IEEE International Symposium on Circuits and Systems ISCAS 2000, May 28–31, 2000, Geneva, Switzerland, pp. IV-489–492.

A SIMPLE THRESHOLD NONLINEARITY FOR BLIND SEPARATION OF SUB-GAUSSIAN SIGNALS

Heinz Mathis, Marcel Joho, and George S. Moschytz

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

ABSTRACT

A computationally simple nonlinearity in the form of a threshold device for the blind separation of sub-Gaussian signals is derived. Convergence is shown to be robust, fast, and comparable to that of more complex polynomial nonlinearities. Together with the known signum nonlinearity for super-Gaussian distributions, which basically is a threshold device with the threshold set to zero, the general threshold nonlinearity (with an appropriate threshold) can separate any non-Gaussian signals.

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. Throughout this paper we assume a linear mixing and separation process as depicted in Fig. 1.

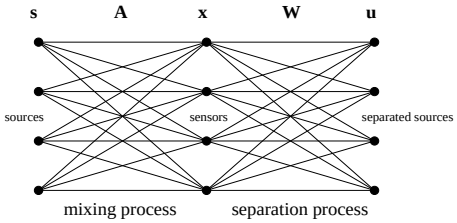

Figure 1: Blind source separation model.

The measured signals $\mathbf{x} = [x_1, \ldots, x_{M_S}]^T$ to be processed are linear combinations of the original source signals $\mathbf{s} = [s_1, \ldots, s_{M_S}]^T$, weighted by scalars, which are the elements of the mixing matrix $\mathbf{A}$. M_S denotes the total number of sources and 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

$$\mathbf{u} = [u_1, \ldots, u_{M_S}]^T = \mathbf{Wx} = \mathbf{WAs} = \mathbf{Ps}. \tag{1}$$

In order to successfully separate the signals, $\mathbf{P} = \mathbf{WA}$ 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, the output entropy maximization [1], the ML estimator [2], the maximum negentropy [3], or using a Bussgang technique [4], and applying the natural gradient [5] to these methods

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

where μ is the step size, $\mathbf{I}$ the identity matrix, and $\mathbf{g(u)}$ is known as the *Bussgang nonlinearity* [6]. This nonlinearity is the central element of blind signal separation. Its role is defined by the objective or contrast function, which often is 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 [5], both the Infomax and the Maximum-Likelihood approach [8] lead to

$$g(u_i) = -\frac{\partial \log p_S(u_i)}{\partial u_i} = -\frac{p'_S(u_i)}{p_S(u_i)} \tag{3}$$

where $p_S(u_i)$ and $p'_S(u_i)$ are the pdf and its derivative, respectively, of the source signals. Eq. (3) is referred to as the score function of a certain pdf $p_S(.)$. We assume the same probabilistic model for all source signals, e.g. $p_{S_i}(.) = p_S(.)$. Note that the nonlinearity is used solely for the update process. The separation itself is obviously linear, since the mixing process is linear by assumption. This is particularly important for acoustic applications, where nonlinear signal processing would generate unacceptable audible distortion.

2. SCALING OF THE NONLINEARITY

By the scaling invariance, which is an inherent property of blind signal separation, it is impossible to recover the original power of the source signals without further knowledge. When the original power of the sources is unknown, it is reasonable to normalize the power after the separation matrix to

$$E\left[\mathbf{uu}^H\right] = \mathbf{I}. \tag{4}$$

This can be achieved by a separate automatic gain control stage (AGC) or by scaling the nonlinearity $g(.)$ properly. The Bussgang property [9] states that an equilibrium point of Eq. (2) is reached when

$$E\left[\mathbf{g(u)u}^H\right] = \mathbf{I}. \tag{5}$$

Hence, for every component u_i of the vector $\mathbf{u}$, $i = 1, \ldots, M_S$, we need to scale $g(u_i)$ such that

$$\int_{-\infty}^{\infty} p_S(u_i) g(u_i) u_i^* du_i = 1 \tag{6}$$

if $p_S(.)$ is a source distribution with $E[|S|^2] = 1$. Note, that for the score function (3) of most common distributions, (6) is satisfied without further scaling.

