearity for the separation of sub-Gaussian signals has been derived in [9] starting from the pdf of a generalized Gaussian signal

$$p_S(s) = \frac{\alpha}{2\beta\Gamma\left(\frac{1}{\alpha}\right)} e^{-\left(\frac{|s|}{\beta}\right)^\alpha} \tag{6}$$

where $\alpha > 2$ for sub-Gaussian signals. The nonlinearity according to (3) is

$$g(u_i) = \alpha \left(\frac{\Gamma\left(\frac{3}{\alpha}\right)}{\Gamma\left(\frac{1}{\alpha}\right)}\right)^{\alpha/2} \mathrm{sign}(u_i) \cdot |u_i|^{\alpha-1}. \tag{7}$$

$\Gamma(.)$ is the gamma function given by $\Gamma(a) = \int_0^\infty x^{a-1} \exp(-x) dx$, and shows a recursive property similar to the factorial function, $\Gamma(a+1) = a\Gamma(a)$. For large α, (7)

yields

$$g(u_i)\Big|_{\alpha\gg 1} \approx \alpha \left(\frac{\sin\left(\frac{\pi}{\alpha}\right)}{\sin\left(\frac{3\pi}{\alpha}\right)}\right)^{\alpha/2} \operatorname{sign}(u_i)\cdot |u_i|^{\alpha-1}. \quad (8)$$

However, for large α the sine functions are well represented by the first term of their Taylor expansions, so (8) becomes

$$g(u_i)\Big|_{\alpha\gg 1} \approx \alpha \left(\frac{1}{3}\right)^{\alpha/2} \operatorname{sign}(u_i)\cdot |u_i|^{\alpha-1} = \alpha\frac{1}{u_i}\left(\frac{u_i^2}{3}\right)^{\alpha/2}. \quad (9)$$

In the limit for $\alpha \to \infty$ (uniform distribution) we get the *threshold nonlinearity*

$$\lim_{\alpha\to\infty} g(u_i) = \begin{cases} 0, & |u_i| < \sqrt{3} \\ \infty\cdot\operatorname{sign}(u_i), & |u_i| \geq \sqrt{3}. \end{cases} \quad (10)$$

A more detailed derivation of Eqs. (7) to (10) can be found in the appendix. The infinite gain in (10) will of course cause convergence problems for a finite step size. The gain can therefore be traded off against a lower threshold ϑ for a specified output power. To this end, we make use of the scaling equation [9]

$$\int_{-\infty}^{\infty} p_S(u_i) g(u_i) u_i du_i = 1 \quad (11)$$

which is a result of the Bussgang property. $p_S(.)$ is a source distribution with unit variance $\sigma_S^2 = 1$. Solving (11) for the uniform distribution and a given threshold ϑ results in a finite gain of

$$a = \frac{2\sqrt{3}}{3-\vartheta^2} \quad (12)$$

for $0 \leq \vartheta < \sqrt{3}$. The resulting threshold nonlinearity is

$$g(u_i) = \begin{cases} 0, & |u_i| < \vartheta \\ a\operatorname{sign}(u_i), & |u_i| \geq \vartheta \end{cases} \quad (13)$$

and is depicted in Fig. 1. Note that a is always positive for the assigned range of ϑ.

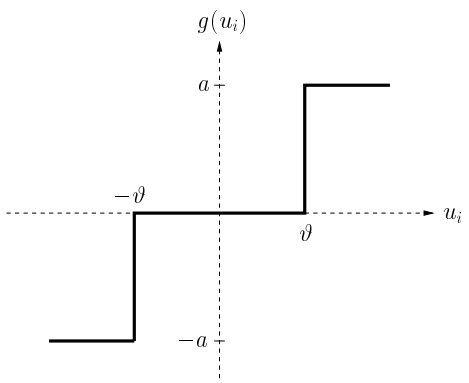

Figure 1: Threshold nonlinearity with parameter ϑ.

In [9], the application of (13) to sub-Gaussian pdfs other than the uniform distribution (e.g. M-PAM or M-QAM) was demonstrated with the possible need for adjusting gain a for normalized output power. By experiment, a good value for the threshold was found to be $\vartheta = 1.5$. On the other hand, for the Laplacian distribution, which is an example of a super-Gaussian distribution and can be written in the form of Eq. (6) with $\alpha = 1$, Eq. (7) simplifies to

$$g(u_i) = \sqrt{2}\operatorname{sign}(u_i) \quad (14)$$

which is the same as (13) for $a = \sqrt{2}$ and $\vartheta = 0$. The signum nonlinearity can hence be regarded as a threshold nonlinearity with threshold ϑ set to zero.

