# A PROOF OF THE NON-EXISTENCE OF UNIVERSAL NONLINEARITIES FOR BLIND SIGNAL SEPARATION

*Heinz Mathis*

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

**ABSTRACT**

The universally applicable nonlinearity for the blind separation of arbitrary source densities is one of the ultimate goals in blind signal processing. This paper provides a proof that such a single, universal nonlinearity for the separation of all non-Gaussian signals cannot exist.

## 1. INTRODUCTION

Blind separation of instantaneously mixed signals using an adaptive algorithm with a nonlinearity implicitly producing higher-order moments has been described by many researchers (e.g. [1]). Many approaches have resulted in similar update equations for the separation matrix. In particular, the maximum likelihood (ML) and the information maximization (InfoMax) [2] criterion used in a stochastic natural gradient algorithm yield

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

where $\boldsymbol{W}$ is the separation matrix used to unwind the mixing process given by the mixing matrix $\boldsymbol{A}$, so that the recovered signals are

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

$\boldsymbol{s}$, $\boldsymbol{x}$, and $\boldsymbol{u}$ denote the source-signal, the mixed-signal, and the separated-signal vectors, respectively. For both the ML and InfoMax approaches, the nonlinearity $g(u)$ is given by the score function

$$g(u) = -\frac{p'(u)}{p(u)} \tag{3}$$

where $p(.)$ and $p'(.)$ are the pdf of the source signals and its derivative, respectively. Interestingly, the minimization of the mutual information between the outputs [1] leads to the same update equation (1), but with a different nonlinearity

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

Eq. (4) is claimed to be model independent—and as such fulfills the requirements of the universal nonlinearity—since the source distributions do not influence the nonlinearity, which is based on the approximation of marginal output densities by the Gram-Charlier expansion [1].

Stability analyses of nonlinearities used in (1) by different authors (e.g., [3]) have resulted in the statement that for local stability around an equilibrium point, the output distribution must satisfy

$$E\left\{g'(U)\right\} E\left\{U^2\right\} - E\left\{g(U)U\right\} > 0. \tag{5}$$

## 2. A PROOF OF THE NON-EXISTENCE OF A UNIVERSAL NONLINEARITY

### 2.1. The Polynomial Nonlinearity

Model-independent nonlinearities such as the one given in (4) have been claimed to separate both super- and sub-Gaussian distributions [1]. An intuitive explanation of why such approaches fail, has been offered in [4] by noticing that the corresponding function to which the nonlinearity is the score function is of the wrong type. In the following, a proof is given, that such polynomial nonlinearities are due to fail.

### 2.2. Problem statement

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

$$\frac{\sigma_X^2 E\left\{g'(X)\right\}}{E\left\{Xg(X)\right\}} > 1 \tag{6}$$

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.

### 2.3. Statistical Moments Prerequisites

Consider $X$ a generalized Gaussian variable

$$p_X(x) = \frac{\alpha}{2\beta\Gamma\left(\frac{1}{\alpha}\right)} e^{-\left(\frac{|x|}{\beta}\right)^\alpha} \tag{7}$$

 

parameterized by $\alpha$ which is larger (smaller) than two for sub- (super-) Gaussian signals. We will also consider a normalized version of $X$, $\breve{X}$. $\beta$ can be found from the general expression for the $m$th-order moment of a generalized Gaussian signal [5]

$$E\left\{|X|^m\right\} = \frac{\Gamma(\frac{m+1}{\alpha})}{\Gamma(\frac{1}{\alpha})}\beta^m. \quad (8)$$

$\Gamma(.)$ is the gamma function given by $\Gamma(a) \triangleq \int_0^\infty x^{a-1}e^{-x}dx$. For $m = 2$, we have

$$\beta = \sqrt{\frac{\Gamma(\frac{1}{\alpha})}{\Gamma(\frac{3}{\alpha})}}\sigma_X. \quad (9)$$

The moments of any unit-variance random variable can be bounded by

$$E\left\{|\breve{X}|^m\right\} > (m-1)E\left\{|\breve{X}|^{m-2}\right\}, \quad X \text{ super-Gaussian distributed} \quad (10)$$

$$E\left\{|\breve{X}|^m\right\} < (m-1)E\left\{|\breve{X}|^{m-2}\right\}, \quad X \text{ sub-Gaussian distributed.} \quad (11)$$

A proof of Eqs. (10) and (11) is provided in Appendix A.1. Eq. (11) reveals immediately (by setting $m = p+1$) that the monomial function $g(u) = au^p$ is a stable nonlinearity for any sub-Gaussian signals. Eqs. (10) and (11) can be extended to distributions whose variance is unequal to one by noting

$$E\left\{|X|^m\right\} = \sigma_X^m E\left\{|\breve{X}|^m\right\} \quad (12)$$

where $\breve{X}$ is the normalized version of $X$. The even moments of generalized Gaussian variables with variance $\sigma_X^2$ are related as

$$\frac{E\{|X|^m\}}{E\left\{|X|^{m-2}\right\}} = \frac{\Gamma\left(\frac{m+1}{\alpha}\right)}{\Gamma\left(\frac{m-1}{\alpha}\right)} \cdot \frac{\Gamma\left(\frac{1}{\alpha}\right)}{\Gamma\left(\frac{3}{\alpha}\right)}\sigma_X^2. \quad (13)$$

