ELSEVIER

Neurocomputing 49 (2002) 61–78

NEUROCOMPUTING

www.elsevier.com/locate/neucom

Blind signal separation in noisy environments using a three-step quantizer

Heinz Mathis[a,*], Marcel Joho[b]

[a] *Signal and Information Processing Laboratory, Swiss Federal Institute of Technology, Zurich, Switzerland*

[b] *Phonak Inc., Champaign, IL, USA*

Received 19 February 2001; accepted 22 October 2001

Abstract

Independent component analysis in noisy channels needs special considerations, since standard solutions lead to a bias in the estimate of the parameters. We show three different approaches to mitigate the effects of additive noise in the transfer medium. A principal component subspace method can reduce the noise to more favorable levels, so that any following algorithm shows reduced bias effects. Although stochastic-gradient algorithms for maximum-likelihood solutions to the problem can easily be found, they are computationally prohibitive. A very successful approach is, therefore, to assume zero noise power for the derivation of the adaptive algorithm and subsequently trying to compensate for any bias introduced by such a solution. The threshold nonlinearity (three-step quantizer) is suitable for the blind separation of a large class of sub-Gaussian distributions. Stability regions are explored followed by algorithmic extensions to suppress the bias in the estimation of the separation matrix. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Unbiased blind signal separation; Noisy ICA; ML solution; Overdetermined blind separation; Threshold nonlinearity

1. Introduction

Blind signal separation using an adaptive algorithm is a technique that has become increasingly important for a vast range of applications in acoustics, communications,

[*] Corresponding author. University of Applied Sciences (HSR), Oberseestrasse 10, CH-8640 Rapperswil, Switzerland. Tel.: +41-55-222-45-95; fax: +41-55-222-44-00 .

E-mail address: heinz.mathis@hsr.ch (H. Mathis).

PII: S0925-2312(02)00513-1

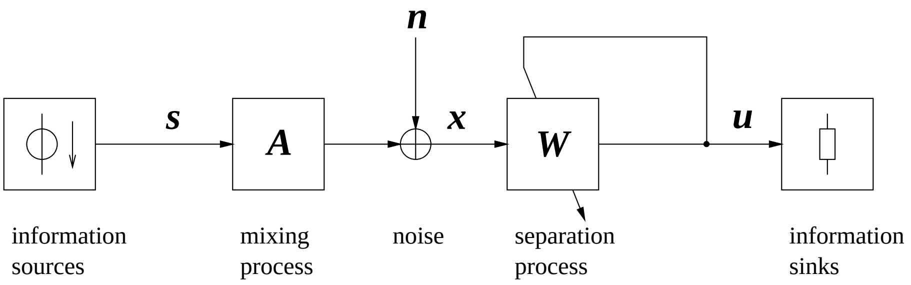

Fig. 1. General blind signal separation model.

biomedical engineering, and so on. The basic issue is to separate a couple of signals from different mixtures of their instances, without knowing the mixing conditions nor any part of the source signals. Analytically, the problem of blind separation of an instantaneous mixture of source signals can be described as follows (see also Fig. 1). Suppose that the information sources generate a number of signals, conveniently described by the vector $\boldsymbol{s} = [s_1, \dots, s_{M_\mathrm{s}}]^\mathrm{T}$. Through a mixing process, usually presumed linear, and therefore represented by an unknown scalar matrix $\boldsymbol{A}$, and an additive noise vector $\boldsymbol{n} = [n_1, \dots, n_{M_\mathrm{s}}]^\mathrm{T}$, we get the observation vector $\boldsymbol{x} = [x_1, \dots, x_{M_\mathrm{s}}]^\mathrm{T}$ at the sensors. M_s here denotes the number of sources as well as the number of sensors. If fewer sensors than sources are available, the problem gets tougher, and generally, a complete separation of all the source signals is no longer possible.

On the other hand, if more sensors than sources are available, the noise suppression capability might be enhanced by using overdetermined separation techniques [5,21,14]. In the communication literature, this situation is referred to as diversity reception. Naturally, the diversity gain is much higher when channel fading occurs. Nonetheless, an improvement is also possible in a static additive-white-Gaussian-noise (AWGN) channel, particularly if some sensors exhibit low signal-to-noise ratios (SNR).

The noisy independent component analysis (ICA) problem has attracted some interest in the literature. It is, e.g., described in a wider framework of Bayesian algorithms by Xu [19]. Hyvärinen [13] considered the maximum-likelihood (ML) solution in the presence of Gaussian noise. A similar approach, although more related to expectation-maximization (EM) algorithms is described by Moulines et al. [18].

Mathematically, we describe the observed signals by

$$\boldsymbol{x} = \boldsymbol{A}\boldsymbol{s} + \boldsymbol{n}. \tag{1}$$

The problem to solve is to find a scalar matrix $\boldsymbol{W}$, describing the separation process, such that the signals in vector $\boldsymbol{u} = [u_1, \dots, u_{M_\mathrm{s}}]^\mathrm{T}$ are noisy replicas of the original source signals up to some invariances.[1] These invariances are:

[1] The literature often refers to such replicas as *wave-preserving signals*, as they maintain the original waveform up to scaling, shifting, and permutation.

- the order of the signals within vector $\boldsymbol{u}$ (permutation),
- the magnitudes of the original source signals (scaling),
- the phases of the original source signals (the signs for real-valued sources).

These invariances are inherently connected to the problem posed rather than to the algorithms solving it, and arise if no assumptions on the variance of the source signals are made. Matrix $\boldsymbol{A}$ may be any invertible square matrix. The assumptions on signals and channels usually are:

- the source signals are mutually independent,
- at most one source signal is Gaussian distributed,
- the sources are stationary and iid (or the underlying process is unknown),
- $\boldsymbol{A}$ is a time-invariant, invertible, square matrix,
- the noise signals are mutually independent,
- the noise signals are independent of the source signals.

As for the noise, sometimes knowledge of σ_{N}^2 is assumed. The recovered signals can be written as

$$\boldsymbol{u} = \boldsymbol{W}\boldsymbol{x} = \boldsymbol{W}(\boldsymbol{A}\boldsymbol{s} + \boldsymbol{n}) = \boldsymbol{P}\boldsymbol{s} + \boldsymbol{W}\boldsymbol{n}. \tag{2}$$

In order to successfully separate the signals, $\boldsymbol{P} = \boldsymbol{W}\boldsymbol{A}$ should approximate a scaled permutation matrix as closely as possible.

The separation process can be modeled as a single-layer neural network (see e.g. [2] or [17]) with an equal number of input and output nodes, where the coefficients w_{ij} of the separation matrix $\boldsymbol{W}$ are simply the weights from the input to the output nodes. The activation functions at the output nodes are used for the training mode only, while the problem itself is linear (since the mixing is a linear process, its inverse operation is linear, too), so that for a successful separation of a linear mixture, a linear combination of the available input signals is adequate. This is particularly important for acoustic applications, where nonlinear signal processing might generate unacceptable audible distortion. In the case of substantial noise, a nonlinear transformation might yield better results (as far as MSE criteria are concerned). In this paper, we restrict ourselves to the linear case.

2. Overdetermined source separation

As mentioned in the last section, overdetermined source separation is a very effective technique to mitigate channel noise, when more mixture observations than source signals are available. Both Douglas [10] and Joho et al. [14] suggest a two-stage blind approach to solve this separation problem. Fig. 2 shows an example of a setup of such a two-stage algorithm with two sources and five sensors. Matrix $\boldsymbol{A}$ is now no longer square but transforms the original source signals into a higher number of mixtures. At this point, sensor noise—or measurement noise—is added. The signals then become the input to the algorithm. In a first stage—the preprocessing step—the original number of source signals is retrieved by a principal component analysis (PCA).