3. STABILITY ANALYSIS

In [10] it was shown that a necessary and sufficient stability criterion for the separation of the ith and jth source is that the eigenvalues of the Hessian sub-matrix

$$\mathbf{\Xi} = \begin{bmatrix} E\left\{g_i'(s_i)\right\} E\left\{s_j^2\right\} & E\left\{g_i(s_i)s_i\right\} \\ E\left\{g_j(s_j)s_j\right\} & E\left\{g_j'(s_j)\right\} E\left\{s_i^2\right\} \end{bmatrix} \quad (15)$$

are positive. For equal source distributions and nonlinearities, the eigenvalues of $\mathbf{\Xi}$ are given by

$$\kappa_+ = E\left\{g_i'(s_i)\right\} E\left\{s_i^2\right\} + E\left\{g_i(s_i)s_i\right\} \quad (16)$$

$$\kappa_- = E\left\{g_i'(s_i)\right\} E\left\{s_i^2\right\} - E\left\{g_i(s_i)s_i\right\}. \quad (17)$$

Note that this analysis concerns local stability, hence the statistics of $\mathbf{s}$ and $\mathbf{u}$ are interchangeable near convergence. Similar analyses have been carried out in [11] and [12]. The source power is assumed to be normalized to one, i.e., $E\left\{s_i^2\right\} = 1$, and we scale the nonlinearity such that $E\left\{g_i(s_i)s_i\right\} = 1$. Although the threshold nonlinearity is not differentiable at $s_i = \pm\vartheta$, we can derive $E\left\{g_i'(s_i)\right\}$ by the use of δ-functions and assuming a symmetric distribution

$$\begin{aligned} E\left\{g_i'(u_i)\right\} &= \int_{-\infty}^{\infty} p_S(u_i) g_i'(u_i) du_i \\ &= \int_{-\infty}^{\infty} p_S(u_i) a(\delta(u_i+\vartheta) + \delta(u_i-\vartheta)) du_i \\ &= 2a\cdot p_S(\vartheta). \end{aligned} \quad (18)$$

(16) and (17) can therefore be written as

$$\kappa_+ = 2a\cdot p_S(\vartheta) + 1 \quad (19)$$

$$\kappa_- = 2a\cdot p_S(\vartheta) - 1. \quad (20)$$

For a positive scaling factor a, (19) is always positive. To make (20) positive, we must ensure that $2a\cdot p_{S_i}(\vartheta) > 1$ by suitable choice of ϑ. For the threshold nonlinearity and symmetric distributions, Eq. (11) can be written as

$$2a\int_{\vartheta}^{\infty} p_S(u_i) u_i du_i = 1 \quad (21)$$

or, if solved for the scaling factor,

$$a = \frac{1}{2\int_{\vartheta}^{\infty} p_S(u_i) u_i du_i}. \quad (22)$$

Thus, the stability condition results in

$$\frac{p_S(\vartheta)}{\int_{\vartheta}^{\infty} p_S(u_i) u_i du_i} > 1. \quad (23)$$

Eq. (23) defines a stable region for ϑ depending on the source distribution. In order to find the optimal values for ϑ (in the sense of quality of separation), we have to minimize the term $\gamma_+/\kappa_+ + \gamma_-/\kappa_-$ [10], with κ_+ and κ_- defined by Eqs. (16) and (17), respectively, and

$$\gamma_+ = E\left\{g_i^2(s_i)\right\} E\left\{s_i^2\right\} + \left(E\left\{g_i(s_i)s_i\right\}\right)^2 \tag{24}$$

$$\gamma_- = E\left\{g_i^2(s_i)\right\} E\left\{s_i^2\right\} - \left(E\left\{g_i(s_i)s_i\right\}\right)^2. \tag{25}$$

For the threshold nonlinearity we can write

$$E\left\{g_i^2(s_i)\right\} = 2a^2 \int_\vartheta^\infty p_S(u_i)du_i. \tag{26}$$

We then get

$$\begin{aligned}\gamma_+/\kappa_+ + \gamma_-/\kappa_- &= \frac{2a^2\int_\vartheta^\infty p_S(u_i)du_i + 1}{2ap_S(\vartheta)+1} \\ &\quad + \frac{2a^2\int_\vartheta^\infty p_S(u_i)du_i - 1}{2ap_S(\vartheta)-1} \\ &= \frac{8a^3 p_S(\vartheta)\int_\vartheta^\infty p_S(u_i)du_i - 2}{4a^2p_S^2(\vartheta)-1}\end{aligned} \tag{27}$$

so the optimal value for the threshold is

$$\vartheta_{\text{opt}} = \arg\min_\vartheta \frac{4a^3 p_S(\vartheta)\int_\vartheta^\infty p_S(u_i)du_i - 1}{4a^2p_S^2(\vartheta)-1}. \tag{28}$$

