On the Existence of Universal Nonlinearities for Blind Source Separation

Heinz Mathis, *Member, IEEE,* and Scott C. Douglas, *Senior Member, IEEE*

***Abstract*—Many density-based methods for blind signal separation employ one or more models for the unknown source distribution(s). This paper considers the issue of density model mismatch in maximum likelihood (ML)-type blind signal separation algorithms. We show that the score function nonlinearity, which was previously derived from the standpoint of statistical efficiency, is also the most robust in maintaining a separation solution for the ML algorithm class. We also consider the existence of a universally applicable nonlinearity for separating all signal types, deriving two results. First, among nonlinearities with a convergent Taylor series, a single fixed nonlinearity for universal separation using the natural gradient algorithm cannot exist. Second, among nonlinearities with a single adjustable parameter, a recently proposed threshold nonlinearity can separate all signals with symmetric amplitude distributions as long as the threshold parameter is properly chosen. The design of "difficult-to-separate" signal distributions is also discussed.**

***Index Terms*—Blind signal separation, nonlinear functions, threshold nonlinearity, universal nonlinearities.**

I. Introduction

MANY researchers have described adaptive algorithms for the blind separation of instantaneously mixed independent signals. Among these methods, density-based approaches employ distribution models for the amplitude statistics of the source signals. For a detailed tutorial on this aspect of blind signal separation, see [1]. For example, the maximum likelihood (ML) and information maximization (InfoMax) approaches to blind signal separation (BSS) assume that the source signal vector $\boldsymbol{s}_t = [s_{1,t}, s_{2,t}, \ldots, s_{M,t}]^T$ has the joint probability density function (p.d.f.)

$$p_{\boldsymbol{s}}(s_1, s_2, \ldots, s_m) = p_1(s_1)p_2(s_2)\cdots p_M(s_M) \tag{1}$$

where $p_i(s_i)$ is the p.d.f. of $s_{i,t}$. These source signals are instantaneously mixed by an unknown $(N \times M)$ mixing matrix $\boldsymbol{A}$ to yield the measured signal vector

$$\boldsymbol{x}_t = \boldsymbol{A}\boldsymbol{s}_t. \tag{2}$$

The goal of BSS is to extract estimates $\{u_{i,t}\}$ of the source signals using a linear demixing model

$$\boldsymbol{u}_t = \boldsymbol{W}_t\boldsymbol{x}_t = \boldsymbol{W}_t\boldsymbol{A}\boldsymbol{s}_t. \tag{3}$$

Gradient-based maximization of the ML cost yields the adaptive algorithm, e.g., [2]

$$\boldsymbol{W}_{t+1} = \boldsymbol{W}_t + \mu\left(\boldsymbol{W}_t^{-1} - \langle \boldsymbol{g}(\boldsymbol{u}_t)\boldsymbol{x}_t^T\rangle\right) \tag{4}$$

where $\langle\cdot\rangle$ is an averaging operator, and $\boldsymbol{g}(\boldsymbol{u})$ is a vector of nonlinearities whose forms depend on the p.d.f.s of the extracted output signals. For this paper, we restrict ourselves to zero-mean, symmetric source p.d.f.s.

These distribution models implicitly rely on higher or lower order moments of the underlying signal distributions. This reliance is indicated by the vector nonlinearity $\boldsymbol{g}(\boldsymbol{u})$ in the update in (4). For both the ML and InfoMax approach, the nonlinearity $g(u)$ is given by the score function

$$g(u) = -\frac{p'(u)}{p(u)} \tag{5}$$

where $p(u)$ and $p'(u)$ are the model distribution of the unknown sources and its derivative, respectively. Interestingly, the minimization of the mutual information between the outputs [3] leads to the same update equation (4) but with a different nonlinearity

$$g(u) = \tfrac{3}{4}\,u^{11} + \tfrac{15}{4}\,u^9 - \tfrac{14}{3}\,u^7 - \tfrac{29}{4}\,u^5 + \tfrac{29}{4}\,u^3. \tag{6}$$

Equation (6) is claimed to be model independent since the source distributions do not influence the nonlinearity, which is based on the approximation of marginal output densities by the Gram–Charlier expansion [3].

The choice of the nonlinearity $g(u)$ thus becomes a critical factor in the success of ML-type BSS methods. Generally, two issues have been of greatest concern in such situations.

- *Generality:* Does a universal nonlinearity exist that separates all distributions regardless of their form?
- *Robustness:* What nonlinearity yields the most robust performance in the presence of numerical effects (such as finite sample averages, model mismatch, and the like)?

In this paper, we focus on the design of $g(u)$ in ML-type BSS algorithms to address the above two issues. In particular, we define a concept of robustness of a nonlinearity and look for the most robust nonlinearity for a given distribution. On the other hand, we are interested in the existence of a fixed nonlinearity for all non-Gaussian signals, as addressed by Amari [3].

Manuscript received December 15, 2000; revised January 29, 2002. The associate editor coordinating the review of this paper and approving it for publication was Dr. Athina Petropulu.

H. Mathis was with the Signal and Information Processing Laboratory, Swiss Federal Institute of Technology, Zurich, Switzerland. He is now with the University of Applied Sciences, Rapperswil, Switzerland (e-mail: h.mathis@ieee.org).

S. C. Douglas is with the Department of Electrical Engineering, School of Engineering and Applied Science, Southern Methodist University, Dallas, TX 75275-0338 USA (e-mail:douglas@engr.smu.edu).

Publisher Item Identifier S 1053-587X(02)03223-3.

1053-587X/02$17.00 © 2002 IEEE

For all our investigations, we consider the natural gradient algorithm [3]

$$\boldsymbol{W}_{t+1} = \boldsymbol{W}_t + \mu\left(\boldsymbol{I} - \boldsymbol{g}(\boldsymbol{u})\boldsymbol{u}^T\right)\boldsymbol{W}_t \tag{7}$$

which is obtained by postmultiplying (4) by $\boldsymbol{W}^T\boldsymbol{W}$. The stability analysis used to examine the algorithms is an approach equivalent to the methods described by [4] and [5]. With regard to the algorithms, there are two main novelties. First, as opposed to [3], we prove that there is no universal fixed nonlinearity. Second, we propose a very simple nonlinearity, which needs one parameter only to adapt it for all distributions.

II. Optimal Nonlinearities

A. Stability Regions of Some Nonlinearities

Global stability is difficult to investigate due to a complicated cost structure in the parameter manifold. Local stability analyses by different authors [4], [6], [7], [5] have resulted in the statement that for local stability around an equilibrium point, the signal must satisfy

$$E\left\{g'(U)\right\}E\left\{U^2\right\} - E\{g(U)U\} > 0. \tag{8}$$

For monomial nonlinearities of the form

$$g(u) = au|u|^{n-1} \tag{9}$$

the stability condition (8) can be written in terms of the pdf $p_{\breve{U}}$ of the corresponding normalized random variable $\breve{U}$, which is a scaled version of the original random variable U. Thus, we have

$$U = \sigma_U \breve{U}. \tag{10}$$

Using (9) and (10) in (8) results in

$$p\sigma^{n+1}aE\left\{\breve{U}^{n-1}\right\} - \sigma^{n+1}aE\left\{\breve{U}^{n+1}\right\} > 0 \tag{11}$$

which is written in terms of the nonlinearities as

$$E\left\{g'(\breve{U})\right\} - E\left\{g(\breve{U})\breve{U}\right\} > 0. \tag{12}$$