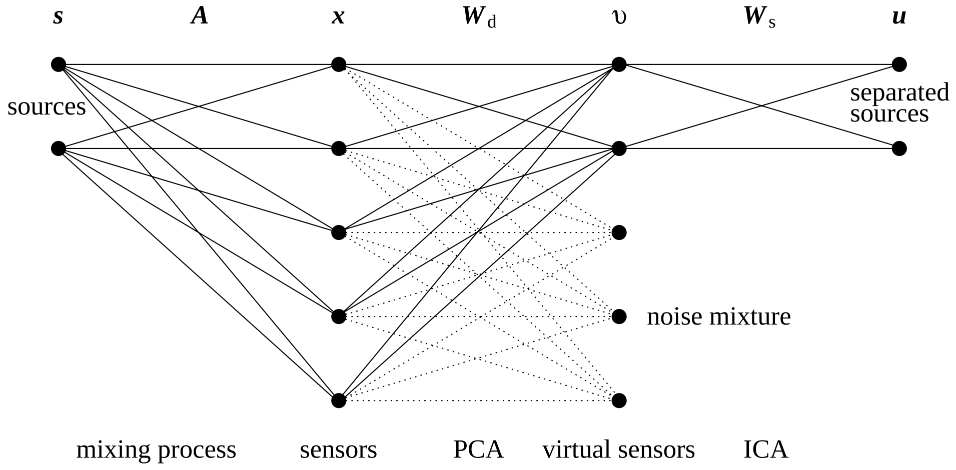

Fig. 2. Overdetermined source separation model using a two-stage approach.

Note that we assume here the signals to be unknown, but the number of source signals to be known in advance. In principle, this might be any subspace decomposition technique, see for example [10], to extract a higher SNR mixture of the signals of interest. The resulting signals $\boldsymbol{v}$ are now treated as signals coming from virtual sensors, so that any ICA technique will separate the signals. Of course, noise is still present after the PCA step, albeit at lower levels, and needs to be addressed by the following stage. As a consequence of a nonsquare matrix $\boldsymbol{A}$, matrix $\boldsymbol{W}_{\mathrm{d}}$ will have the transposed dimension of $\boldsymbol{A}$. The ICA stage is represented by the square matrix $\boldsymbol{W}_{\mathrm{s}}$ with the dimension of the original number of sources. Simulation results in [14] show that diversity gains close to the theoretical optimum of MMSE solutions are possible.

3. Maximum-likelihood solution

One possible solution to the blind signal separation problem can be found by answering the question of what mixing matrix has most likely led to the current observation $\boldsymbol{x}$. Our goal is to find the inverse of the mixing matrix, $\boldsymbol{W} = \boldsymbol{A}^{-1}$. This is a zero-forcing solution, since it nulls any contribution from other sources than the source of interest. We assign as the likelihood the probability of the observation, parameterized by $\boldsymbol{A}$, $p_X(\boldsymbol{x}; \boldsymbol{A})$. If the noise signals were known, we could write the conditional probability

$$p_{X|N}(\boldsymbol{x}|\boldsymbol{n}) = \frac{p_S(\boldsymbol{s})}{|\det \boldsymbol{A}|} = \frac{p_S(\boldsymbol{A}^{-1}(\boldsymbol{x}-\boldsymbol{n}))}{|\det \boldsymbol{A}|} = p_S(\boldsymbol{W}(\boldsymbol{x}-\boldsymbol{n}))|\det \boldsymbol{W}|. \tag{3}$$

The noise vector $\boldsymbol{n}$ is a latent variable we want to get rid of. By integrating over it we get the unconditional probability

$$p_{\boldsymbol{X}}(\boldsymbol{x}) = \int_{-\infty}^{\infty} p_{\boldsymbol{X}|\boldsymbol{N}}(\boldsymbol{x}|\boldsymbol{n})\,p_{\boldsymbol{N}}(\boldsymbol{n})\,\mathrm{d}\boldsymbol{n}$$

$$= \int_{-\infty}^{\infty} p_{\boldsymbol{S}}(\boldsymbol{u} - \boldsymbol{W}n)|\det \boldsymbol{W}|\,p_{\boldsymbol{N}}(\boldsymbol{n})\,\mathrm{d}\boldsymbol{n}. \tag{4}$$

Owing to the mutual independence assumption of the sources, we can factorize the probability density function (pdf) of the source signals $p_{\boldsymbol{S}}(\boldsymbol{s}) = \prod_{i=1}^{M_\mathrm{s}} p_{S_i}(s_i)$ with $p_{S_i}(\cdot)$ denoting the pdf of the ith source. Likewise, for the noise vector we have $p_{\boldsymbol{N}}(\boldsymbol{n}) = \prod_{k=1}^{M_\mathrm{s}} p_{N_k}(n_k)$ with $p_{N_k}(\cdot)$ being the pdf of the noise source at sensor k. In many communication applications, the dominant noise is thermal noise, whose distribution is known. The log-likelihood function L is then given by the logarithm of this probability density

$$L = \ln p_{\boldsymbol{X}}(\boldsymbol{x})$$

$$= \ln|\det \boldsymbol{W}| + \ln \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \prod_{i=1}^{M_\mathrm{s}} p_{S_i}(u_i - \boldsymbol{w}_i^\mathrm{T}\boldsymbol{n}) \prod_{k=1}^{M_\mathrm{s}} p_{N_k}(n_k)\,\mathrm{d}n_1 \ldots \mathrm{d}n_{M_\mathrm{s}}, \tag{5}$$

where $\boldsymbol{w}_i^\mathrm{T}$ is the ith row of matrix $\boldsymbol{W}$. As with most ML-related solutions, Eq. (5) is difficult to solve directly. An adaptive solution using a gradient-search method is usually sought to overcome this problem. In order to find a gradient leading to the ML solution, we have to differentiate L w.r.t. matrix $\boldsymbol{W}$. We write this gradient elementwise

$$\frac{\partial L}{\partial w_{mn}} = [\boldsymbol{W}^{-\mathrm{T}}]_{mn}$$

$$+ \frac{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p'_{S_m}(u_m - \boldsymbol{w}_m^\mathrm{T}\boldsymbol{n}) \cdot (x_n - n_n) \prod_{\substack{i=1\\ i\neq m}}^{M_\mathrm{s}} p_{S_i}(u_i - \boldsymbol{w}_i^\mathrm{T}\boldsymbol{n}) \prod_{k=1}^{M_\mathrm{s}} p_{N_k}(n_k)\,\mathrm{d}n_1 \ldots \mathrm{d}n_{M_\mathrm{s}}}{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \prod_{i=1}^{M_\mathrm{s}} p_{S_i}(u_i - \boldsymbol{w}_i^\mathrm{T}\boldsymbol{n}) \prod_{k=1}^{M_\mathrm{s}} p_{N_k}(n_k)\,\mathrm{d}n_1 \ldots \mathrm{d}n_{M_\mathrm{s}}}, \tag{6}$$

where $[\boldsymbol{W}^{-\mathrm{T}}]_{mn}$ is the (m,n)th entry in $\boldsymbol{W}^{-\mathrm{T}}$. In practice, Eq. (6) is too complicated for a real-time implementation. If the noise is negligible, however, Eq. (6) turns into a much easier form

$$\frac{\partial L}{\partial w_{mn}} = [\boldsymbol{W}^{-\mathrm{T}}]_{mn} + \frac{p'_{S_m}(u_m)}{p_{S_m}(u_m)} \cdot x_n \tag{7}$$

$$= [\boldsymbol{W}^{-\mathrm{T}}]_{mn} - g_m(u_m)x_n, \tag{8}$$

where

$$g_i(u_i) = -\frac{\partial\, log\; p_{S_i}(u_i)}{\partial u_i} = -\frac{p'_{S_i}(u_i)}{p_{S_i}(u_i)}, \quad i = 1, \ldots, M_\mathrm{s} \tag{9}$$

with $p_{S_i}(u_i)$ and $p'_{S_i}(u_i)$ being the source pdf and its derivative, respectively. $g(\cdot)$ is called the score function associated with a certain pdf. A possible update equation for the separation matrix using a stochastic gradient can now be formulated from Eq. (8) as

$$\boldsymbol{W}_{t+1} = \boldsymbol{W}_t + \mu(\boldsymbol{W}_t^{-\mathrm{T}} - \boldsymbol{g}(\boldsymbol{u})\boldsymbol{x}^{\mathrm{T}}), \tag{10}$$

where μ is a small step size, and $\boldsymbol{g}(\boldsymbol{u}) = [g_1(u_1), \ldots, g_{M_s}(u_{M_s})]^{\mathrm{T}}$ is the vector of score functions. The direct derivation of Eq. (10) in the noiseless case was also given by Yang [20] using an ML approach. Rather interestingly, other approaches, such as the InfoMax [4] or the minimization of the mutual information [3] lead to the same solution.