Fig. 2 shows the stability region and the optimal values of ϑ of the threshold nonlinearity for the generalized Gaussian distribution. For signals with $\alpha \leq 1$, the stable range for ϑ is roughly between 0 and 0.5 with an optimal value of 0. For this threshold value, the threshold nonlinearity is the true score function for the Laplace distribution. Interestingly, the optimal value for $1 < \alpha < 2$ is slightly higher than 0. At the other extreme is the uniform distribution with $\alpha = \infty$. While the upper limit of the stability range is $\sqrt{3}$, the lower limit is 1. The optimal value approaches $\sqrt{3}$ as the pdf gets close to the uniform distribution. Note that only for the uniform distribution the threshold nonlinearity is the score function (except for the finite gain).

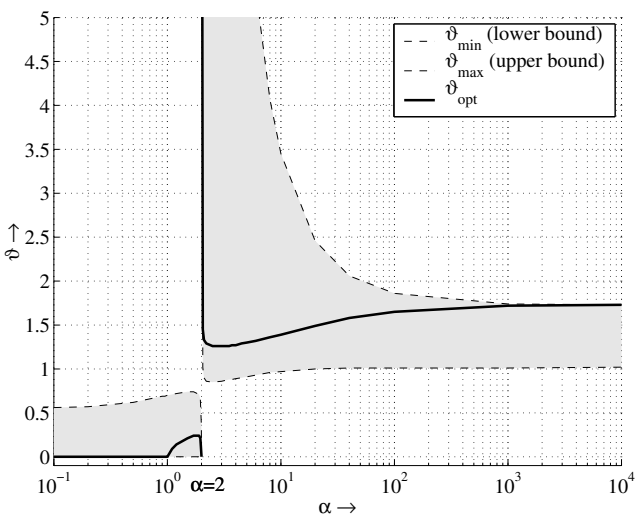

Figure 2: Range of threshold ϑ (shaded area) as a function of the generalized Gaussian parameter α.

Generally speaking, for sub-Gaussian signals, the stable range is between 1 and $\sqrt{3}$. By decreasing α we approach the normal distribution ($\alpha = 2$), which clearly poses a singularity in the stability plot. This can be seen by the broken stability regions around $\alpha = 2$.

In the following we take a closer look at two distributions, specifically the Laplacian and the uniform distribution as representatives of super- and sub-Gaussian distributions, respectively, and work out the optimal values for the threshold ϑ. For the Laplacian distribution, which is given by (6) with $\alpha = 1$, or, for unit variance

$$p_S(s) = \frac{1}{\sqrt{2}} e^{-\sqrt{2}|s|} \tag{29}$$

we get with Eq. (22) for the gain of the threshold nonlinearity

$$a = \frac{1}{\left(\vartheta + \frac{\sqrt{2}}{2}\right) e^{-\sqrt{2}\vartheta}}. \tag{30}$$

Using (30) in (28) we find after some calculations the optimal threshold as

$$\vartheta_{\text{opt}} = \arg\min_\vartheta \frac{\sqrt{2}e^{\sqrt{2}\vartheta} - \left(\vartheta + \frac{\sqrt{2}}{2}\right)^3}{2\left(\vartheta + \frac{\sqrt{2}}{2}\right) - \left(\vartheta + \frac{\sqrt{2}}{2}\right)^3} \tag{31}$$

which by numerical inspection reveals $\vartheta_{\text{opt}} = 0$. A similar calculation can be carried out for the uniform distribution, given by

$$p_S(x) = \begin{cases} \frac{1}{2\sqrt{3}}, & |x| \leq \sqrt{3} \\ 0, & \text{otherwise} \end{cases} \tag{32}$$

where the scaling factor a has already been computed by Eq. (12). Eq. (28) then results in

$$\vartheta_{\text{opt}} = \arg\min_\vartheta \frac{8\sqrt{3} - \left(\sqrt{3}+\vartheta\right)\left(3-\vartheta^2\right)^2}{\left(\sqrt{3}+\vartheta\right)\left(4-\left(3-\vartheta^2\right)^2\right)}. \tag{33}$$

Numerical inspection of (33) on the interval $\left[0, \sqrt{3}\right]$ shows that $\vartheta_{\text{opt}} = \sqrt{3}$.