If the source distribution is restricted to symmetric functions, it makes sense to choose an odd nonlinear function for $g(.)$, i.e., the product $g(u)u$ is always positive. We can then write, instead of (12)

$$\frac{E\left\{g'(\breve{U})\right\}}{E\left\{g(\breve{U})\breve{U}\right\}} > 1. \tag{13}$$

This basically means that the scaling of monomial nonlinearities does not affect the stability region, which is entirely defined by the exponent of the monomial and the normalized distribution. Note that such a conclusion is not generally true for polynomials. However, a similar simplification of the stability condition can be carried out for the sign function. For continuous distributions, we know that under scalar multiplication of a unit-variance distribution, the mode of a pdf is inversely proportional to its standard deviation

$$p_U(0) \sim \frac{1}{\sigma_U} \tag{14}$$

and therefore, the stability condition can be written as

$$\frac{E\left\{g'(U)\right\}\sigma_U^2}{E\{g(U)U\}} = \frac{2\sigma_U^2 p_U(0)}{E\{|U|\}} = \frac{2\sigma_U^2\frac{1}{\sigma_U}p_{\breve{U}}(0)}{\sigma_U E\left\{\left|\breve{U}\right|\right\}} = \frac{2p_{\breve{U}}(0)}{E\left\{\left|\breve{U}\right|\right\}} > 1. \tag{15}$$

On the other hand, for general nonlinear functions, if we scale the nonlinearity properly such that

$$E\left\{g(\breve{U})\breve{U}\right\} = 1 \tag{16}$$

with $\breve{U}$, again, as the normalized version of U, then the stability condition (8) simplifies to

$$E\left\{g'(\breve{U})\right\} > 1. \tag{17}$$

Note that (17) is conditioned on the scaling constraint (16). However, it has to be pointed out that the scaling condition is not a necessary condition for stability. It merely ensures unit-variance output signals and simplifies the stability condition equation, albeit not necessarily its satisfaction.

The stability condition (8) has been evaluated for frequently applied nonlinearities, and the resulting stability regions are given in Table I. For those nonlinear functions with two entries in the stability-condition column, the first one is an unconditional stability condition, whereas the second entry is conditioned on satisfying the scaling constraint.

B. Form of the Nonlinearity

If the separation of signals of a certain class of distributions is the goal, the literature suggests to apply nonlinearities of the form $g(u) = au^3$ for sub-Gaussian signals and $g(u) = a\tanh(bu)$ for super-Gaussian signals, where a is a scalar used to adjust the output power. These nonlinearity choices can be refined according to the stability conditions given earlier, as summarized in Table I.

An intuitive explanation of the appropriate form of the nonlinearity can be given as follows. If the nonlinearity is properly scaled, i.e., $E\{g(\breve{U})\breve{U}\} = 1$, the stability condition $E\{g'(\breve{U})\} > 1$ determines if the separating points of the nonlinearity are locally stable. To ensure stability, we aim at making $E\{g'(\breve{U})\}$ as large as possible. For peaky distributions (super-Gaussian) where a large proportion of the pdf lies around zero, the derivative of $g(.)$ should be large around this value, whereas with a flatter distribution, the contrary is the case. This means that super-Gaussian distributions need "sigmoid"-looking nonlinearities for their separation, which are concave functions for their arguments greater than zero, whereas sub-Gaussian distributions need nonlinearities of the form $g(u) = u|u|^{n-1}$ with $n > 1$ showing a convex shape for $u > 0$.

Because sub- (super-) Gaussian signals have a negative (positive) kurtosis κ_4, these expressions are often used interchangeably, although the inverse direction of reasoning is not strictly applicable. Since the nonlinearities for super-Gaussian signals, e.g., sign$(.)$, $a\tanh(.)$, do not exhibit stability for the entire

TABLE I
STABILITY REGIONS OF SOME NONLINEARITIES

Nonlinearity	Scaling condition	Stability condition
au^3	$a = \frac{1}{\kappa_4+3}$	$\kappa_4 < 0$
$au\|u\|^{n-1}$	$a = \frac{1}{E\{\breve{U}^{n+1}\}}$	$p\frac{E\{\breve{U}^{n-1}\}}{E\{\breve{U}^{n+1}\}} > 1$
$a\,\mathrm{sign}(u)$	$a = \frac{1}{E\{\|\breve{U}\|\}}$	$\frac{2p_{\breve{U}}(0)}{E\{\|\breve{U}\|\}} > 1$
$a\tanh(u)$	$a = \frac{1}{E\{\breve{U}\tanh(\breve{U})\}}$	$\sigma_U^2 \frac{1-E\{\tanh^2(U)\}}{E\{U\tanh(U)\}} > 1$
		$\frac{1-E\{\tanh^2(\breve{U})\}}{E\{\breve{U}\tanh(\breve{U})\}} > 1$
threshold NL	$a = \frac{1}{2\int_\vartheta^\infty p_{\breve{U}}(\breve{u})\breve{u}d\breve{u}}$	$\frac{\sigma_U^2 p_U(\vartheta)}{\int_\vartheta^\infty p_U(u)udu} > 1$
		$\frac{p_{\breve{U}}(\vartheta)}{\int_\vartheta^\infty p_{\breve{U}}(\breve{u})\breve{u}d\breve{u}} > 1$

positive kurtosis plane, distributions might be constructed, for which both nonlinearities $g(u) = au^3$ and $g(u) = a\tanh(bu)$ fail [8]. This will be further discussed in Section III.

C. Optimization of the Nonlinearity

The fact that the stability of blind separation algorithms depends on a nonlinear moment being greater than one implies that robustness of the algorithms can be obtained by making this nonlinear moment as large as possible. We wish to maximize the left-hand side of the stability condition for a scaled nonlinearity according to (17). The scaling constraint

$$\int_{-\infty}^{\infty} g(u)p(u)u\,du = 1 \tag{18}$$

alone is not sufficient. Any even part of $g(.)$ would show up neither in the constraint nor in the integral to maximize. Clearly, due to symmetry, we wish to restrict $g(.)$ to odd functions. The optimization problem can be formulated as follows:

$$\text{maximize} \int_{-\infty}^{\infty} g'(u)p(u)\,du \tag{19}$$

$$\text{subject to} \int_{-\infty}^{\infty} g^2(u)p(u)\,du = c \tag{20}$$

where c is a constant. Now, an even part of $g(.)$ would increase the constraint unnecessarily without contributing to the integral to maximize. We are attempting to find the optimal nonlinearity by calculus of variations. To this end, we define

$$f = g'(u)p(u) + \lambda(g^2(u)p(u)) \tag{21}$$

where λ is a Lagrange multiplier. To find the optimal $g(u)$, we have to solve the Euler–Lagrange equation [9]

$$\frac{\partial f}{\partial g} - \frac{d}{du}\frac{\partial f}{\partial g'} = \frac{\partial f}{\partial g} - \frac{\partial}{\partial u}\frac{\partial f}{\partial g'} - \frac{\partial}{\partial g}\frac{\partial f}{\partial g'}g' - \frac{\partial}{\partial g'}\frac{\partial f}{\partial g'}g'' = 0 \tag{22}$$

where we abridged $p \triangleq p(u)$, $p' \triangleq p'(u) = (\partial/\partial u)p(u)$, $g \triangleq g(u)$, and $g' \triangleq g'(u) = (\partial/\partial u)g(u)$. Working out the different terms of (22) for (21) yields

$$\frac{\partial f}{\partial g} = 2\lambda gp \tag{23}$$

$$\frac{\partial}{\partial u}\frac{\partial f}{\partial g'} = p' \tag{24}$$

$$\frac{\partial}{\partial g}\frac{\partial f}{\partial g'}g' = 0 \tag{25}$$

$$\frac{\partial}{\partial g'}\frac{\partial f}{\partial g'}g'' = 0. \tag{26}$$

Using (23)–(26) in (22) results in

$$g = \frac{1}{2\lambda}\frac{p'}{p}. \tag{27}$$

λ can now be found by the constraint on the output power of the nonlinearity. For that, we would have to determine the constant c. Alternately, we know that a further constraint is the one given originally. Inserting the solution (27) into (18) gives us

$$\int_{-\infty}^{\infty}\frac{1}{2\lambda}\frac{p'(u)}{p(u)}up(u)\,du = \frac{1}{2\lambda}\int_{-\infty}^{\infty}up'(u)\,du = 1. \tag{28}$$