The convergence of Eq. (10) is not very fast and depends on the mixing matrix $\boldsymbol{A}$ as well as on the initial $\boldsymbol{W}_{t=0}$. Besides, the implementation of Eq. (10) involves a matrix inversion, an operation that should be avoided for fast real-time algorithms. Possible ways out of these problems were presented by Amari [1] and Cardoso [7] by using the natural gradient and the relative gradient, respectively. The natural gradient corrects for the nonisotropic gradient magnitude structure—called Riemannian structure in information geometry—in the parameter space of the standard-gradient adaptation, but at the same time preserves local minima of the cost function. For the blind separation problem the natural gradient (as well as the relative gradient) method involves a post-multiplication of the matrix update by $\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W}$, hence Eq. (10) becomes

$$\boldsymbol{W}_{t+1} = \boldsymbol{W}_t + \mu(\boldsymbol{I} - \boldsymbol{g}(\boldsymbol{u})\boldsymbol{u}^{\mathrm{T}})\boldsymbol{W}_t, \tag{11}$$

thereby getting rid of the matrix inversion. Moreover, the convergence speed of Eq. (11) is considerably improved over the original update equation, Eq. (10). A comparison between Eqs. (10) and (11) and further details on the natural gradient and its properties are given in [11].

4. The threshold nonlinearity

4.1. Derivation

Many source signals, particularly in communications are modeled by a uniform distribution. In the following, we derive a suitable nonlinearity for uniformly distributed source signals. The uniform distribution is a special case of a larger family of distributions, the generalized Gaussian distributions, whose pdf is given by

$$p_S(u_i) = \frac{\alpha}{2\beta\Gamma(\frac{1}{\alpha})}\, \mathrm{e}^{-(|u_i|/\beta)^{\alpha}}. \tag{12}$$

$p_S(u_i)$ models super-Gaussian distributions for $0 < \alpha < 2$ and sub-Gaussian distributions for $\alpha > 2$, respectively.

Differentiating Eq. (12) with respect to u_i leads to

$$p_S'(u_i) = -\alpha \left(\frac{|u_i|}{\beta}\right)^{\alpha-1} \frac{\mathrm{sign}(u_i)}{\beta}\, \frac{\alpha}{2\beta\Gamma\left(\frac{1}{\alpha}\right)}\, \mathrm{e}^{-(|u_i|/\beta)^{\alpha}}. \tag{13}$$

If we divide Eq. (13) by Eq. (12) and flip the sign we get

$$\begin{aligned} g_i(u_i) &= -\frac{p_S'(u_i)}{p_S(u_i)} = \alpha \left(\frac{|u_i|}{\beta}\right)^{\alpha-1} \frac{\text{sign}(u_i)}{\beta} \\ &= \frac{\alpha}{\beta^\alpha} |u_i|^{\alpha-1} \,\text{sign}(u_i). \end{aligned} \tag{14}$$

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

$$E\{|X|^m\} = \frac{\Gamma((m+1)/\alpha)}{\Gamma(1/\alpha)}\, \beta^m. \tag{15}$$

$\Gamma(\cdot)$ is the gamma function given by $\Gamma(a) = \int_0^\infty x^{a-1} \mathrm{e}^{-x}\,\mathrm{d}x$, and shows a recursive property similar to the factorial function, $\Gamma(a+1) = a\Gamma(a)$. For $m = 2$, Eq. (15) yields

$$\beta = \sqrt{\frac{\Gamma(1/\alpha)}{\Gamma(3/\alpha)}}. \tag{16}$$

Inserting this value for β into Eq. (14) yields the nonlinear function

$$g_i(u_i) = \alpha \left(\frac{\Gamma(3/\alpha)}{\Gamma(1/\alpha)}\right)^{\alpha/2} \text{sign}(u_i) \cdot |u_i|^{\alpha-1}. \tag{17}$$

Eq. (17) is the score function for any generalized unit-variance Gaussian distribution. Using $\Gamma(x) \cdot \Gamma(1-x) = \pi/\sin(\pi x)$ (see for example [6]) leads to

$$g_i(u_i) = \alpha \left(\frac{\pi/\sin 3\pi/\alpha}{\pi/\sin \pi/\alpha} \cdot \frac{\Gamma(1-1/\alpha)}{\Gamma(1-3/\alpha)}\right)^{\alpha/2} \text{sign}(u_i) \cdot |u_i|^{\alpha-1}. \tag{18}$$

Both terms $\Gamma(1-1/\alpha)$ and $\Gamma(1-3/\alpha)$ are close to $\Gamma(1) = 1$ for large values of α, so that simplification of Eq. (18) yields

$$g_i(u_i)|_{\alpha \gg 1} \approx \alpha \left(\frac{\sin(\pi/\alpha)}{\sin(3\pi/\alpha)}\right)^{\alpha/2} \text{sign}(u_i) \cdot |u_i|^{\alpha-1}. \tag{19}$$

The first term of the Taylor expansion of a sine function for a small argument is just the argument itself, leading to

$$g_i(u_i)|_{\alpha \gg 1} \approx \alpha \left(\frac{1}{3}\right)^{\alpha/2} \text{sign}(u_i) \cdot |u_i|^{\alpha-1} = \alpha \frac{1}{u_i} \left(\frac{u_i^2}{3}\right)^{\alpha/2}. \tag{20}$$

We are now interested in the form of $g_i(\cdot)$ as α approaches infinity, in which case Eq. (12) corresponds to a uniform distribution. As a consequence of the behavior of $\lim_{b\to\infty} a^b$ depending on $|a|$ being less or greater than one, we can write the *threshold nonlinearity* as

$$\lim_{\alpha\to\infty} g_i(u_i) = \begin{cases} 0, & |u_i| < \sqrt{3}, \\ \infty \cdot \text{sign}(u_i), & |u_i| \geqslant \sqrt{3}. \end{cases} \tag{21}$$

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, $\boldsymbol{W}_{t+1}$ in Eq. (11) 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 Eq. (21) will of course cause convergence problems for a finite learning rate parameter μ. The gain can, therefore, be traded off against a lower threshold ϑ for a specified output power. Again, if we aim at unity output power, we need to scale the nonlinearity. Hence, for every component u_i of the vector $\boldsymbol{u}$, $i = 1, \dots, M_s$, we need to scale $g_i(u_i)$ such that the scaling constraint of the nonlinearity

$$\int_{-\infty}^{\infty} p_S(u) g(u) u \, \mathrm{d}u = 1 \tag{22}$$

is satisfied if $p_S(\cdot)$ is a source distribution with unit variance $\sigma_S^2 = 1$. By satisfying Eq. (22), the output power of $\boldsymbol{u}$ will become normalized after convergence

$$E\{\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}\} = \boldsymbol{I}. \tag{23}$$

Replacing Eq. (21) with

$$g_i(u_i) = \begin{cases} 0, & |u_i| < \vartheta, \\ a \, \mathrm{sign}(u_i), & |u_i| \geqslant \vartheta \end{cases} \tag{24}$$

we get the gain a of the threshold nonlinearity for uniform distributions by solving Eq. (22) as

$$a = \frac{2\sqrt{3}}{3 - \vartheta^2} \tag{25}$$

for $0 \leqslant \vartheta < \sqrt{3}$. The resulting threshold nonlinearity is depicted in Fig. 3. Its form represents a three-step quantizer. Note that a is always positive for the assigned range of ϑ. Although the threshold nonlinearity has been derived for the uniform distribution, [17] shows that, in fact, all sub-Gaussian signals can be separated using this nonlinearity. By adapting the threshold parameter, even super-Gaussian distributions might be separated. Moreover, it can be shown [15] that the threshold nonlinearity separates any non-Gaussian distribution, provided the threshold value is set correctly.