4. BLIND SEPARATION OF ARBITRARY SOURCES

4.1. Known Methods

In practice, the distributions of the sources are often identical, since the nature of their origin is related. In that case, the nonlinearity will be the same for all outputs. However, we may find the situation where some sources have different distributions, possibly with a different sign of their respective kurtoses. If in a mixture some sources are sub-Gaussian and some super-Gaussian distributed, the appropriate nonlinearity might be chosen in advance, as long as the number of sub- and super-Gaussian sources is known. Doing so, the system is deprived of some degree of freedom, due to the restriction of permutation within the group of equal kurtosis sign. In other words, once a nonlinearity is chosen, only a signal with the appropriate kurtosis can be separated at that specific output. Other signals are forced away to outputs with the

appropriate nonlinearity. Global convergence can be greatly accelerated by letting the system choose its permutation closest to some initial mixing condition. This can be achieved by an adaptive nonlinearity. If the number of sub-Gaussian and the number of super-Gaussian sources is unknown, adaptive nonlinearities are a necessity.

Douglas *et al.* [13] switch between two nonlinearities, namely

$$g_1(u_i) = u_i^3 \qquad \text{and} \qquad g_2(u_i) = \tanh(10u_i) \tag{34}$$

where $g_1(.)$ and $g_2(.)$ separate sub- and super-Gaussian signals, respectively. The algorithm does not try to normalize its output power regardless of the distribution. A sufficient stability condition is therefore

$$E\left\{g_i'(u_i)\right\} E\left\{u_i^2\right\} - E\left\{g_i(u_i)u_i\right\} > 0. \tag{35}$$

The left-hand side of (35) is constantly evaluated for the two nonlinearities $g_1(.)$ and $g_2(.)$. The larger value decides which nonlinearity is applied.

Similarly, Lee *et al.* [14] present an extended Infomax algorithm, where the update equation for the separation matrix is formulated as

$$\mathbf{W}_{t+1} = \mathbf{W}_t + \mu\left(\mathbf{I} - \mathbf{K}\tanh(\mathbf{u})\mathbf{u}^T - \mathbf{u}\mathbf{u}^T\right)\mathbf{W}_t \tag{36}$$

with $\mathbf{K} = \mathrm{diag}[k_1, \ldots, k_{M_\mathrm{s}}]^T$ being the vector of signs. k_i is positive for a super-Gaussian and negative for a sub-Gaussian signal, respectively. If the distributions are unknown, the sign might be switched according to a kurtosis estimation at the output or some parameter expressing the stability of the nonlinearity currently used. Similarly to (35) it follows

$$k_i\left((1 - E\left\{\tanh(u_i)^2\right\})E\left\{u_i^2\right\} - E\left\{\tanh(u_i)u_i\right\}\right) > 0. \tag{37}$$

By choosing k_i the same sign as the rest of (37), the algorithm is stabilized. Thus, the sign must be adapted as

$$k_i = \mathrm{sign}\left((1 - E\left\{\tanh(u_i)^2\right\})E\left\{u_i^2\right\} - E\left\{\tanh(u_i)u_i\right\}\right). \tag{38}$$

Again, output powers are not normalized, and depend on the source distributions.

4.2. Adaptive Threshold Nonlinearity

Since we know that any non-Gaussian distribution can be separated by the threshold nonlinearity with either $\vartheta = 0$ or $\vartheta \approx 1.5$, we can set up an algorithm in which the update equation for the separation matrix is given by (2) with

$$g_i(u_i) = \begin{cases} 0, & |u_i| < \vartheta_i \\ a_i\,\mathrm{sign}(u_i), & |u_i| \geq \vartheta_i. \end{cases} \tag{39}$$

Each threshold ϑ_i is chosen from $\{0, 1.5\}$ as that value which maximizes the right-hand side of (17) with $g_i(.)$ of (39). The use of the stability equation to switch between the two threshold values has two important disadvantages. First, the value of $p_S(\vartheta)$ is difficult to work out since the function $p_S(.)$ is unknown. Second, although the threshold nonlinearity successfully separates discrete distributions, $p_S(\vartheta)$ is generally zero for discrete distributions as used in data communications, making the switching criterion invalid.