Integrating by parts yields, for the integral in (28)

$$\int_{-\infty}^{\infty}up'(u)\,du = p(u)u\Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty}p(u)\,du = -1. \tag{29}$$

The desired solution is thus $\lambda = -(1/2)$, leading to

$$g(u) = -\frac{p'(u)}{p(u)} \tag{30}$$

which is exactly the score function. This is a further justification for the score function in addition to the ones already known, such as ML and InfoMax (see also [10]).

III. PROOF OF THE NONEXISTENCE OF A UNIVERSAL NONLINEARITY

A. Polynomial Nonlinearity

Model-independent nonlinearities such as the one given in (6) have been claimed to separate both super- and sub-Gaussian distributions [3]. An intuitive explanation of why such approaches fail has been offered in [11] by noticing that the corresponding function to which the nonlinearity is the score function is of the wrong type. Indeed, the evaluation of (6) by (8) reveals its failure. In the following, a proof is given that such polynomial nonlinearities are bound to fail.

B. Problem Statement

We wish to prove that there is no general nonlinear function $g(.)$ that is stable in the sense that

$$\frac{\sigma_X^2 E\{g'(X)\}}{E\{Xg(X)\}} > 1 \tag{31}$$

for both X a sub- and super-Gaussian distributed random variable. In the following, we restrict ourselves to the family of generalized Gaussian signals. This allows us to use some properties of higher order moments, as given in the next subsection. The restriction does not result in a loss of generality since if we can prove that no single nonlinearity can separate all sub- and super-Gaussian signals, then this result also means that no single nonlinearity can separate any non-Gaussian signals.

C. Statistical Moments Prerequisites

Consider $\breve{N}$ to be a unit-variance Gaussian variable $\breve{N} \sim \mathcal{N}(0, 1)$ and X the generalized Gaussian variable

$$p_X(x) = \frac{\alpha}{2\beta\Gamma\left(\frac{1}{\alpha}\right)} e^{-(|x|/\beta)^\alpha} \tag{32}$$

parameterized by α, which is larger (smaller) than 2 for sub- (super-) Gaussian signals. We will also consider a normalized version of X ($\breve{X}$). β can be found from the general expression for the mth-order moment of a generalized Gaussian signal [12]

$$E\{|X|^m\} = \frac{\Gamma\left(\frac{m+1}{\alpha}\right)}{\Gamma\left(\frac{1}{\alpha}\right)} \beta^m. \tag{33}$$

$\Gamma(.)$ is the gamma function given by $\Gamma(a) = \int_0^\infty x^{a-1}e^{-x}\,dx$ and shows a recursive property similar to the factorial function $\Gamma(a+1) = a\Gamma(a)$. For $m = 2$, (33) leads to

$$\beta = \sqrt{\frac{\Gamma\left(\frac{1}{\alpha}\right)}{\Gamma\left(\frac{3}{\alpha}\right)}}\, \sigma_X. \tag{34}$$

We have the following relationship between the even moments of the unit-variance Gaussian variable

$$E\left\{\left|\breve{N}\right|^m\right\} = (m-1)E\left\{\left|\breve{N}\right|^{m-2}\right\}. \tag{35}$$

Similarly, for the even moments of generalized Gaussian variables with unit variance, we can find

$$\frac{E\left\{\left|\breve{X}\right|^m\right\}}{E\left\{\left|\breve{X}\right|^{m-2}\right\}} = \frac{\Gamma\left(\frac{m+1}{\alpha}\right)}{\Gamma\left(\frac{m-1}{\alpha}\right)} \cdot \frac{\Gamma\left(\frac{1}{\alpha}\right)}{\Gamma\left(\frac{3}{\alpha}\right)}. \tag{36}$$

The moments of any unit-variance random variable are a monotonic increasing function in m

$$E\left\{\left|\breve{X}\right|^m\right\} \geq E\left\{\left|\breve{X}\right|^{m-2}\right\}, \qquad m \geq 4 \tag{37}$$

with equality if and only if $\breve{X}$ is binary distributed. A proof of (37) can be found in Appendix A. Furthermore, if $\breve{X}$ is super-Gaussian distributed, we find [13]

$$E\left\{\left|\breve{X}\right|^m\right\} > E\left\{\left|\breve{N}\right|^m\right\}, \qquad m > 2. \tag{38}$$

Vice versa, for $\breve{X}$ being sub-Gaussian distributed

$$E\left\{\left|\breve{X}\right|^m\right\} < E\left\{\left|\breve{N}\right|^m\right\}, \qquad m > 2. \tag{39}$$

Although stochastic ordering is also possible for fractional moments [13], we only consider even integers for m in what follows. In fact, using (35) and (37) in (38) and (39), we realize that

$$E\left\{\left|\breve{X}\right|^m\right\} > (m-1)E\left\{\left|\breve{N}\right|^{m-2}\right\} \quad X \text{ super-Gaussian distributed} \tag{40}$$

$$E\left\{\left|\breve{X}\right|^m\right\} < (m-1)E\left\{\left|\breve{N}\right|^{m-2}\right\} \quad X \text{ sub-Gaussian distributed.} \tag{41}$$

Tighter bounds can be obtained as

$$E\left\{\left|\breve{X}\right|^m\right\} > (m-1)E\left\{\left|\breve{X}\right|^{m-2}\right\} \quad X \text{ super-Gaussian distributed} \tag{42}$$

$$E\left\{\left|\breve{X}\right|^m\right\} < (m-1)E\left\{\left|\breve{X}\right|^{m-2}\right\} \quad X \text{ sub-Gaussian distributed} \tag{43}$$

where in the right-hand sides of (40) and (41), $\breve{N}$ has been replaced by $\breve{X}$. A proof of (42) and (43) is provided in Appendix B. Equation (43) reveals immediately (by setting $m = n+1$) that the monomial function $g(u) = au^n$ is a stable nonlinearity for any sub-Gaussian signals [see (11) or Table I].

Most of the equalities and inequalities from (35) to (43) can be extended to distributions whose variance is unequal to one by noting

$$E\{|X|^m\} = \sigma_X^m E\left\{\left|\breve{X}\right|^m\right\} \tag{44}$$

if $\breve{X}$ is the normalized version of X. In particular, the moments of a Gaussian variable $N \sim \mathcal{N}(0, \sigma_N^2)$ are related as

$$E\{|N|^m\} = (m-1)E\{|N|^{m-2}\}\sigma_N^2. \tag{45}$$

The even moments of generalized Gaussian variables with variance σ_X^2 are related as

$$\frac{E\{|X|^m\}}{E\{|X|^{m-2}\}} = \frac{\Gamma\left(\frac{m+1}{\alpha}\right)}{\Gamma\left(\frac{m-1}{\alpha}\right)} \cdot \frac{\Gamma\left(\frac{1}{\alpha}\right)}{\Gamma\left(\frac{3}{\alpha}\right)}\, \sigma_X^2. \tag{46}$$

For arbitrary distributions, we have

$$E\{|X|^m\} \geq E\{|X|^{m-2}\}\sigma_X^2, \qquad m \geq 4 \tag{47}$$

with equality if and only if X is binary distributed. Finally

$$E\{|X|^m\} > (m-1)E\{|X|^{m-2}\}\sigma_X^2 \quad X \text{ super-Gaussian distributed} \tag{48}$$

$$E\{|X|^m\} < (m-1)E\{|X|^{m-2}\}\sigma_X^2 \quad X \text{ sub-Gaussian distributed.} \tag{49}$$

In the following, only considering even moments allows us to drop the modulus operator $|.|$.