4.2. Stability regions of the threshold nonlinearity

The local stability of the threshold nonlinearity has only been proven explicitly for continuous distributions [17]. In the following, this local stability analysis is extended to discrete distributions. The condition for local stability for the threshold nonlinearity under the assumption of equal source distribution and nonlinearities is [17]

$$\frac{p_{U_i}(\vartheta)}{\int_{\vartheta}^{\infty} p_{U_i}(u_i) u_i \, \mathrm{d}u_i} > 1, \quad i = 1 \dots M_s \tag{26}$$

with M_s denoting the number of sources. Whereas the integral in the denominator of Eq. (26) can be written as a sum for discrete distributions

$$\int_{\vartheta}^{\infty} p_{U_i}(u_i) u_i \, \mathrm{d}u_i = \sum_{k, A_k \geqslant \vartheta} \Pr(u_i = A_k) A_k, \tag{27}$$

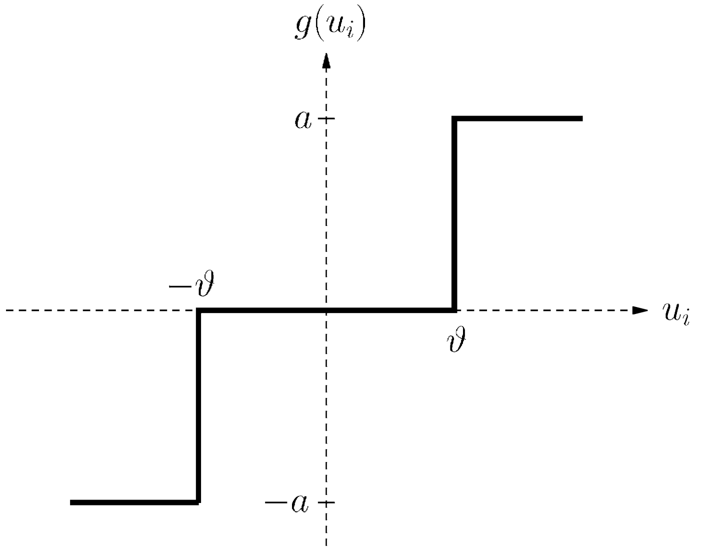

Fig. 3. Threshold nonlinearity with parameter ϑ.

the evaluation of a probability density, as appearing in the numerator of Eq. (26), needs a closer look. Close to an equilibrium point we may model the output distribution as a convolution of the discrete probability model of the source signals by some mixing noise distribution, which is Gaussian distributed. The probability density at a certain constellation point is, therefore, the discrete probability of that point multiplied by the mode of the Gaussian kernel $1/\sqrt{2\pi}\sigma_{\mathrm{N}}$, with σ_{N}^2 being the variance of the mixing noise. In other words, the discrete-level distribution is convolved with the probability density function (pdf) of a Gaussian noise signal. Examples of such discrete-level distributions are M-ary pulse amplitude modulation (M-PAM), essentially data communication signaling schemes, which have an alphabet size of M different, equally spaced, and equally probable amplitudes (cf. top of Fig. 4). The resulting pdf for a 4-PAM (pulse amplitude modulation) signal with an SNR=25 dB is depicted at the bottom of Fig. 4. The stability regions are thus dependent on the mixing noise. Figs. 5 and 6 show the stable regions as derived from the evaluation of Eq. (26) for binary phase shift keying (BPSK) and 4-PAM, respectively. It is interesting to note that in addition to the region around the outer symbols, which looks similar for BPSK and 4-PAM, there is a further stable region around the inner symbols in the case of 4-PAM.

It becomes apparent that for a stable update equation for BPSK signals, the threshold ϑ has to be in the neighborhood of the symbol amplitude, otherwise the algorithm becomes unstable. A closer look at Eq. (26) reveals that the mixing noise keeps the algorithm stable through a finite pdf in the neighborhood of the symbol amplitude. In other words, if the threshold ϑ is chosen too far away from the symbol amplitude, more mixing noise is needed to satisfy Eq. (26). For BPSK, the threshold ϑ should, therefore, be chosen directly at the symbol amplitude $A_1 = 1$. For this choice, with

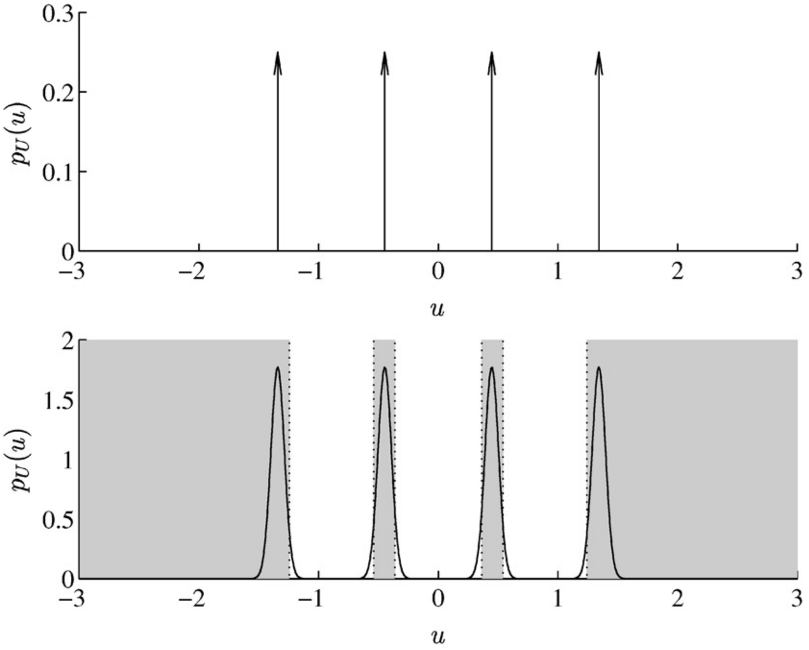

Fig. 4. Top: discrete distribution of 4-PAM signal with unit variance. Bottom: pdf of 4-PAM signal with additive Gaussian noise, SNR=25 dB. The shaded regions indicate the stable region of the threshold parameter ϑ as derived from Fig. 6.

probability 0.5 the signal will be larger (smaller) than the threshold, enforcing a choice of the scaling factor $a = 2$ in order to satisfy the scaling condition, Eq. (22). For all choices of the threshold ϑ smaller than $A_1 = 1$ and low residual mixing, a scaling factor of $a = 1$ is needed. For larger threshold values, the gain gets impractically high due to Eq. (22). For M-PAM signals with $M > 2$, stable algorithms can be obtained by setting the threshold to the outermost symbol amplitude

$$\vartheta = \sqrt{\frac{3(M-1)}{M+1}}. \tag{28}$$

The corresponding gain is

$$a = M\sqrt{\frac{M+1}{3(M-1)}}. \tag{29}$$

4.3. Bias removal for the threshold nonlinearity

Algorithms of the form given by Eq. (11) lead to a biased solution, if additive noise is present at the sensors. Particularly in communication environments we often have this situation of additive noise. By linearly combining the signals in order to separate them, the noise signals get correlated at the output, introducing dependencies between