An alternative is to adapt the threshold vector $\boldsymbol{\vartheta} = [\vartheta_1, \ldots, \vartheta_{M_\mathrm{s}}]^T$ according to

$$\boldsymbol{\vartheta}_{t+1} = \boldsymbol{\vartheta}_t - \mu_\vartheta \hat{\boldsymbol{\kappa}}_t \tag{40}$$

where $\hat{\boldsymbol{\kappa}}_t$ is an estimate of the output kurtoses of the vector $\mathbf{u}$ at sample time t. Additionally, $\boldsymbol{\vartheta}_{t+1}$ is clipped at 0 and 1.5 to keep it inside a meaningful region.

4.3. Output normalization

In the analysis of the stability we have shown that the scaling factors in $\mathbf{a} = [a_1, \ldots, a_{M_\mathrm{s}}]^T$ have to be chosen according to Eq. (21) in order to obtain output signals with unit variance. In an environment where the probability distributions are given, Eq. (21) can be evaluated off-line and $\mathbf{a}$ is thus fixed for the adaptation part. If the distributions are unknown, however, $\mathbf{a}$ itself has to be found adaptively. To this end, we note that for unimodal, symmetric distributions, a_i is a monotonously decreasing function of the output standard deviation σ_{U_i}. Vice versa, σ_{U_i} can be written as $\sigma_{U_i} = f_i(a_i)$, where $f_i(.)$ is a monotonously decreasing function for $a_i > 0$, hence

$$\frac{\partial f_i(a_i)}{\partial a_i} < 0. \tag{41}$$

The exact course of $f_i(.)$ depends on the pdf of the ith source. For convenience we denote $\mathbf{f}(.) = [f_1(.), \ldots, f_{M_\mathrm{s}}(.)]^T$. We define our error function $\mathbf{e} = [e_1, \ldots, e_{M_\mathrm{s}}]^T$ by the deviation from unit variance

$$e_i = 1 - \hat{\sigma}_{U_i}^2 \tag{42}$$

and its sum of squares as the cost function

$$J(\mathbf{a}) = \mathbf{e}^T\mathbf{e} = \sum_{i=1}^{M_\mathrm{s}} e_i^2 = \sum_{i=1}^{M_\mathrm{s}} (1 - \hat{\sigma}_{U_i}^2)^2. \tag{43}$$

$\hat{\sigma}_{U_i}^2$ denotes the estimation of the output power $\sigma_{U_i}^2$. The derivative of the cost function $J(\mathbf{a})$ with respect to the gain $\mathbf{a}$ is

$$\begin{aligned} \nabla_\mathbf{a} J(\mathbf{a}) &= 2\mathbf{e} \odot [\frac{\partial e_1}{\partial a_1}, \ldots, \frac{\partial e_{M_\mathrm{s}}}{\partial a_{M_\mathrm{s}}}]^T \\ &= -4\left[(1 - \hat{\sigma}_{U_1}^2)\hat{\sigma}_{U_1}\frac{\partial f_1(a_1)}{\partial a_1}, \ldots, \right. \\ &\qquad \left. , \ldots, (1 - \hat{\sigma}_{U_{M_\mathrm{s}}}^2)\hat{\sigma}_{U_{M_\mathrm{s}}}\frac{\partial f_{M_\mathrm{s}}(a_{M_\mathrm{s}})}{\partial a_{M_\mathrm{s}}}\right]^T. \end{aligned} \tag{44}$$

$\odot$ denotes element-wise multiplication of two vectors. We can now develop a stochastic gradient algorithm to adapt the gain vector

$$\mathbf{a}_{t+1} = \mathbf{a}_t - \mu\nabla_\mathbf{a} J(\mathbf{a}). \tag{45}$$

Using Eq. (41) and the fact that $\hat{\sigma}_{U_i} > 0$, we can incorporate $\hat{\sigma}_{U_i}$ and $\frac{\partial f_i(a_i)}{\partial a_i}$ into a different step size μ_a and write

$$\mathbf{a}_{t+1} = \mathbf{a}_t - \mu_a(\mathbf{1} - \hat{\boldsymbol{\sigma}}_U^2) \tag{46}$$

with $\hat{\boldsymbol{\sigma}}_U^2 = [\hat{\sigma}_{U_1}^2, \ldots, \hat{\sigma}_{U_{M_\mathrm{s}}}^2]^T$ being the vector of power estimates and $\mathbf{1}$ a vector of ones, respectively. (46) is a simple AGC (automatic gain control) algorithm, which normalizes the outputs of the separation process. It runs along with the adaptation of $\mathbf{W}$

and ϑ. Alternately, the normalization can of course be performed by a separate AGC stage after the separation process. This is for example necessary if the mixture contains binary sources. It is intuitively clear that a normalized source with symbol values ± 1 produces zero output after the threshold nonlinearity with $\vartheta = 1.5$.