3. THE FORM OF THE NONLINEARITY

The basic form of the nonlinearity depends on the distribution of the source signal. Signals with negative kurtosis (normalized fourth-order cumulant) are referred to as sub-Gaussian, positive-kurtosis signals as super-Gaussian. For sub-Gaussian distributed real signals, a nonlinearity of the form $g(u_i) = a \cdot u_i^3$ is often selected [10]. a is determined by (6). In fact, any nonlinearity of the form

$$g(u_i) = a \cdot u_i^p, \quad p \text{ odd}, p \geq 3 \tag{7}$$

will separate a mixture of sub-Gaussian signals. p being odd ensures the validity of the sign after the nonlinearity. If (7) is rewritten as

$$g(u_i) = a \cdot |u_i|^{p-1} u_i, \qquad p > 1 \tag{8}$$

p is no longer restricted to odd integers, but can be any rational number greater than one. (8) also has the advantage that it is directly applicable to complex signals, as will be seen in Section 6.

For distributions which belong to the family of generalized Gaussian distributions

$$p_S(s) = \frac{\alpha}{2\beta\Gamma\left(\frac{1}{\alpha}\right)} e^{-\left(\frac{|s|}{\beta}\right)^\alpha} \tag{9}$$

with $\alpha > 2$ for sub-Gaussian signals, the nonlinearity according to (3) is

$$g(u_i) = \alpha \left(\frac{\Gamma\left(\frac{3}{\alpha}\right)}{\Gamma\left(\frac{1}{\alpha}\right)} \right)^{\alpha/2} \text{sign}(u_i) \cdot |u_i|^{\alpha-1}. \tag{10}$$

$\Gamma(.)$ 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, $\Gamma(a+1) = a\Gamma(a)$. The uniform distribution can be modeled as a generalized Gaussian distribution in the limit $\alpha \to \infty$. For large α, (10) yields

$$g(u_i)\Big|_{\alpha \gg 1} \approx \alpha \left(\frac{\sin\left(\frac{\pi}{\alpha}\right)}{\sin\left(\frac{3\pi}{\alpha}\right)} \right)^{\alpha/2} \text{sign}(u_i) \cdot |u_i|^{\alpha-1}. \tag{11}$$

However, for large α the sine functions are well represented by the first term of their Taylor expansions, so (11) becomes

$$g(u_i)\Big|_{\alpha \gg 1} \approx \alpha \left(\frac{1}{3} \right)^{\alpha/2} \text{sign}(u_i) \cdot |u_i|^{\alpha-1} = \alpha \frac{1}{u_i} \left(\frac{u_i^2}{3} \right)^{\alpha/2}. \tag{12}$$

In the limit for α approaching infinity we get a *threshold nonlinearity*

$$\lim_{\alpha \to \infty} g(u_i) = \begin{cases} 0, & |u_i| < \sqrt{3} \\ \infty \cdot \text{sign}(u_i), & |u_i| \geq \sqrt{3} \end{cases} \tag{13}$$

with a threshold of $\sqrt{3}$ for a uniform distribution with unity power. The normalized uniform distribution only has a finite probability density for $|u_i| < \sqrt{3}$; outside it is zero. With $g(u_i)$ being zero for small u_i, $\mathbf{W}_{t+1}$ in (2) grows gradually, thereby increasing u_i. When u_i 'hits' the threshold, it is pushed back hard (infinite gain) into the region where $g(u_i) = 0$, so that the amplitude of u_i is clearly controlled. The infinite gain in (13) 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. Solving (6) for a given threshold ϑ results in a finite gain of

$$a = \frac{2\sqrt{3}}{3 - \vartheta^2} \tag{14}$$

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

$$g(u_i) = \begin{cases} 0, & |u_i| < \vartheta \\ \frac{2\sqrt{3}}{3-\vartheta^2} \text{sign}(u_i), & |u_i| \geq \vartheta \end{cases} \tag{15}$$

is very simple to implement and reduces computational complexity, compared with the evaluation of polynomials. An alternative nonlinearity for sub-Gaussian signals has been derived by Lambert [11], also using the MAP rule. Fig. 2 shows the threshold nonlinearity in comparison with other nonlinearities of the form given in (10) with proper scaling applied.