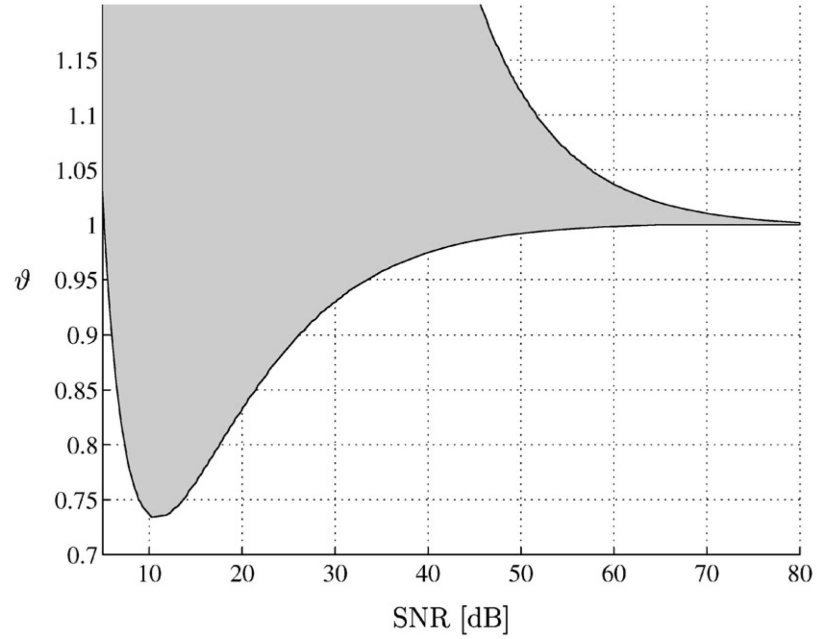

Fig. 5. Stable region (shaded) for noisy BPSK signals and the threshold nonlinearity.

the sensor signals. Any criterion that searches for the minimum dependence among the output signals will therefore deviate from this solution, thereby introducing a bias. A combined learning process involving unsupervised learning for the separation and supervised learning for noise reduction was presented in [9]. The lack of a noise reference in practice, however, makes this approach inapplicable to most common problems. It is possible to devise an update equation with an additional term in the update equation, which involves either some expectation of the signal derivatives [8] or their stochastic versions [12]

$$\boldsymbol{W}_{t+1} = \boldsymbol{W}_t + \mu(\boldsymbol{I} - \boldsymbol{g}(\boldsymbol{u})\boldsymbol{u}^{\mathrm{T}} + \boldsymbol{B}\boldsymbol{W}_t\boldsymbol{R}_{\mathrm{N}}\boldsymbol{W}_t^{\mathrm{T}})\boldsymbol{W}_t, \tag{30}$$

where $\boldsymbol{B}$ is a diagonal matrix with entries

$$b_{ii} = E\left\{\frac{\mathrm{d}g(u_i)}{\mathrm{d}u_i}\right\} \tag{31}$$

and $\boldsymbol{R}_{\mathrm{N}}$ is the covariance matrix of the noise contribution. To see the mechanism behind Eq. (30), we define an unbiased estimate of the source signal as

$$\hat{\boldsymbol{u}} = \boldsymbol{W}\boldsymbol{A}\boldsymbol{s}. \tag{32}$$

If we use the original algorithm, Eq. (11), to separate a noisy mixture of signals, we get an equilibrium point when the expectation of the parenthesis is zero, hence

$$E\{\boldsymbol{I} - \boldsymbol{g}(\boldsymbol{u})\boldsymbol{u}^{\mathrm{T}}\} = \boldsymbol{0}. \tag{33}$$

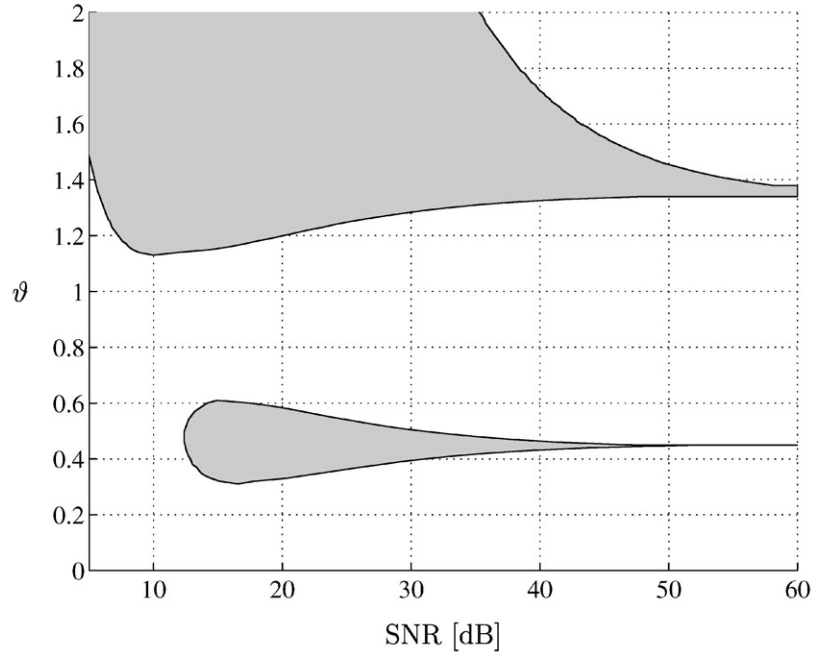

Fig. 6. Stable regions (shaded) for noisy 4-PAM signals and the threshold nonlinearity.

But since the output is noisy, i.e.,

$$\boldsymbol{u} = \hat{\boldsymbol{u}} + \boldsymbol{W}\boldsymbol{n}, \tag{34}$$

we get from Eq. (33)

$$E\{\boldsymbol{I} - \boldsymbol{g}(\hat{\boldsymbol{u}} + \boldsymbol{W}\boldsymbol{n})(\hat{\boldsymbol{u}} + \boldsymbol{W}\boldsymbol{n})^{\mathrm{T}}\} = \boldsymbol{0}. \tag{35}$$

A first-order truncated Taylor series expansion of the nonlinearity around $\hat{\boldsymbol{u}}$ yields

$$\boldsymbol{g}(\hat{\boldsymbol{u}} + \boldsymbol{W}\boldsymbol{n}) = \boldsymbol{g}(\hat{\boldsymbol{u}}) + \operatorname{diag}(\boldsymbol{g}'(\hat{\boldsymbol{u}}))\boldsymbol{W}\boldsymbol{n}, \tag{36}$$

where $\operatorname{diag}(\boldsymbol{g}'(\hat{\boldsymbol{u}}))$ is a diagonal matrix with the elements $g'(\hat{u}_i)$ located on the diagonal. Inserted into Eq. (35), this results in

$$\begin{aligned} &E\{\boldsymbol{I} - \boldsymbol{g}(\hat{\boldsymbol{u}} + \boldsymbol{W}\boldsymbol{n})(\hat{\boldsymbol{u}} + \boldsymbol{W}\boldsymbol{n})^{\mathrm{T}}\} \\ &\quad = \boldsymbol{I} - E\{\boldsymbol{g}(\hat{\boldsymbol{u}})\hat{\boldsymbol{u}}^{\mathrm{T}}\} - E\{\operatorname{diag}(\boldsymbol{g}'(\hat{\boldsymbol{u}}))\boldsymbol{W}\boldsymbol{n}\hat{\boldsymbol{u}}^{\mathrm{T}}\} \\ &\qquad - E\{\boldsymbol{g}(\hat{\boldsymbol{u}})\boldsymbol{n}^{\mathrm{T}}\boldsymbol{W}^{\mathrm{T}}\} - E\{\operatorname{diag}(\boldsymbol{g}'(\hat{\boldsymbol{u}}))\boldsymbol{W}\boldsymbol{n}\boldsymbol{n}^{\mathrm{T}}\boldsymbol{W}^{\mathrm{T}}\} = \boldsymbol{0}. \end{aligned} \tag{37}$$