In summary, the adaptive threshold nonlinearity algorithm is given as

Adaptive Threshold Nonlinearity Algorithm

$$\mathbf{W}_{t+1} = \mathbf{W}_t + \mu \left(\mathbf{I} - \mathbf{g}(\mathbf{u})\mathbf{u}^T \right) \mathbf{W}_t \tag{47}$$

$$g_i(u_i) = \begin{cases} 0, & |u_i| < \vartheta_i \\ a_i \operatorname{sign}(u_i), & |u_i| \geq \vartheta_i \end{cases} \tag{48}$$

$$\boldsymbol{\vartheta}_{t+1} = \operatorname{clip}\left(\boldsymbol{\vartheta}_t - \mu_\vartheta \hat{\boldsymbol{\kappa}}_t \right) \tag{49}$$

$$\mathbf{a}_{t+1} = \mathbf{a}_t - \mu_a (\mathbf{1} - \hat{\boldsymbol{\sigma}}_U^2). \tag{50}$$

5. SIMULATIONS

For the following simulations of the convergence behavior of blind signal separation using the adaptive threshold device, $M_\mathrm{s} = 10$ independent source signals were mixed by random matrix $\mathbf{A}$, whose condition number is chosen $\chi(\mathbf{A}) = 100$ (the singular values of $\mathbf{A}$ are logarithmically distributed). Block processing with a block length $L = 64$ was applied. With this length the kurtosis estimation for the purpose of threshold adaptation is accurate enough, and inter-block memory does not offer any advantage.

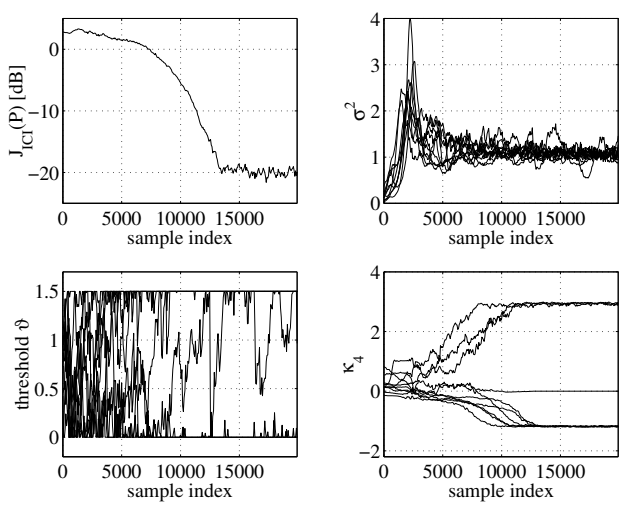

Figure 3: Course of some statistics during the separation process of mixed-kurtosis signals. Top left: Separation performance. Top right: Output powers. Bottom left: Adaptive threshold values. Bottom right: Kurtoses of output signals.

In the first computer experiment we mixed three Laplacian, three uniform, three 16-PAM, and one Gaussian source. If more sources are Gaussian distributed, they can still be separated from other sources by the adaptive threshold nonlinearity, but remain mixed among themselves, leading to a disturbed permutation matrix. This is an inherent limitation of blind separation using higher-order statistics, and is usually circumvented by the restriction to at most one Gaussian source. The adaptive threshold nonlinearity algorithm (47) to (50) was then used to separate the signals in a block-processing manner. The step size of the adaptation was adjusted for a residual mixing of $J_\mathrm{ICI}(\mathbf{P}) = -20\,\mathrm{dB}$, where the performance measure

$$J_\mathrm{ICI}(\mathbf{P}) = \frac{1}{M_\mathrm{s}} \left(\sum_{i=1}^{M_\mathrm{s}} \frac{\sum_{k=1}^{M_\mathrm{s}} p_{ik}^2}{\max_k p_{ik}^2} \right) - 1 \tag{51}$$

is the average interchannel interference and is described in [15].

Fig. 3 shows the adaptation process. The effect of the AGC can be observed as well as the convergence of the kurtoses of the output signals to the respective values 3 for Laplacian, -1.2 for uniform, -1.209 for 16-PAM, and 0 for Gaussian distributions. The threshold values approach either 0 or 1.5, depending on their kurtoses, and converge around 10000 samples, except for the output with the Gaussian distribution where the threshold remains undecided.