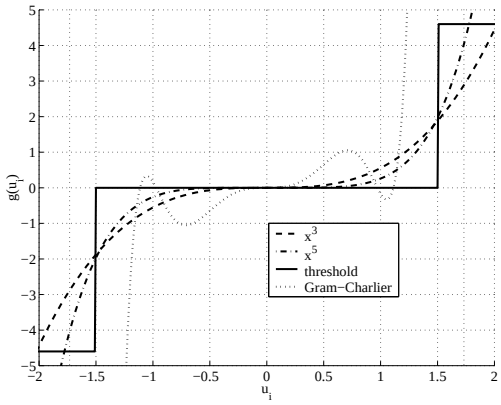

Figure 2: Threshold nonlinearity ($\vartheta = 1.5$) and other scaled nonlinearities suitable for the separation of uniformly distributed signals.

A similar, simple nonlinearity is available for the Laplace distribution ($\alpha = 1$), which is a super-Gaussian distribution, where (10) leads to

$$g(u_i) = \sqrt{2}\,\text{sign}(u_i). \tag{16}$$

4. UNIFORMLY DISTRIBUTED SIGNALS

To separate uniformly distributed real signals, very often a nonlinearity of the form given by (7) is used. Scaling according to (6) leads to

$$g(u_i) = \frac{p+2}{3^{\frac{p+1}{2}}} u_i^p, \qquad p \text{ odd}, p \geq 3. \tag{17}$$

For the following simulation of the convergence behavior of blind signal separation using the threshold device, $M_S = 10$ independent, uniformly distributed source signals are mixed by matrix $\mathbf{A}$, whose condition number is chosen $\chi(\mathbf{A}) = 100$ (the singular values of $\mathbf{A}$ are logarithmically distributed). The step size μ is tuned such as to reach a residual mixing of $J_{\text{ICI}}(\mathbf{P}) = -20\,\text{dB}$, where the performance measure

$$J_{\text{ICI}}(\mathbf{P}) = \frac{1}{M_S} \left(\sum_{i=1}^{M_S} \frac{\sum_{k=1}^{M_S} p_{ik}^2}{\max_k p_{ik}^2} \right) - 1 \tag{18}$$

is the average interchannel interference and is described in [12]. Fig. 3 displays the performance curves for different nonlinearities, such as (17) for different p, and the nonlinearity derived from the application of the Gram-Charlier expansion [5]

$$g(u_i) = \frac{3}{4}u_i^{11} + \frac{15}{4}u_i^9 - \frac{14}{3}u_i^7 - \frac{29}{4}u_i^5 + \frac{29}{4}u_i^3. \qquad (19)$$

Clearly, the threshold device shows a convergence behavior comparable to that of more complicated nonlinearities. Best results were achieved with a threshold of around $\vartheta = 1.5$.

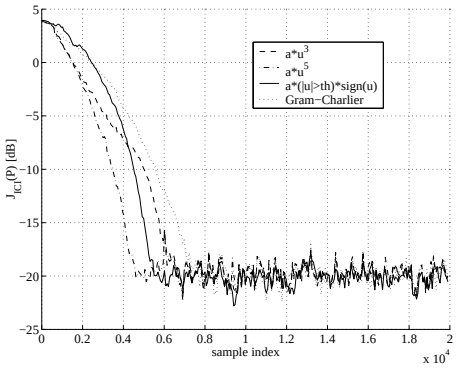

Figure 3: Convergence curves of blind separation with different nonlinearities for uniformly distributed signals.