D. Stability Analysis

Assume that the nonlinearity $g(.)$ is either an odd polynomial or that we can replace it by its Taylor series (with odd powers only)

$$g(X) = \sum_{k\,\mathrm{odd}} t_k X^k. \tag{50}$$

The Taylor series used in (50) assumes the existence of all derivatives of $g(.)$. If the nonlinearity is not smooth, however, the use of a polynomial can still make sense since the expectation operator allows discontinuities in the nonlinearity since its evaluation involves an integration over such singularities. Clearly, differentiation of $g(X)$ w.r.t. X yields

$$g'(X) = \sum_{k\,\mathrm{odd}} t_k k X^{k-1}. \tag{51}$$

If (50) and (51) are inserted into (31), we get the stability condition expressed as a function of the Taylor series coefficients

$$\frac{\sigma_X^2 \sum_{k\,\mathrm{odd}} t_k k E\left\{X^{k-1}\right\}}{\sum_{k\,\mathrm{odd}} t_k E\left\{X^{k+1}\right\}} > 1. \tag{52}$$

Because of (45), we now see that for Gaussian variables, any nonlinearity will lead to

$$\frac{\sigma_X^2 E\{g'(N)\}}{E\{Xg(N)\}} = 1 \not> 1 \tag{53}$$

which was, of course, to be expected. In addition, for strictly non-negative values of t_k and X a sub-Gaussian variable, stability is guaranteed due to (49), and hence

$$\sum_{k\,\mathrm{odd}} t_k E\left\{X^{k+1}\right\} < \sum_{k\,\mathrm{odd}} t_k k E\left\{X^{k-1}\right\} \sigma_X^2. \tag{54}$$

In the following, we will show that no polynomial nonlinearity can be found to separate both sub- and super-Gaussian signals. We assume symmetric unbiased signals leading to anti-symmetric nonlinearities.

Lemma: For the natural gradient update equation (7), there does not exist a single fixed nonlinearity $g(.)$ that separates arbitrary mixtures of sub- and super-Gaussian signals.

Proof: We carry out this proof by induction. First, we note that with only $t_1 \neq 0$, we have a linear function for $g(.)$, which is, of course, unable to separate any distribution. We will therefore have to add at least one further coefficient t_k unequal to zero. We then show that the choice of this coefficient is contradictive and, hence, cannot lead to stability. By induction, we can add as many coefficients as we like, but we will never reach stability for all distributions.

The basis of the induction is to show that $g(X) = t_1 X + t_3 X^3$ cannot separate both sub- and super-Gaussian signals because

$$G(X) \triangleq \frac{\sigma_X^2(t_1 + 3t_3\sigma_X^2)}{\sigma_X^2 t_1 + t_3 E\{X^4\}} \tag{55}$$

is always smaller than one either for sub-Gaussian or for super-Gaussian signals. We first note that t_1 and t_3 need to have different signs; otherwise, $G(X)$ is smaller than one for super-Gaussian signals, as can be easily verified using (48). Furthermore, we may restrict ourselves to positive values of t_1, implicating negative values of t_3, since the reverse case leads to identical values of $G(X)$ as both numerator and denominator are linear expressions in both t_1 and t_3. We now distinguish between two cases.

Case 1) $t_1 < -3t_3\sigma_X^2$. This makes the numerator of $G(X)$ negative. For super-Gaussian signals, the denominator is negative as well but is smaller than the numerator due to (48) and negative t_3. Thus, $G(X) < 1$ for super-Gaussian signals.

Case 2) $t_1 > -3t_3\sigma_X^2$. Here, we have a positive numerator. The denominator is greater than the numerator for sub-Gaussian signals due to (49) and negative t_3. Hence, $G(X) < 1$ for sub-Gaussian signals.

Suppose that a nonlinearity with only terms up to order K does not yield a stable solution. In the following, we assume that a solution with $g(X) = \sum_{k=1,\,k\,\mathrm{odd}}^{K} t_k X^k$ will be stable for either sub- or super-Gaussian signals, but not both. We now show that a solution that is unstable for either signal distribution class cannot, in one step, be extended to an everywhere-stable solution. To stabilize the nonlinearity for all distributions, we add a further term in the nonlinearity: $t_{K+2}X^{K+2}$. This results in the extended nonlinearity

$$g(X) = \sum_{\substack{k=1\\ k\,\mathrm{odd}}}^{K+2} t_k X^k. \tag{56}$$

The stability criterion can now be written as

$$\frac{\sigma_X^2 \left(\sum_{\substack{k=1\\ k\,\mathrm{odd}}}^{K} t_k k E\left\{X^{k-1}\right\} + t_{K+2}(K+2) E\left\{X^{K+1}\right\} \right)}{\sum_{\substack{k=1\\ k\,\mathrm{odd}}}^{K} t_k E\left\{X^{k+1}\right\} + t_{K+2} E\left\{X^{K+3}\right\}} > 1. \tag{57}$$

If we want to multiply both sides of (57) by the denominator, we have to distinguish two cases, depending on the sign of the denominator of (57).

Case 1) [Positive Denominator of (57)]: From (57), it follows that

$$\begin{aligned} t_{K+2} &\left(\sigma_X^2 (K+2) E\left\{X^{K+1}\right\} - E\left\{X^{K+3}\right\}\right) \\ &> - \sum_{\substack{k=1\\ k\,\mathrm{odd}}}^{K} t_k \left(\sigma_X^2 k E\left\{X^{k-1}\right\} - E\left\{X^{k+1}\right\}\right). \end{aligned} \tag{58}$$

Because of (49), $\sigma_X^2 (K+2)\, E\{X^{K+1}\} - E\{X^{K+3}\} > 0$ for sub-Gaussian signals; hence, we have

$$t_{K+2} > \frac{- \sum_{\substack{k=1\\ k\,\mathrm{odd}}}^{K} t_k \left(\sigma_X^2 k E\left\{X^{k-1}\right\} - E\left\{X^{k+1}\right\}\right)}{\sigma_X^2 (K+2) E\left\{X^{K+1}\right\} - E\left\{X^{K+3}\right\}}. \tag{59}$$

However, due to (48), for super-Gaussian signals $\sigma_X^2(K+2)E\{X^{K+1}\} - E\{X^{K+3}\} < 0$, and therefore

$$t_{K+2} < \frac{-\sum_{\substack{k=1\\k\,\mathrm{odd}}}^{K} t_k\left(\sigma_X^2 kE\left\{X^{k-1}\right\} - E\left\{X^{k+1}\right\}\right)}{\sigma_X^2(K+2)E\left\{X^{K+1}\right\} - E\left\{X^{K+3}\right\}}. \quad (60)$$

In the limit for $\alpha \to 2$ for generalized Gaussian distributions, the right-hand sides of (59) and (60) will become equal, making it impossible for t_{K+2} to satisfy both conditions. The choice of t_{K+2} restricts the stable range of α.