In the next simulation we wanted to compare the adaptive threshold nonlinearity algorithm with the algorithms found by Douglas *et al.* [13] and Lee *et al.* [14]. To this end, we mixed 5 Laplacian and 5 uniform sources. The three algorithms were then run with as similar parameters as possible to allow a fair comparison. A block processing with the block size $L = 64$ was used for all algorithms. The step sizes of the adaptation algorithms were adjusted individually for a residual mixing of $J_\mathrm{ICI}(\mathbf{P}) = -20\,\mathrm{dB}$. Averaging over 50 runs with different matrices (all with the characteristics as described above) were carried out to get typical behavior. Fig. 4 shows the separation performance for all tested algorithms. The adaptive threshold nonlinearity algorithm and the algorithm by Douglas *et al.* reach the -20 dB point at exactly the same time on average, whereas the extended Infomax algorithm needs considerably more time.

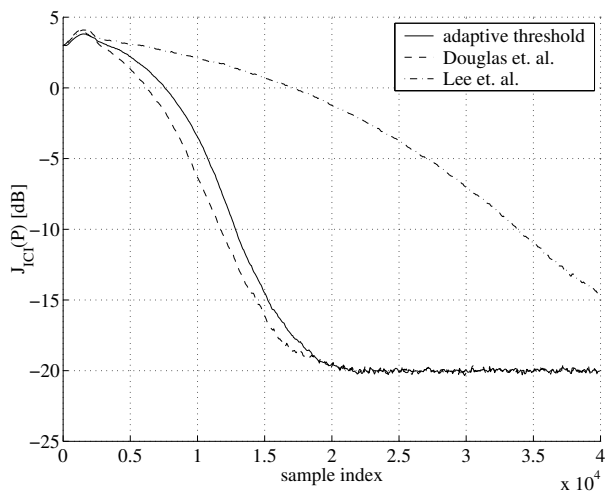

Figure 4: Separation performance of adaptive threshold nonlinearity.

A comparison with simulations of Laplacian sources only [5] shows, that the convergence time associated with Laplacian sources only is about the same as the convergence time obtained here. Simulations also confirm that the adaptive threshold concept is advantageous for mixed-kurtosis signals even if the distributions are known, since by fixing the nonlinearity in advance we deprive

the system of some degree of freedom by restricting the distributions to the outputs with the appropriate nonlinearity, whereas with the adaptive threshold, the system is free to choose among more permutations, thus reducing convergence time.

6. CONCLUSIONS

A threshold nonlinearity for the blind separation of any non-Gaussian sources has been derived. The threshold nonlinearity is not just a simplification of polynomial functions but the true score function for the uniform distribution. As shown in a stability analysis, it separates any sub-Gaussian signals and, if the threshold is reduced to zero, even super-Gaussian signals.

Using the kurtosis of the output signal to control the threshold parameter, the adaptive threshold nonlinearity might be used for the blind separation of mixed-kurtosis signals.

The threshold nonlinearity offers very simple implementation options, since the set of possible output values of the nonlinearity only contains three values, $\pm a$ and 0 for negative-kurtosis signals, and two values, $\pm a$ for positive-kurtosis signals, respectively. The threshold operation can easily be implemented by two comparators only.

7. APPENDIX: DETAILED DERIVATION OF THE THRESHOLD NONLINEARITY

Differentiating (6) with respect to s leads to

$$p_S'(s) = -\alpha \left(\frac{|s|}{\beta}\right)^{\alpha-1} \frac{\operatorname{sign}(s)}{\beta} \frac{\alpha}{2\beta\Gamma\left(\frac{1}{\alpha}\right)} e^{-\left(\frac{|s|}{\beta}\right)^{\alpha}}. \quad (52)$$

If we divide (52) by (6) and flip the sign we get

$$g(s) = -\frac{p_S'(s)}{p_S(s)} = \alpha \left(\frac{|s|}{\beta}\right)^{\alpha-1} \frac{\operatorname{sign}(s)}{\beta} = \frac{\alpha}{\beta^{\alpha}} |s|^{\alpha-1} \operatorname{sign}(s). \quad (53)$$

For unit variance, we can find β from the general expression for the nth-order moment of a generalized Gaussian signal [16]

$$E(|X|^n) = \frac{\Gamma(\frac{n+1}{\alpha})}{\Gamma(\frac{1}{\alpha})} \beta^n. \quad (54)$$

For $n = 2$, Eq. (54) gives

$$E(|X|^2) = \frac{\Gamma(\frac{3}{\alpha})}{\Gamma(\frac{1}{\alpha})} \beta^2. \quad (55)$$

or, for unit variance, we have

$$\beta = \sqrt{\frac{\Gamma(\frac{1}{\alpha})}{\Gamma(\frac{3}{\alpha})}}. \quad (56)$$