Since the noiseless estimate is uncorrelated to the noise, the third and the fourth term of the RHS of Eq. (37) are zero, hence

$$\begin{aligned} &E\{\boldsymbol{I} - \boldsymbol{g}(\hat{\boldsymbol{u}} + \boldsymbol{W}\boldsymbol{n})(\hat{\boldsymbol{u}} + \boldsymbol{W}\boldsymbol{n})^{\mathrm{T}}\} \\ &\quad = \boldsymbol{I} - E\{\boldsymbol{g}(\hat{\boldsymbol{u}})\hat{\boldsymbol{u}}^{\mathrm{T}}\} - E\{\operatorname{diag}(\boldsymbol{g}'(\hat{\boldsymbol{u}}))\boldsymbol{W}\boldsymbol{R}_N\boldsymbol{W}^{\mathrm{T}}\} = \boldsymbol{0}. \end{aligned} \tag{38}$$

The equilibrium point is therefore the point where the above equation is satisfied, and not the point at which the unbiased estimate of the source signals $\hat{\boldsymbol{u}}$ are independent. The third term of the RHS of Eq. (38) is now identified as the bias term and has to be subtracted in the original update equation, leading to Eq. (30). If we then make the same analysis on Eq. (30), for which we know that at the equilibrium we have

$$E\{\boldsymbol{I} - \boldsymbol{g}(\boldsymbol{u})\boldsymbol{u}^{\mathrm{T}} + \boldsymbol{B}\boldsymbol{W}_t\boldsymbol{R}_{\mathrm{N}}\boldsymbol{W}_t^{\mathrm{T}}\} = \boldsymbol{0}, \tag{39}$$

we get

$$\begin{aligned} &E\{\boldsymbol{I} - \boldsymbol{g}(\boldsymbol{u})\boldsymbol{u}^{\mathrm{T}} + \boldsymbol{B}\boldsymbol{W}_t\boldsymbol{R}_N\boldsymbol{W}_t^{\mathrm{T}}\} && (40)\\ &\quad = E\{\boldsymbol{I} - \boldsymbol{g}(\hat{\boldsymbol{u}} + \boldsymbol{W}\boldsymbol{n})(\hat{\boldsymbol{u}} + \boldsymbol{W}\boldsymbol{n})^{\mathrm{T}} + \boldsymbol{B}\boldsymbol{W}_t\boldsymbol{R}_{\mathrm{N}}\boldsymbol{W}_t^{\mathrm{T}}\}\\ &\quad = \boldsymbol{I} - E\{\boldsymbol{g}(\hat{\boldsymbol{u}})\hat{\boldsymbol{u}}^{\mathrm{T}}\} - E\{\mathrm{diag}(\boldsymbol{g}'(\hat{\boldsymbol{u}}))\boldsymbol{W}R_{\mathrm{N}}\boldsymbol{W}^{\mathrm{T}}\} + \boldsymbol{B}W_t\boldsymbol{R}_{\mathrm{N}}\boldsymbol{W}_t^{\mathrm{T}}\\ &\quad = \boldsymbol{I} - E\{\boldsymbol{g}(\hat{\boldsymbol{u}})\hat{\boldsymbol{u}}^{\mathrm{T}}\} = \boldsymbol{0}. && (41) \end{aligned}$$

Hence, the equilibrium means that the elements of $\hat{\boldsymbol{u}}$ will be mutually independent.

Although the threshold nonlinearity is nondifferentiable, its expectation can be expressed by integration over a Dirac impulse

$$\begin{aligned} E\{g'(\hat{u}_i)\} &= \int_{-\infty}^{\infty} p_{\hat{U}_i}(\hat{u}_i)g'(\hat{u}_i)\,\mathrm{d}\hat{u}_i\\ &= \int_{-\infty}^{\infty} p_{\hat{U}_i}(\hat{u}_i)a(\delta(u_i + \vartheta) + \delta(\hat{u}_i - \vartheta))\,\mathrm{d}\hat{u}_i\\ &= 2a \cdot p_{\hat{U}_i}(\vartheta). \end{aligned} \tag{42}$$

In the following we assume equal noise power σ_{N}^2 at each of the sensors, but uncorrelated noise signals, so that the sensor noise vector is described by $\mathcal{N}(0, \sigma_{\mathrm{N}}^2 \cdot \boldsymbol{I})$, or by $\boldsymbol{R}_{\mathrm{N}} = \sigma_{\mathrm{N}}^2 \cdot \boldsymbol{I}$. This is a reasonable assumption, as very often noise is of thermal origin, therefore, given by temperature and noise figure and as such of equal variance but mutually uncorrelated for all the channels. Furthermore, the noise power σ_{N}^2 is presumed to be known, be that from theoretical calculations of thermal noise or by estimating it, e.g., using minor component analysis in an overdetermined separation case [14].

For identical distributions of all source signals, Eq. (30) can be simplified to

$$\boldsymbol{W}_{t+1} = \boldsymbol{W}_t + \mu(\boldsymbol{I} - \boldsymbol{g}(\boldsymbol{u})\boldsymbol{u}^{\mathrm{T}} + \sigma_{\mathrm{N}}^2 b\boldsymbol{W}_t\boldsymbol{W}_t^{\mathrm{T}})\boldsymbol{W}_t, \tag{43}$$

where

$$b = E\{g'(\hat{u})\}. \tag{44}$$

For the uniform distribution, which is a good approximation for M-ary distributions where M is high, with unit variance, implying that the threshold function is properly scaled according to Eq. (25) we get

$$b = E\{g'(\hat{u})\} = \frac{2}{3 - \vartheta^2}. \tag{45}$$

If the source signals have discrete distributions rather than continuous ones, the update equation, Eq. (43), is not accurate, as it is based on the assumption of uniform distribution. Owing to its discrete distribution, $p_{\hat{U}_i}(\cdot)$, which should be used in

Eq. (42), is a probability mass function (pmf) rather than a pdf. However, close to the real solution we may approximate $p_{\tilde{U}_i}(\cdot)$ by $p_{U_i}(\cdot)$, which is a true pdf. Since the noise carried through to the outputs determines the probability density at the threshold level ϑ, $p_{U_i}(\vartheta)$ depends on the separation matrix. For an M-ary signaling scheme (e.g., M-PAM) we can write for the probability density at the ith output

$$p_{U_i}(\vartheta) = \frac{1}{M}\,\frac{1}{\sqrt{2\pi}\sigma_{\mathrm{N}}}\,\frac{1}{\sqrt{\sum_{k=1}^{M_s} w_{ik}^2}} \tag{46}$$

with w_{ik} being the i,kth element of the separation matrix $\boldsymbol{W}$, describing the path from the kth sensor to the ith output and ϑ given by Eq. (28). For M-PAM signals, using Eqs. (29) and (46) in Eq. (42) and the update equation, Eq. (30), we get

$$\boldsymbol{W}_{t+1} = \boldsymbol{W}_t + \mu\left(\boldsymbol{I} - \boldsymbol{g}(\boldsymbol{u})\boldsymbol{u}^{\mathrm{T}} + \sqrt{\frac{2(M+1)}{3\pi(M-1)}}\sigma_{\mathrm{N}}(\mathrm{diag}(\boldsymbol{W}_t\boldsymbol{W}_t^{\mathrm{T}}))^{-1/2}\boldsymbol{W}_t\boldsymbol{W}_t^{\mathrm{T}}\right)\boldsymbol{W}_t, \tag{47}$$

where diag$(\boldsymbol{W}_t\boldsymbol{W}_t^{\mathrm{T}})$ means here the matrix $\boldsymbol{W}_t\boldsymbol{W}_t^{\mathrm{T}}$ with suppressed off-diagonal terms and may also be written by the use of the Hadamard or Schur product: $\mathrm{diag}(\boldsymbol{W}_t\boldsymbol{W}_t^{\mathrm{T}}) = \boldsymbol{I}\circ(\boldsymbol{W}_t\boldsymbol{W}_t^{\mathrm{T}})$, where $\circ$ denotes elementwise multiplication.