Case 2) [Negative Denominator of (57)]: According to (57)

$$t_{K+2}\left(\sigma_X^2(K+2)E\left\{X^{K+1}\right\} - E\left\{X^{K+3}\right\}\right) < -\sum_{\substack{k=1\\k\,\mathrm{odd}}}^{K} t_k\left(\sigma_X^2 kE\left\{X^{k-1}\right\} - E\left\{X^{k+1}\right\}\right). \quad (61)$$

Since $\sigma_X^2(K+2)E\{X^{K+1}\} - E\{X^{K+3}\} > 0$ for sub-Gaussian signals, we have

$$t_{K+2} < \frac{-\sum_{\substack{k=1\\k\,\mathrm{odd}}}^{K} t_k\left(\sigma_X^2 kE\left\{X^{k-1}\right\} - E\left\{X^{k+1}\right\}\right)}{\sigma_X^2(K+2)E\left\{X^{K+1}\right\} - E\left\{X^{K+3}\right\}} \quad (62)$$

but for super-Gaussian signals, $\sigma_X^2(K+2)E\{X^{K+1}\} - E\{X^{K+3}\} < 0$, and therefore

$$t_{K+2} > \frac{-\sum_{\substack{k=1\\k\,\mathrm{odd}}}^{K} t_k\left(\sigma_X^2 kE\left\{X^{k-1}\right\} - E\left\{X^{k+1}\right\}\right)}{\sigma_X^2(K+2)E\left\{X^{K+1}\right\} - E\left\{X^{K+3}\right\}}. \quad (63)$$

Along the same line of argument as in Case 1, we notice again the contradiction for the choice of t_{K+2}.

We have now shown that if a nonlinearity

$$g(X) = \sum_{\substack{k=1\\k\,\mathrm{odd}}}^{K} t_k X^k \quad (64)$$

is capable of only separating either sub- or super-Gaussian signals (but not both), then its extended version [additional $(K+2)$th term]

$$g(X) = \sum_{\substack{k=1\\k\,\mathrm{odd}}}^{K+2} t_k X^k \quad (65)$$

will also only separate either sub- or super-Gaussian signals (but not necessarily the same distribution class as the original nonlinearity) because t_{K+2} would have to be chosen greater and smaller than a certain a value at the same time.

By induction, this applies to any arbitrarily long odd power series. □

IV. Stabilization of Mixed Distributions

In the previous section, we showed the nonexistence of a single universal nonlinearity. Different solutions to the blind separation problem of unknown mixed distributions have been suggested, such as the dynamic change between two different nonlinearities [14]–[16] or the transpose form of (7) by [7]. Related to the transpose form is a change of sign of the nonlinearity in the EASI algorithm [6]. In the following, we investigate the extended stability regions of the former methods.

A. Difficult Distributions

From Table I, it becomes clear that if a non-Gaussian distribution exists that is neither separable by $g(u) = u^3$ nor by $g(u) = a\tanh(u)$, it has to show a positive kurtosis since $g(u) = u^3$ covers all negative-kurtosis distributions, but the stability region of $g(u) = a\tanh(u)$ does not include all positive-kurtosis distributions. One such peculiar distribution was given by Douglas [8]. It is a symmetric, discrete, quaternary signaling scheme with symbols $\in [\pm A_1, \pm A_2]$, where $A_2 = 3.8A_1$ and $\Pr(x = A_2) = 0.035$. A_1 is adjusted for unit variance resulting in $A_1 = 0.718$. The kurtosis of this distribution is $\kappa_4 = 1.12$. As can be checked, this distribution does not satisfy the stability condition for any of the two nonlinearities. In fact, we can create more of those challenging distributions by making the following considerations. We only consider quaternary symmetric signals. With more levels of signaling, it is, of course, possible to create these kinds of distributions too. However, using four symmetric levels provides enough degrees of freedom to construct the desired distributions. We are looking at combinations of A_1, A_2, $p_1 = \Pr(x = A_1)$, and $p_2 = \Pr(x = A_2)$ that result in instability for both $g(u) = u^3$ and $g(u) = a\tanh(u)$. We assume that a is adjusted for unit-variance outputs. We have the following constraints:

C1) distributional sum

$$p_1 + p_2 = \tfrac{1}{2} \quad (66)$$

C2) unit variance

$$p_1 A_1^2 + p_2 A_2^2 = \tfrac{1}{2} \quad (67)$$

C3) unstable for $g(u) = u^3$

$$\kappa_4 \geq 0 \iff p_1 A_1^4 + p_2 A_2^4 \geq \tfrac{3}{2} \quad (68)$$

C4) unstable for $g(u) = a\tanh(u)$

$$\begin{aligned} &1 - E\{\tanh^2(U)\} > E\{U\tanh(U)\} \iff \\ &1 - 2p_1\tanh^2(A_1) - 2p_2\tanh^2(A_2) \\ &\quad > 2p_1A_1\tanh(A_1) + 2p_2A_2\tanh(A_2). \end{aligned} \quad (69)$$

Combining these four conditions—for a detailed derivation see [17]—leads to the possible range of A_2 and p_2, as depicted in Fig. 1. One example of a "difficult" distribution can be extracted from Fig. 1 as $A_2 = 5$, $p_2 = 0.005$, and therefore, $p_1 = 0.495$, and $A_1 = 0.87$.

B. Threshold Nonlinearity

The threshold nonlinearity [18]

$$g(u) = \begin{cases} 0, & |u| < \vartheta \\ a\,\mathrm{sign}(u), & |u| \geq \vartheta \end{cases} \quad (70)$$

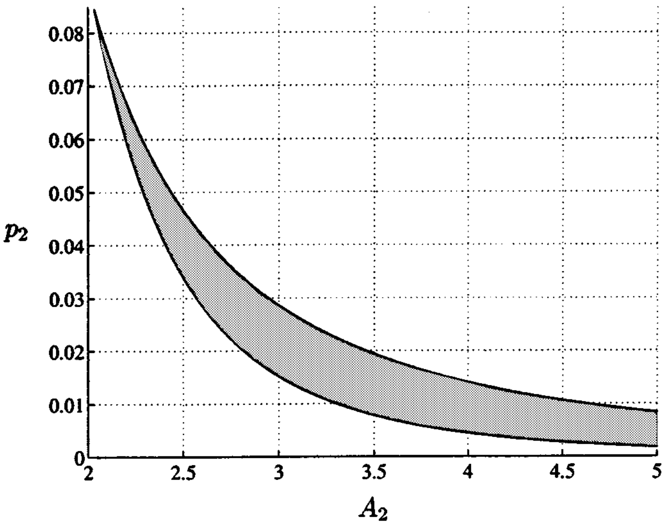

Fig. 1. Possible region of A_2 and p_2 for generating "challenging" distributions.