5. M-PAM SIGNALS

M-ary pulse amplitude modulation, or equivalently M-ary amplitude shift keying is a linear digital modulation scheme whose pdf may be approximated by a uniform distribution, particularly for large M. Applying the scaling equation (6) to Eq. (8) yields a suitable nonlinearity

$$g(u_i) = \frac{\frac{M}{2}\left(\frac{M^2-1}{3}\right)^{\frac{p+1}{2}}}{\sum_{m=1}^{M/2}(2m-1)^{p+1}} \cdot |u_i|^{p-1}u_i, \qquad p > 1. \qquad (20)$$

Note, that the term in (20) that is written as a fraction is merely the scaling factor, and is e.g. reduced to 1 for $M = 2$. Again, Eq. (20) may be replaced by a threshold nonlinearity as in Eq. (15) with an appropriate scaling factor.

6. M-QAM SIGNALS

M-ary quadrature amplitude modulation is an important digital modulation format, which can be regarded as an extension of pulse amplitude modulation to the complex baseband representation. For the common case that M is an even power of two, e.g. $M = 16$, I and Q parts are independent of each other. The distribution is sub-Gaussian, and similarly to the uniform distribution and the M-PAM signals, nonlinearities of the form

$$g(u_i) = \frac{4 \cdot 5^{\frac{p+1}{2}}}{1 + 2 \cdot 5^{\frac{p+1}{2}} + 3^{p+1}} |u_i|^{p-1}u_i, \qquad p > 1 \qquad (21)$$

lead to separated, normalized output signals. The mixing matrix $\mathbf{A}$ of a complex baseband representation is generally complex, in order to model amplitude and phase variations. The separation matrix $\mathbf{W}$ must therefore also be chosen to be complex. In addition to the amplitude and permutation invariance, we get a phase invariance, or in other words, the entries of $\mathbf{P}$ in (1) are complex. As a consequence, the signal vectors $\mathbf{s}$, $\mathbf{x}$ and $\mathbf{u}$ are now assumed to be complex. The threshold nonlinearity for QAM signals is chosen as

$$g(u_i) = \begin{cases} 0, & |u_i| < \vartheta \\ a\frac{u_i}{|u_i|}, & |u_i| \geq \vartheta. \end{cases} \qquad (22)$$

It has been found by simulations, that $\vartheta = 1.3$ is an effective threshold value for 16-QAM. The scaling factor $a = 2.98$ is obtained from evaluating the scaling equation (6) for the discrete 16-QAM distribution. To adjust for residual mixing, we chose $a = 4$ to obtain unit variance signals.

The nonlinearities so far leave the initial rotation of the signal constellation unchanged. Rather than applying the nonlinearity to the complex signal, it can be applied individually to the real and imaginary part of the signal, which, by a further use of the scaling equation (6), leads to nonlinearities of the form

$$\tilde{g}(u_i) = \frac{10^{\frac{p+1}{2}}}{1+3^{p+1}}\left(u_{\mathrm{R},i}^p + j u_{\mathrm{I},i}^p\right) = g(u_{\mathrm{R},i}) + j g(u_{\mathrm{I},i}) \qquad (23)$$

with $u_{\mathrm{R},i} = \mathrm{Re}(u_i)$ and $u_{\mathrm{I},i} = \mathrm{Im}(u_i)$. Likewise, for the vector form we assign $\mathbf{u}_\mathrm{R} = \mathrm{Re}(\mathbf{u})$ and $\mathbf{u}_\mathrm{I} = \mathrm{Im}(\mathbf{u})$. The Bussgang property can then be written as