4.4. MMSE vs. zero-forcing solution

Very often in data communications we are not interested in the solution of $\boldsymbol{W}$ that directly inverts $\boldsymbol{A}$—the so-called zero-forcing solution—due to problems associated with noise enhancement at frequencies close to zeros of the system transfer function. In terms of signal purity—the essence of low bit-error rates—we do not care where unwanted contributions to the signal comes from; signals from other channels or thermal noise. This is of course only the case if channels are not jointly detected. For single-channel detection, the proper criterion to choose is the minimum mean squared error (MMSE). If we have a zero-forcing solution $\boldsymbol{W}_{\mathrm{ZF}}$ we can, by looking at the MMSE solution for unit-power source signals [14]

$$\boldsymbol{W}_{\mathrm{MMSE}} = \boldsymbol{A}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}} + \sigma_n^2\boldsymbol{I})^{-1} \tag{48}$$

reformulate the MMSE solution in terms of the zero-forcing solution. To this end, we note that the zero-forcing solution is the inverse of the system matrix but for some permutation and sign flippings

$$\boldsymbol{W}_{\mathrm{ZF}} = \boldsymbol{J}\boldsymbol{P}\boldsymbol{A}^{-1}. \tag{49}$$

$\boldsymbol{J}$ is a matrix of ±1s and $\boldsymbol{P}$ is a permutation matrix. Using Eq. (49) in Eq. (48) leads to

$$\boldsymbol{W}_{\mathrm{MMSE}} = \boldsymbol{P}^{\mathrm{T}}\boldsymbol{J}^{\mathrm{T}}\boldsymbol{W}_{\mathrm{ZF}}^{-\mathrm{T}}(\boldsymbol{W}_{\mathrm{ZF}}^{-1}\boldsymbol{J}\boldsymbol{P}\boldsymbol{P}^{\mathrm{T}}\boldsymbol{J}^{\mathrm{T}}\boldsymbol{W}_{\mathrm{ZF}}^{-\mathrm{T}} + \sigma_n^2\boldsymbol{I})^{-1}. \tag{50}$$

$\boldsymbol{J}\boldsymbol{P}\boldsymbol{P}^{\mathrm{T}}\boldsymbol{J}^{\mathrm{T}}=\boldsymbol{I}$, and by premultiplying the solution in Eq. (50) by $\boldsymbol{J}\boldsymbol{P}$ we do not challenge its validity, so we get

$$\boldsymbol{W}_{\mathrm{MMSE}} = \boldsymbol{W}_{\mathrm{ZF}}^{-\mathrm{T}}(\boldsymbol{W}_{\mathrm{ZF}}^{-1}\boldsymbol{W}_{\mathrm{ZF}}^{-\mathrm{T}} + \sigma_n^2\boldsymbol{I})^{-1}. \tag{51}$$

4.5. Computer simulations

In the following, results of computer simulations of the blind separation using the bias-removal method suggested above are shown. Some important parameters influencing the performance were taken from [12], such as the number of sources and sensors $M_{\mathrm{s}} = 3$, the mixing matrix

$$\boldsymbol{A} = \begin{bmatrix} 0.4 & 1.0 & 0.7 \\ 0.6 & 0.5 & 0.5 \\ 0.3 & 0.7 & 0.2 \end{bmatrix} \tag{52}$$

and the condition $\boldsymbol{W}_{\mathrm{o}}\boldsymbol{W}_{\mathrm{o}}^{\mathrm{T}}=0.25\cdot\boldsymbol{I}$ (implying that $\boldsymbol{W}_{\mathrm{o}}$ is a scaled orthogonal matrix) for the one hundred trials with a different initial separation matrix. In the first experiment three uniformly distributed source signals were mixed, and noise was added at the sensors with $\sigma_{\mathrm{N}}^2 = 0.01$. The noise level was assumed to be known to the algorithm. The mixed noisy signals were then separated using the threshold nonlinearity with $\vartheta = 1.5$ and the update equation, Eq. (11). The step size was adjusted without noise to obtain an interchannel interference level of −35 dB and then fixed to $\mu = 0.00032$ for the other simulations. The performance measure used in the plots is calculated as a function of the global system matrix $\mathrm{P} = [p_{ik}]$

$$J_{\mathrm{ICI}}(\boldsymbol{P}) = \frac{1}{M_{\mathrm{s}}}\left(\sum_{i=1}^{M_{\mathrm{s}}} \frac{\sum_{k=1}^{M_{\mathrm{s}}} p_{ik}^2}{\max_k p_{ik}^2}\right) - 1 \tag{53}$$

and expresses the average interchannel interference. Fig. 7 reveals the convergence improvement of the modified algorithm compared to the standard algorithm without bias removal. Since mixing matrix and noise power are identical to the parameters chosen in [8,12], we can compare the convergence with those results directly. The curves shown in Fig. 7 look almost identical to the curves given in [8,12]. The advantage in this method here lies in the application of a much simpler nonlinearity, essentially a three-step quantizer.

Still better results were obtained for binary-distributed source signals. Three binary-distributed source signals were mixed using the same mixing matrix as above. To show clearer differences between the algorithms, the noise was increased by 5 dB, resulting in $\sigma_{\mathrm{N}}^2 = 0.0316$. Fig. 8 shows that the modified algorithm is in fact capable of completely removing any bias, albeit at a lower convergence speed. Again, step sizes were chosen equal ($\mu = 0.0018$) for all three cases. It was also observed that the modified algorithm with certain noise levels (e.g., $\sigma_{\mathrm{N}}^2 = 0.01$) consistently outperformed the standard algorithm with no noise. This surprising effect is due to an increased stability region (see Fig. 5) for lower SNRs. The additive noise has then a positive dithering effect. With other nonlinearities (e.g., $g(u) = au^3$) or other distributions (e.g., uniform distribution), this effect cannot be observed. It is only the special arrangement

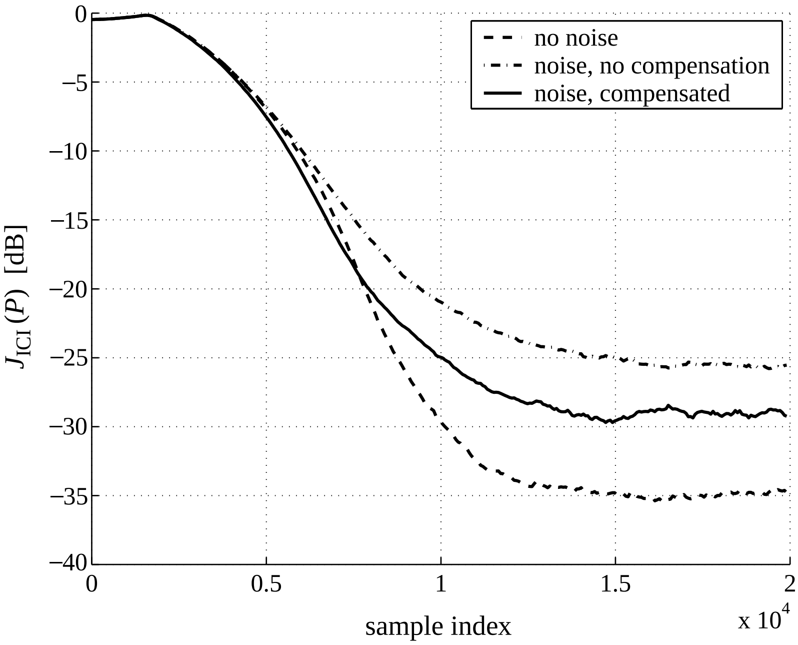

Fig. 7. Separation convergence of bias removal algorithm for uniform distributions.