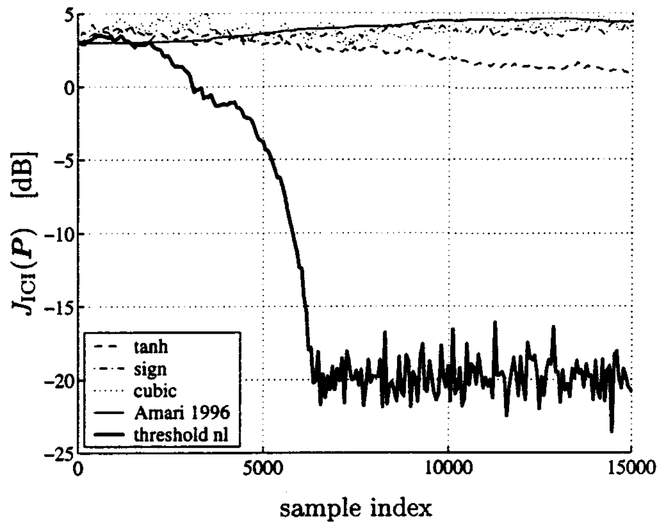

Fig. 2. Convergence of different nonlinearities for a mixture of signals exhibiting a "challenging" distribution.

with $\vartheta = A_1$ and $a = 2$ successfully separates the distribution given above, which was verified both by inspection of the stability condition as well as experimental simulation. Fig. 2 shows the convergence performance of different nonlinearities for ten sources with the "challenging" distribution using the natural gradient algorithm [3] given by (7). The fidelity criterion used is the average interchannel interference given by [19]

$$J_{\rm ICI}(\boldsymbol{P}) = \frac{1}{M_{\rm s}} \left(\sum_{i=1}^{M_{\rm s}} \frac{\sum_{k=1}^{M_{\rm s}} p_{ik}^2}{\max_k p_{ik}^2} \right) - 1 \tag{71}$$

where p_{ik} are elements of the global transfer matrix given by $\boldsymbol{P} = \boldsymbol{WA}$. All but the threshold nonlinearity fail to separate the signals, including the nonlinearity motivated by Gram–Charlier approximation, as given by (6). This leads to the question if the threshold nonlinearity is capable of separating any non-Gaussian distribution for an appropriate threshold parameter ϑ. The answer is given by the following lemma. In contrast to (70), we omit scaling and obtain a more general case.

Lemma: The threshold nonlinearity given by

$$g(u) = \begin{cases} 0, & |u| < \vartheta \\ \text{sign}(u), & |u| \geq \vartheta \end{cases} \tag{72}$$

satisfies the local stability condition

$$\sigma_X^2 p_X(\vartheta) - \int_\vartheta^\infty p_X(x) x\, dx > 0 \tag{73}$$

for some appropriately chosen $\vartheta \geq 0$ and any continuous, differentiable, non-Gaussian, symmetric output distribution $p_X(.)$. In addition, we have that

$$\sigma_N^2 p_N(\vartheta) - \int_\vartheta^\infty p_N(x) x\, dx \equiv 0, \qquad \forall \vartheta \in \mathbb{R}_0^+ \tag{74}$$

if and only if $p_N(.)$ is Gaussian.

The proof is one of existence rather than of construction in that it shows that there is a threshold parameter ϑ for which the update equation (7) is stable, but it does not necessarily give an explicit solution for ϑ.

Proof: We consider real, symmetric, continuous, differentiable distributions. The result for other distributions can be obtained by convolving discrete distributions by low-variance Gaussian kernels. We have to show that to satisfy the stability condition (8), the inequality

$$\sigma_X^2 p_X(\vartheta) > \int_\vartheta^\infty p_X(x) x\, dx \tag{75}$$

has to be satisfied for at least one value of $\vartheta \in \mathbb{R}_0^+$, given a non-Gaussian distribution. We assume that no value of ϑ can satisfy (75); therefore

$$\sigma_X^2 p_X(\vartheta) \leq \int_\vartheta^\infty p_X(x) x\, dx, \qquad \forall \vartheta \in \mathbb{R}_0^+ \tag{76}$$

and lead the proof by contradiction.

First, we show that for a normal distribution $p_N(.) = \mathcal{N}(0, \sigma_N^2)$, we have

$$\sigma_N^2 p_N(\vartheta) \equiv \int_\vartheta^\infty p_N(x) x\, dx, \qquad \forall \vartheta \in \mathbb{R}_0^+. \tag{77}$$

To this end, we assume that

$$\sigma_X^2 p_X(\vartheta) - \int_\vartheta^\infty p_X(x) x\, dx = c \tag{78}$$

for some nonpositive constant c. Taking derivatives of both sides of (78) with respect to ϑ gives the differential equation

$$\sigma_X^2 \frac{dp_X(\vartheta)}{d\vartheta} + \vartheta p_X(\vartheta) = 0. \tag{79}$$

Equation (79) is a simple first-order differential equation whose parametric solution is

$$p_X(\vartheta) = K \exp\left(-\frac{\vartheta^2}{2\sigma_X^2}\right), \qquad K \geq 0. \tag{80}$$

Because $p_X(.)$ is a pdf, the value of K must be $K = 1/(\sqrt{2\pi}\sigma_X)$, meaning that $c = 0$. This proves the uniqueness of the Gaussian distribution as the pdf that minimizes the left-hand side of the stability condition inequality. All other continuously valued and differentiable distributions should therefore satisfy the inequality.

In the following, we make use of crossing points of the different pdfs, which are given as points where these two functions are identical. By taking ϑ as the last (right-most) crossing point of the distribution under consideration and the normal distribution, we have either (75), which is already in contradiction to (76), or

$$\sigma_X^2 p_X(\vartheta) < \int_{\vartheta}^{\infty} p_X(x)\, x\, dx \tag{81}$$

for some region around that particular ϑ. By integrating both sides of (76) over $\mathbb{R}_0^+$, we get

$$\sigma_X^2 \int_0^{\infty} p_X(\vartheta)\, d\vartheta = \frac{\sigma_X^2}{2} < \int_0^{\infty}\int_{\vartheta}^{\infty} p_X(x) x\, dx\, d\vartheta \tag{82}$$

where the strict inequality results from the region where (81) is valid. The right-hand side of (82) can be solved by exchanging the integrals

$$\begin{aligned}\int_0^{\infty}\int_{\vartheta}^{\infty} p_X(x) x\, dx\, d\vartheta &= \int_0^{\infty}\int_0^{x} d\vartheta\, p_X(x) x\, dx \\ &= \int_0^{\infty} p_X(x) x^2\, dx = \frac{\sigma_X^2}{2}.\end{aligned} \tag{83}$$

Equation (83) is a contradiction to (82). This means that if there are values of ϑ satisfying (81), due to (83), there must also be values satisfying (75) and vice versa, which is in contradiction to (76). □

Note that we have not used the notion of sub- and super-Gaussian to refer to the tail of the distribution since multimodal distributions are also considered, where the tail observation is still valid, but the use of the terms sub- and super-Gaussian is not appropriate.

V. Computer Simulations

As a demonstration of the local stability of the threshold nonlinearity for different distributions, a computer simulation was carried out, where a mixture of different distributions, requiring different threshold values, was blindly separated using the threshold nonlinearity. The application area was chosen data communications; therefore, three BPSK signals, three 4-PAM signals, three uniformly distributed signals (simulating large-alphabet constellations), and one Gaussian signal (simulating some undesired noise signal) were added using a mixing matrix with diagonal-dominant entries. The reason for this choice was that the threshold values could be fixed according to the distribution of the output signal. In this case, we chose $\vartheta = 1$ for BPSK, $\vartheta = 1.3416$ for 4-PAM, $\vartheta = 1.5$ for uniform, and $\vartheta = 0$ for Gaussian distributions. The corresponding gain factors a in (70) were chosen such that the output signals were of unit variance. If an unknown permutation of the overall transfer function occurs, the threshold values have to be adapted [20]. Such approaches are outside the scope of this paper. Since only local stability is of concern here, such a choice of the mixing matrix is not a restriction. The step size was adjusted for residual mixing of $J_{\mathrm{ICI}}(\boldsymbol{P}) = -20$ dB. Fig. 3 shows clear convergence of the separation process. Fig. 4 illustrates the development of the output kurtoses as they approach -2 for BPSK, -1.36 for 4-PAM, -1.2 for uniform signals, and zero for the Gaussian signal.