$$\begin{aligned} E[\mathbf{I} - \tilde{\mathbf{g}}(\mathbf{u})\mathbf{u}^H] &= E\left[\mathbf{I} - \left(g(\mathbf{u}_\mathrm{R}) + j g(\mathbf{u}_\mathrm{I})\right)\left(\mathbf{u}_\mathrm{R}^T - j\mathbf{u}_\mathrm{I}^T\right)\right] \\ &= E\left[\mathbf{I} - g(\mathbf{u}_\mathrm{R})\mathbf{u}_\mathrm{R}^T - g(\mathbf{u}_\mathrm{I})\mathbf{u}_\mathrm{I}^T \right. \\ &\quad \left. + j\left(g(\mathbf{u}_\mathrm{R})\mathbf{u}_\mathrm{I}^T - g(\mathbf{u}_\mathrm{I})\mathbf{u}_\mathrm{R}^T\right)\right] = \mathbf{0}. \end{aligned} \qquad (24)$$

The Bussgang property now requires for the diagonal elements

$$E[g(u_{\mathrm{R},i})u_{\mathrm{R},i} + g(u_{\mathrm{I},i})u_{\mathrm{I},i}] = 1 \qquad (25)$$

$$E[g(u_{\mathrm{R},i})u_{\mathrm{I},i} - g(u_{\mathrm{I},i})u_{\mathrm{R},i}] = 0 \qquad (26)$$

and for the off-diagonal elements

$$E[g(u_{\mathrm{R},i})u_{\mathrm{R},k} + g(u_{\mathrm{I},i})u_{\mathrm{I},k}] = 0, \quad i \neq k \qquad (27)$$

$$E[g(u_{\mathrm{R},i})u_{\mathrm{I},k} - g(u_{\mathrm{I},i})u_{\mathrm{R},k}] = 0, \quad i \neq k. \qquad (28)$$

(27) and (28) make the output signals mutually independent, while (25) ensures that the signals are properly scaled. (26) makes the real and imaginary part of an output signal independent of each other, forcing the signal constellation to rotate into its right position, or into a mirrored, or rotated-by-a-multiple-of-$\pi/2$, version of itself. This effect can be exploited for phase synchronization.

Similarly to the last experiment for uniform distribution, ten 16-QAM sources were mixed using a complex mixing matrix $\mathbf{A}$ with $\chi(\mathbf{A}) = 100$ and logarithmically spaced singular values. Performance curves are depicted in Fig. 4. It can be observed that nonlinearities applied individually to the real and imaginary parts tend to separate the signals faster than complex nonlinearities. The effect to the back-rotation of the signal constellation is shown in Fig. 5.

Similar results can be obtained for the convergence speed of 256-QAM signals (see Fig. 6). Overall convergence is slightly slower for this modulation scheme than for that of 16-QAM.

It is relatively straightforward to extend the blind separation of instantaneously mixed sources using the threshold nonlinearity to do single-channel or multichannel blind deconvolution, which will be the subject of future work. For an efficient implementation, the methods described in [12] may be applied.

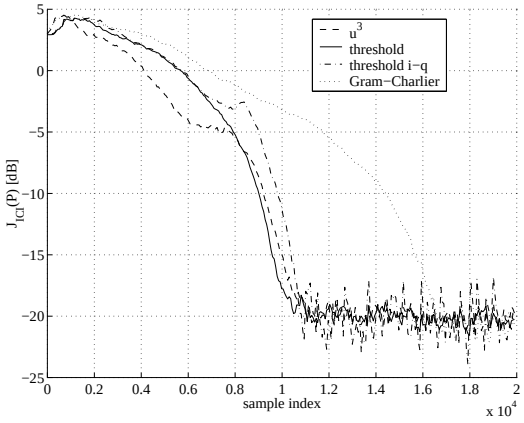

Figure 4: Convergence curves of blind separation with different nonlinearities for 16-QAM signals.