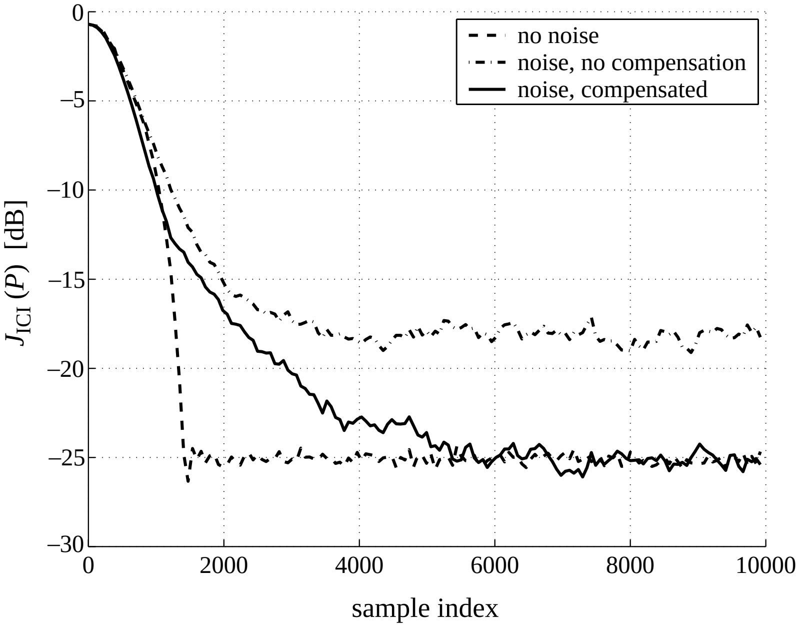

Fig. 8. Separation convergence of bias removal algorithm for binary distributions.

of high derivative of the nonlinearity at the level of spikes in the pdf that benefits from additional noise.

5. Conclusions

We have presented ways of overcoming the problems associated with excessive noise on the transfer channel of an instantaneous mixture of signals, when trying to blindly separate them. When more mixture observations are available, the noise space should be suppressed using preprocessing steps such as PCA. Algorithms based on simple nonlinearities such as a three-step quantizer can be extended to take into account additive noise, resulting in solutions of the estimate of the separation matrix with suppressed bias. Simulation results support the theory presented. From an unbiased separation solution, which satisfies a zero-forcing criterion, an MMSE solution can be readily obtained by simple matrix operations. The methods for bias removal shown can easily be extended to complex quadrature signals. Mathis et al. [16] give some hints as to how the update equations have to be modified.

Acknowledgements

The authors would like to thank all three reviewers for constructive comments on improving this manuscript.

References

[1] S.-I. Amari, Natural gradient works efficiently in learning, Neural Comput. 10 (1998) 251–276.

[2] S.-I. Amari, A. Cichocki, Adaptive blind signal processing—neural network approaches, Proc. IEEE 86 (10) (1998) 2026–2048.

[3] S.-I. Amari, A. Cichocki, H.H. Yang, A new learning algorithm for blind signal separation, Adv. Neural Inform. Process. Systems 8 (1996) 757–763.

[4] A.J. Bell, T.J. Sejnowski, An information-maximization approach to blind separation and blind deconvolution, Neural Comput. 7 (1995) 1129–1159.

[5] A. Belouchrani, K. Abed-Meraim, Constant modulus blind source separation technique: a new approach, in: International Symposium on Signal Processing and its Application (ISSPA), Goldcoast, Australia, August 1996, pp. 232–235.

[6] I.N. Bronshtein, K.A. Semendyayev, Handbook of Mathematics, 3rd Edition, Springer, Berlin, 1997.

[7] J.-F. Cardoso, The invariant approach to source separation, in: International Symposium on Nonlinear Theory and its Applications (NOLTA), Las Vegas, NV, December 10–14, 1995, pp. 55–60.

[8] A. Cichocki, S.C. Douglas, A. Amari, Robust techniques for independent component analysis (ICA) with noisy data, Neurocomputing 22 (1998) 113–129.

[9] A. Cichocki, W. Kasprzak, S. Amari, Adaptive approach to blind source separation with cancellation of additive and convolutional noise, in: International Conference on Signal Processing, Beijing, China, September 1996, pp. 412–415.

[10] S.C. Douglas, Combined subspace tracking, prewhitening, and contrast optimization for noisy blind signal separation, in: Proceedings of International Conference on Independent Component Analysis and Blind Signal Separation (ICA), Helsinki, Finland, June 19–22, 2000, pp. 579–584.

[11] S.C. Douglas, S.-I. Amari, Natural-gradient adaptation, in: S. Haykin (Ed.), Unsupervised Adaptive Filtering, Blind Source Separation, Vol. I, Wiley, New York, 2000, pp. 13–61.

[12] S.C. Douglas, A. Cichocki, S. Amari, Bias removal technique for blind source separation with noisy measurements, Electron. Lett. 34 (14) (1998) 1379–1380.

[13] A. Hyvärinen, Independent component analysis in the presence of Gaussian noise by maximizing joint likelihood, Neurocomputing 22 (1–3) (1998) 49–67.

[14] M. Joho, H. Mathis, R.H. Lambert, Overdetermined blind source separation: using more sensors than source signals in a noisy mixture, in: Proceedings of International Conference on Independent Component Analysis and Blind Signal Separation (ICA), Helsinki, Finland, June 19–22, 2000, pp. 81–86.

[15] H. Mathis, S.C. Douglas, On optimal and universal nonlinearities for blind signal separation, in: IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Salt Lake City, UT, May 7–11, 2001.

[16] H. Mathis, M. Joho, G.S. Moschytz, A simple threshold nonlinearity for blind signal separation, in: IEEE International Symposium on Circuits and Systems (ISCAS), Vol. IV, Geneva, Switzerland, May 28–31, 2000, pp. 489–492.

[17] H. Mathis, T.P. von Hoff, M. Joho, Blind separation of signals with mixed kurtosis signs using threshold activation functions, IEEE Trans. Neural Networks 12 (3) (2001) 618–624.

[18] E. Moulines, J.-F. Cardoso, E. Gassiat, Maximum likelihood for blind separation and deconvolution of noisy signals using mixture models, in: IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Munich, Germany, April 21–24, 1997, pp. 3617–3620.

[19] L. Xu, Bayesian Kullback Ying-Yang dependence reduction theory, Neurocomputing 22 (1–3) (1998) 81–111.

[20] H.H. Yang, Serial updating rule for blind separation derived from the method of scoring, IEEE Trans. Signal Process. 47 (8) (1999) 2279–2285.

[21] L.-Q. Zhang, A. Cichocki, S. Amari, Natural gradient algorithm for blind separation of overdetermined mixture with additive noise, IEEE Signal Process. Lett. 6 (11) (1999) 293–295.

Heinz Mathis was born in Zurich, Switzerland, in 1968. He received the Diploma in electrical engineering from the Swiss Federal Institute of Technology (ETH), Zurich, in 1993, and the Ph.D. in EE at the same University in 2001, respectively. From 1993 to 1997 he held jobs as a DSP and RF Engineer with different companies in Switzerland and England. From 1997 to 2001 he was a Research Assistant at the Signal and Information Processing Laboratory at ETH Zurich. He is currently a Professor for mobile communications at the University of Applied Sciences, Rapperswil, Switzerland. His research interests include blind separation and equalization for communications systems.

Marcel Joho was born in Zurich, Switzerland, in 1967. He received his Diploma degree in electrical engineering from the Swiss Federal Institute of Technology (ETH), Zurich, in 1993. From 1993 to 2000 he was a Research Assistant with the Signal and Information Processing Laboratory (ISI), ETH, Zurich, where he also received his Ph.D. degree. Since 2001, he has been an R&D Engineer with Phonak Inc., Champaign, IL, where he currently works on hearing aids. His research interests include adaptive beamforming, echo cancellation, and blind source separation.