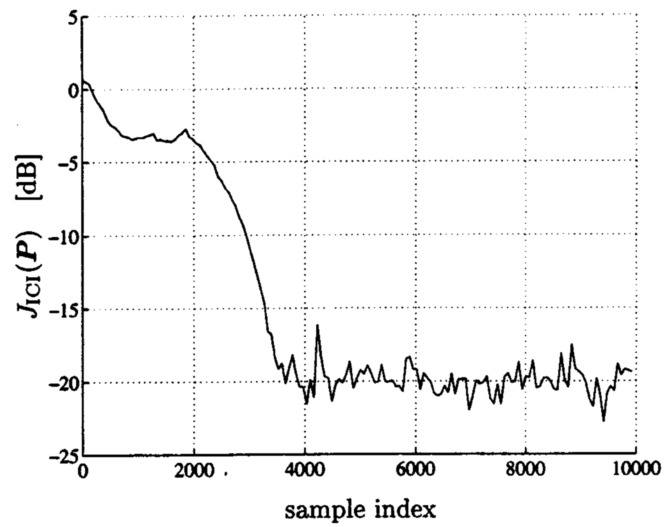

Fig. 3. Convergence of blind separation of mixed-distribution signals using the threshold nonlinearity.

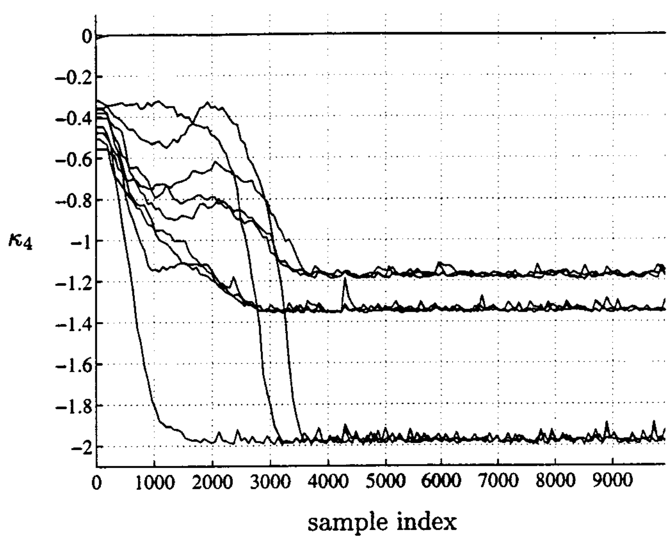

Fig. 4. Output kurtoses of blind separation of mixed-distribution signals using the threshold nonlinearity.

VI. Conclusions

The score function is a robust choice for model mismatch as long as the kurtosis sign does not change. The single, universal nonlinearity, however, which separates mixed distributions, regardless of their kurtosis signs, does not exist. There are even special distributions (with positive kurtosis) that are not separable by the "standard" hyperbolic tangent function. A remedy is at hand in the form of the threshold nonlinearity, which, by suitable choice of the threshold parameter, blindly separates any non-Gaussian distributed signals.

APPENDIX A
PROOF OF (37)

Equation (37) can be proved trivially for super-Gaussian signals by (42). A more general proof of (37) for all symmetric distributions with unit variance can be constructed as follows.

Proof: The proof is based on induction. As the basis of the induction, we presume that we have shown

$$E\left\{\left|\breve{X}\right|^{m-2}\right\} \geq E\left\{\left|\breve{X}\right|^{m-4}\right\} \geq 1 \tag{84}$$

which is true for $m = 4$ by the assumption of unit variance. Reformulation of (84) leads to

$$\int_0^1 (x^{m-2} - x^{m-4})p(x)\,dx + \int_1^\infty (x^{m-2} - x^{m-4})p(x)\,dx \geq 0. \tag{85}$$

Due to negative values in the integrand, the first integral of (85) will be negative. By multiplying the first integrand with values smaller than or equal to one and the second integrand with values greater than or equal to one, we can write

$$\int_0^1 x^2(x^{m-2} - x^{m-4})p(x)\,dx + \int_1^\infty x^2(x^{m-2} - x^{m-4})p(x)\,dx \geq 0 \tag{86}$$

or

$$\int_0^\infty x^m p(x)\,dx \geq \int_0^\infty x^{m-2} p(x)\,dx \tag{87}$$

which, for symmetric distributions, is equivalent to (37). Equality applies only if $\breve{X}$ is restricted to ± 1, hence, the binary case. □

APPENDIX B
PROOF OF (42) AND (43)

In the following, we will make use of the $\lfloor . \rfloor$ and the $\lceil . \rceil$ operators to denote the integer not greater than and the integer not smaller than, respectively. We assume even $m \geq 4$ so that $\lfloor m/2 \rfloor = \lceil m/2 \rceil = m/2$.

Proof: We can write (36) as

$$\frac{E\left\{\left|\breve{X}\right|^{m}\right\}}{E\left\{\left|\breve{X}\right|^{m-2}\right\}} = \frac{\Gamma\left(\frac{m+1}{\alpha}\right)}{\Gamma\left(\frac{m-1}{\alpha}\right)} \cdot \frac{\Gamma\left(\frac{1}{\alpha}\right)}{\Gamma\left(\frac{3}{\alpha}\right)} = \frac{\Gamma\left(\frac{3}{\alpha} + \frac{m-2}{\alpha}\right)\Gamma\left(\frac{1}{\alpha}\right)}{\Gamma\left(\frac{1}{\alpha} + \frac{m-2}{\alpha}\right)\Gamma\left(\frac{3}{\alpha}\right)}. \tag{88}$$

Consider $\alpha < 2$ first. Using the recursive property $\Gamma(a+1) = a\Gamma(a)$, we can lower bound (88) by

$$\frac{\Gamma\left(\frac{3}{\alpha} + \frac{m-2}{\alpha}\right)\Gamma\left(\frac{1}{\alpha}\right)}{\Gamma\left(\frac{1}{\alpha} + \frac{m-2}{\alpha}\right)\Gamma\left(\frac{3}{\alpha}\right)} > \frac{\Gamma\left(\frac{3}{\alpha}\right) \cdot \frac{3}{\alpha}\left(\frac{3}{\alpha}+1\right)\left(\frac{3}{\alpha}+2\right) \cdot \cdots \cdot \left(\frac{3}{\alpha}+\left\lfloor\frac{m-2}{\alpha}\right\rfloor - 1\right) \cdot \Gamma\left(\frac{1}{\alpha}\right)}{\Gamma\left(\frac{1}{\alpha}\right) \cdot \frac{1}{\alpha}\left(\frac{1}{\alpha}+1\right)\left(\frac{1}{\alpha}+2\right) \cdot \cdots \cdot \left(\frac{1}{\alpha}+\left\lfloor\frac{m-2}{\alpha}\right\rfloor - 1\right) \cdot \Gamma\left(\frac{3}{\alpha}\right)} = \prod_{k=0}^{\lfloor (m-2)/\alpha \rfloor - 1} \frac{\frac{3}{\alpha} + k}{\frac{1}{\alpha} + k}. \tag{89}$$

Using

$$\frac{\frac{3}{\alpha} + k}{\frac{1}{\alpha} + k} \geq \frac{\frac{3}{2} + k}{\frac{1}{2} + k} = \frac{3 + 2k}{1 + 2k}, \qquad \alpha < 2,\ k \geq 0 \tag{90}$$

with equality if and only if $k = 0$, and

$$\left\lfloor \frac{m-2}{\alpha} \right\rfloor \geq \left\lfloor \frac{m}{2} - 1 \right\rfloor = \frac{m}{2} - 1, \qquad \alpha < 2 \tag{91}$$

we can write

$$\frac{E\left\{\left|\breve{X}\right|^{m}\right\}}{E\left\{\left|\breve{X}\right|^{m-2}\right\}} > \prod_{k=0}^{\lfloor (m-2)/\alpha \rfloor - 1} \frac{\frac{3}{\alpha} + k}{\frac{1}{\alpha} + k} \geq \prod_{k=0}^{(m/2)-2} \frac{3+2k}{1+2k} = 3 + 2\left(\frac{m}{2} - 2\right) = m - 1 \tag{92}$$

where the second but last equality is by realizing that the denominator factors are cancelled by previous numerator factors so that the last numerator factor remains, and the last equality is by the assumption of even m, which completes the proof of (42).