Inserting (56) into (53) yields (7). Using $\Gamma(x) \cdot \Gamma(1-x) = \pi / \sin(\pi x)$ from [17] leads to

$$g(s) = \alpha \left(\frac{\frac{\pi}{\sin \frac{3\pi}{\alpha}}}{\frac{\pi}{\sin \frac{\pi}{\alpha}}} \cdot \frac{\Gamma(1-\frac{\pi}{\alpha})}{\Gamma(1-\frac{3\pi}{\alpha})} \right)^{\frac{\alpha}{2}} |s|^{\alpha-1} \operatorname{sign}(s). \quad (57)$$

Both terms $\Gamma(1-\frac{\pi}{\alpha})$ and $\Gamma(1-\frac{3\pi}{\alpha})$ are close to $\Gamma(1) = 1$ for large values of α, so that simplification of (57) yields (8). The first term of the Taylor expansion of a sine function for a small argument is just the argument itself, leading to (9). Eq. (10) finally is a consequence of the behavior of $\lim_{b\to\infty} a^b$ depending on a being less or greater than one.

8. REFERENCES

[1] S.-I. Amari, A. Cichocki, and H. H. Yang, "A new learning algorithm for blind signal separation," *Advances in Neural Information Processing Systems*, vol. 8, pp. 757–763, 1996.

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

[3] J.-F. Cardoso, "Blind signal separation: Statistical principles," *Proc. IEEE*, vol. 86, no. 10, pp. 2009–2025, Oct. 1998.

[4] M. Girolami and C. Fyfe, "Negentropy and kurtosis as projection pursuit indices provide generalised ICA algorithms," in *Proc. NIPS*, Aspen, CO, Dec, 7, 1996.

[5] M. Joho and H. Mathis, "Performance comparison of combined blind/non-blind source separation algorithms," in *Proc. ICA*, Aussois, France, Jan. 11–15, 1999, pp. 139–142.

[6] R. H. Lambert and A. J. Bell, "Blind separation of multiple speakers in a multipath environment," in *Proc. ICASSP*, Munich, Germany, Apr. 21–24, 1997, pp. 423–426.

[7] A. Hyvärinen and E. Oja, "Independent component analysis by general nonlinear Hebbian-like learning rules," *Signal Processing*, vol. 64, no. 3, pp. 301–313, Feb. 1998.

[8] T.-W. Lee, M. Girolami, A. J. Bell, and T. J. Sejnowski, "A unifying information-theoretic framework for independent component analysis," *Int. J. Math. and Comp. Modeling*, in press, 1999.

[9] H. Mathis, M. Joho, and G. S. Moschytz, "A simple threshold nonlinearity for blind signal separation," in *Proc. ISCAS*, Geneva, Switzerland, May 28–31, 2000, accepted for publication.

[10] T. P. von Hoff, A. G. Lindgren, and A. N. Kaelin, "Transpose properties in the stability and performance of the classic adaptive algorithms for blind source separation and deconvolution," *Signal Processing*, vol. 80, no. 9, 2000, to appear.

[11] J.-F. Cardoso and B. H. Laheld, "Equivariant adaptive source separation," *IEEE Trans. Signal Processing*, vol. 44, no. 12, pp. 3017–3030, Dec. 1996.

[12] S.-I. Amari, T.-P. Chen, and A. Cichocki, "Stability analysis of adaptive blind source separation," *Neural Networks*, vol. 10, no. 8, pp. 1345–1351, Aug. 1997.

[13] S. C. Douglas, A. Cichocki, and S. Amari, "Multichannel blind separation and deconvolution of sources with arbitrary distributions," in *IEEE Workshop on Neural Networks for Signal Processing*, Almelia Island Plantation, FL, Sept. 1997, pp. 436–445.

[14] T.-W. Lee, M. Girolami, and T. J. Sejnowski, "Independent component analysis using an extended infomax algorithm for mixed sub-Gaussian and super-Gaussian sources," *Neural Computation*, vol. 11, no. 2, pp. 417–441, 1999.

[15] M. Joho, H. Mathis, and G. S. Moschytz, "An FFT-based algorithm for multichannel blind deconvolution," in *Proc. ISCAS*, Orlando, FL, May 30 – June 2, 1999, pp. III–203–206.

[16] R. H. Lambert, *Multichannel Blind Deconvolution: FIR Matrix Algebra and Separation of Multipath Mixtures*, Ph.D. thesis, University of Southern California, 1996.

[17] I. N. Bronshtein and K. A. Semendyayev, *Handbook of Mathematics*, Springer Verlag, 3rd edition, 1997.