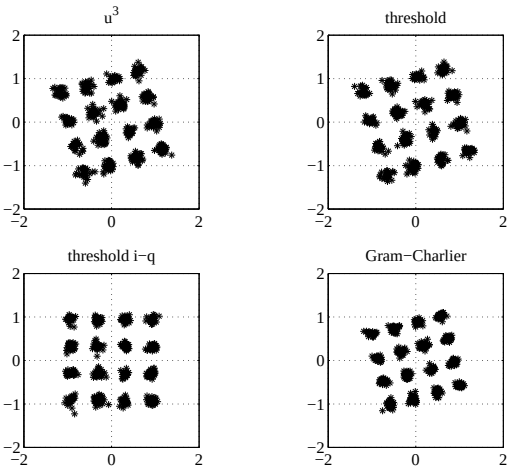

Figure 5: Signal constellation diagrams of 16-QAM signals after separation with different nonlinearities.

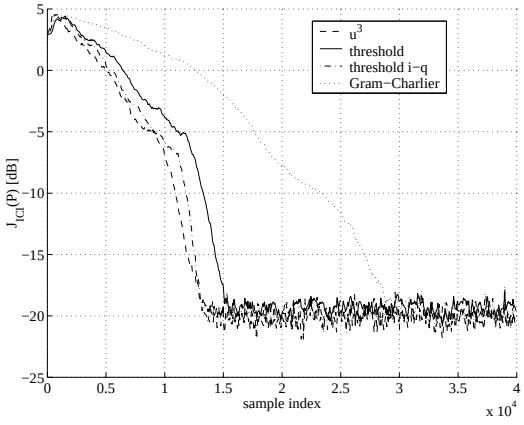

Figure 6: Convergence curves of blind separation with different nonlinearities for 256-QAM signals.

7. CONCLUSIONS

A simple threshold device has been shown to separate sub-Gaussian signals, e.g. M-ary PAM and QAM. This nonlinearity is simple to implement, very robust, and results in a convergence speed that is comparable to that of other known nonlinearities. If real and imaginary parts of the signals are treated separately, phase equalization up to an ambiguity of $\pi/2$ can be achieved.

8. REFERENCES

[1] 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.

[2] J.-F. Cardoso, "Blind signal separation: Statistical principles," *Proceedings of the IEEE*, vol. 86, no. 10, pp. 2009–2025, October 1998.

[3] M. Girolami and C. Fyfe, "Negentropy and kurtosis as projection pursuit indices provide generalised ICA algorithms," in *Advances in Neural Information Processing Systems*, Aspen, CO, Dec, 7, 1996.

[4] M. Joho and H. Mathis, "Performance comparison of combined blind/non-blind source separation algorithms," in *Proc. International Conference on Independent Component Analysis and Blind Signal Separation*, Aussois, France, January 11–15, 1999, pp. 139–142.

[5] 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.

[6] R. H. Lambert and A. J. Bell, "Blind separation of multiple speakers in a multipath environment," in *International Conference on Acoustics, Speech & Signal Processing*, Munich, Germany, April 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, February 1998.

[8] T.-W. Lee, M. Girolami, A. J. Bell, and T. J. Sejnowski, "A unifying information-theoretic framework for independent component analysis," *International Journal on Mathematical and Computer Modeling*, in press, 1999.

[9] S. Bellini, "Bussgang techniques for blind deconvolution and equalization," in *Blind Deconvolution*, S. Haykin, Ed. 1994, pp. 8–59, Prentice Hall.

[10] S.-I. Amari, S. C. Douglas, A. Cichocki, and H. H. Yang, "Multichannel blind deconvolution and equalization using the natural gradient," in *First IEEE Signal Processing Workshop on Signal Processing Advances in Wireless Communications*, Paris, France, April 16–18 1997, pp. 101–104.

[11] R. H. Lambert and C. L. Nikias, "Optimal blind equalization cost functions and maximum a posteriori estimation," in *IEEE Military Communications Conference*, Ft. Monmouth, NJ, 1994, pp. 291–295.

[12] M. Joho, H. Mathis, and G. S. Moschytz, "An FFT-based algorithm for multichannel blind deconvolution," in *IEEE International Symposium on Circuits and Systems*, Orlando, FL, May 30 – June 2, 1999, pp. III–203–206.