Now, we consider $\alpha > 2$. Similarly to (89), we can now upper bound (88) by

$$\frac{\Gamma\left(\frac{3}{\alpha} + \frac{m-2}{\alpha}\right)\Gamma\left(\frac{1}{\alpha}\right)}{\Gamma\left(\frac{1}{\alpha} + \frac{m-2}{\alpha}\right)\Gamma\left(\frac{3}{\alpha}\right)} < \frac{\frac{3}{\alpha}\left(\frac{3}{\alpha}+1\right)\left(\frac{3}{\alpha}+2\right) \cdot \cdots \cdot \left(\frac{3}{\alpha}+\left\lceil\frac{m-2}{\alpha}\right\rceil - 1\right)}{\frac{1}{\alpha}\left(\frac{1}{\alpha}+1\right)\left(\frac{1}{\alpha}+2\right) \cdot \cdots \cdot \left(\frac{1}{\alpha}+\left\lceil\frac{m-2}{\alpha}\right\rceil - 1\right)} = \prod_{k=0}^{\lceil (m-2)/\alpha \rceil - 1} \frac{\frac{3}{\alpha} + k}{\frac{1}{\alpha} + k}. \tag{93}$$

Using

$$\frac{\frac{3}{\alpha} + k}{\frac{1}{\alpha} + k} \leq \frac{\frac{3}{2} + k}{\frac{1}{2} + k} = \frac{3 + 2k}{1 + 2k}, \qquad \alpha > 2,\ k \geq 0 \tag{94}$$

with equality if and only if $k = 0$, and

$$\left\lceil \frac{m-2}{\alpha} \right\rceil \leq \left\lceil \frac{m}{2} - 1 \right\rceil = \frac{m}{2} - 1, \qquad \alpha > 2 \tag{95}$$

we can write

$$\frac{E\left\{\left|\breve{X}\right|^{m}\right\}}{E\left\{\left|\breve{X}\right|^{m-2}\right\}} \leq \prod_{k=0}^{\lceil (m-2)/\alpha \rceil - 1} \frac{\frac{3}{\alpha} + k}{\frac{1}{\alpha} + k} < \prod_{k=0}^{(m/2)-2} \frac{3+2k}{1+2k} = m - 1. \tag{96}$$

This completes the proof of (43). □

REFERENCES

[1] J.-F. Cardoso, "Blind signal separation: Statistical principles," *Proc. IEEE*, vol. 86, pp. 2009–2025, Oct. 1998.

[2] H. H. Yang, "Serial updating rule for blind separation derived from the method of scoring," *IEEE Trans. Signal Processing*, vol. 47, pp. 2279–2285, Aug. 1999.

[3] S.-I. Amari, A. Cichocki, and H. H. Yang, "A new learning algorithm for blind signal separation," *Adv. Neural Inform. Process. Syst.*, vol. 8, pp. 757–763, 1996.

[4] 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.

[5] J.-F. Cardoso, "On the stability of source separation algorithms," in *Proc. IEEE Int. Workshop Neural Networks Signal Process.*, Cambridge, U.K., Sept. 1998, pp. 13–22.

[6] J.-F. Cardoso and B. H. Laheld, "Equivariant adaptive source separation," *IEEE Trans. Signal Processing*, vol. 44, pp. 3017–3030, Dec. 1996.

[7] 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 Process.*, vol. 80, no. 9, pp. 1807–1822, Aug. 2000.

[8] S. C. Douglas, "Self-stabilized gradient algorithms for blind source separation with orthogonality constraints," *IEEE Trans. Neural Networks*, vol. 11, pp. 1490–1497, Nov. 2000.

[9] I. N. Bronshtein and K. A. Semendyayev, *Handbook of Mathematics*, 3rd ed. New York: Springer Verlag, 1997.

[10] J.-F. Cardoso, "Infomax and maximum likelihood for source separation," *IEEE Signal Processing Lett.*, vol. 4, pp. 112–114, Apr. 1997.

[11] A. J. Bell, "Information theory, independent-component analysis, and applications," in *Unsupervised Adaptive Filtering, Volume I: Blind Source Separation*, S. Haykin, Ed: Wiley, 2000, pp. 237–264.

[12] R. H. Lambert, "Multichannel blind deconvolution: FIR matrix algebra and separation of multipath mixtures," Ph.D. dissertation, Univ. Southern Calif., Los Angeles, 1996.

[13] M. M. Ali, "Stochastic ordering and kurtosis measure," *Amer. Statist.*, vol. 69, no. 346, pp. 543–545, June 1974.

[14] S. C. Douglas, A. Cichocki, and S. Amari, "Multichannel blind separation and deconvolution of sources with arbitrary distributions," in *Proc. IEEE Int. Workshop Neural Networks Signal Process.*, Amelia Island Plantation, FL, Sept. 1997, pp. 436–445.

[15] 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 Comput.*, vol. 11, no. 2, pp. 417–441, Feb. 1999.

[16] H. Mathis, T. P. von Hoff, and M. Joho, "Blind separation of mixed-kurtosis signals using an adaptive threshold nonlinearity," in *Proc. Int. Independent Component Anal. Blind Signal Separation*, Helsinki, Finland, June 19–22, 2000, pp. 221–226.

[17] H. Mathis and S. C. Douglas, "On optimal and universal nonlinearities for blind signal separation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 5, Salt Lake City, UT, May 7–11, 2001, pp. 2777–2780.

[18] H. Mathis, M. Joho, and G. S. Moschytz, "A simple threshold nonlinearity for blind signal separation," in *Proc. IEEE Int. Symp. Circuits Syst.*, vol. IV, Geneva, Switzerland, May 28–31, 2000, pp. 489–492.

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

[20] H. Mathis, T. P. von Hoff, and M. Joho, "Blind separation of signals with mixed kurtosis signs using threshold activation functions," *IEEE Trans. Neural Networks*, vol. 12, pp. 618–624, May 2001.

Heinz Mathis (M'02) was born in Zurich, Switzerland, in 1968. He received the B.E. and Ph.D. degrees in electrical engineering from the Swiss Federal Institute of Technology (ETH), Zurich, in 1993 and 2001, respectively.

From 1993 to 1997, he was 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. 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.

Scott C. Douglas (SM'98) received the B.S. (with distinction), M.S., and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, in 1988, 1989, and 1992, respectively.

From 1992 to 1998, he was an Assistant Professor with the Department of Electrical Engineering, University of Utah, Salt Lake City. Since August 1998, he has been with the Department of Electrical Engineering, School of Engineering and Applied Science, Southern Methodist University, Dallas, TX, as an Associate Professor. His research activities include adaptive filtering, active noise control, blind deconvolution and source separation, and VLSI/hardware implementations of digital signal processing systems.