For arbitrary distributions we have

$$E\left\{|X|^m\right\} \geq E\left\{|X|^{m-2}\right\}\sigma_X^2, \quad m \geq 4 \quad (14)$$

with equality if and only if $X$ is binary distributed. And finally,

$$E\left\{|X|^m\right\} > (m-1)E\left\{|X|^{m-2}\right\}\sigma_X^2, \quad X \text{ super-Gaussian distributed} \quad (15)$$

$$E\left\{|X|^m\right\} < (m-1)E\left\{|X|^{m-2}\right\}\sigma_X^2, \quad X \text{ sub-Gaussian distributed.} \quad (16)$$

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

### 2.4. 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)

$$g(X) = \sum_{k \text{ odd}} t_k X^k. \quad (17)$$

The Taylor series used in (17) 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

$$g'(X) = \sum_{k \text{ odd}} t_k k X^{k-1}. \quad (18)$$

If (17) and (18) are inserted into (6) we get the stability condition expressed as a function of the Taylor series coefficients

$$\frac{\sigma_X^2 \sum_{k \text{ odd}} t_k k E\left\{X^{k-1}\right\}}{\sum_{k \text{ odd}} t_k E\left\{X^{k+1}\right\}} > 1. \quad (19)$$

For strictly non-negative values of $t_k$ and $X$ a sub-Gaussian variable, stability is guaranteed due to (16) and hence

$$\sum_{k \text{ odd}} t_k E\left\{X^{k+1}\right\} < \sum_{k \text{ odd}} t_k k E\left\{X^{k-1}\right\}\sigma_X^2. \quad (20)$$

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 1**
*For the natural gradient update equation* (1)*, 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 $g(X) = t_1X + t_3X^3$ cannot separate both sub- and super-Gaussian signals, because

$$G(X) \triangleq \frac{\sigma_X^2(t_1 + 3t_3\sigma_X^2)}{\sigma_X^2 t_1 + t_3 E\left\{X^4\right\}} \quad (21)$$

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 (15). 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, too, but smaller than the numerator due to (15) 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 (16) and negative $t_3$. Hence, $G(X) < 1$ for sub-Gaussian signals.

We have seen, that a nonlinearity with only $K$ terms does not yield a stable solution. In the following we assume, that a solution with $g(X) = \sum_{\substack{k=1 \\ k \text{ odd}}}^{K} t_k X^k$ will be stable for either sub- or super-Gaussian signals but not both. A solution that is unstable for either signal distribution class cannot in one step be extended to an overall stable solution as can easily be verified, because in such a case $t_{K+2}$ would have to be chosen as positive and negative at the same time. 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

$$g(X) = \sum_{\substack{k=1 \\ k \text{ odd}}}^{K+2} t_k X^k. \tag{22}$$

The stability criterion can now be written as

$$\frac{\sigma_X^2 \left( \sum_{\substack{k=1 \\ k \text{ odd}}}^{K} t_k k E\left\{X^{k-1}\right\} + t_{K+2}(K+2)E\left\{X^{K+1}\right\} \right)}{\sum_{\substack{k=1 \\ k \text{ odd}}}^{K} t_k E\left\{X^{k+1}\right\} + t_{K+2}E\left\{X^{K+3}\right\}} > 1. \tag{23}$$

If we want to multiply both sides of (23) by the denominator, we have to distinguish two cases depending on the sign of the denominator of (23).

Case 1 (positive denominator of (23)): From (23) it follows

$$t_{K+2}(\sigma_X^2(K+2)E\left\{X^{K+1}\right\} - E\left\{X^{K+3}\right\}) > -\sum_{\substack{k=1 \\ k \text{ odd}}}^{K} t_k \left( \sigma_X^2 k E\left\{X^{k-1}\right\} - E\left\{X^{k+1}\right\} \right). \tag{24}$$

Because of (16), $\sigma_X^2(K+2)E\left\{X^{K+1}\right\} - E\left\{X^{K+3}\right\} > 0$ for sub-Gaussian signals, hence we have

$$t_{K+2} > \frac{-\sum_{\substack{k=1 \\ k \text{ odd}}}^{K} t_k \left( \sigma_X^2 k E\left\{X^{k-1}\right\} - E\left\{X^{k+1}\right\} \right)}{\sigma_X^2(K+2)E\left\{X^{K+1}\right\} - E\left\{X^{K+3}\right\}}. \tag{25}$$

But due to (15), for super-Gaussian signals we have $\sigma_X^2(K+2)E\left\{X^{K+1}\right\} - E\left\{X^{K+3}\right\} < 0$, so

$$t_{K+2} < \frac{-\sum_{\substack{k=1 \\ k \text{ odd}}}^{K} t_k \left( \sigma_X^2 k E\left\{X^{k-1}\right\} - E\left\{X^{k+1}\right\} \right)}{\sigma_X^2(K+2)E\left\{X^{K+1}\right\} - E\left\{X^{K+3}\right\}}. \tag{26}$$

In the limit for $\alpha \to 2$ for generalized Gaussian distributions, the RHS of (25) and (26) 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 $\alpha$.

Case 2 (negative denominator of (23)): According to (23)

$$t_{K+2}(\sigma_X^2(K+2)E\left\{X^{K+1}\right\} - E\left\{X^{K+3}\right\}) < -\sum_{\substack{k=1 \\ k \text{ odd}}}^{K} t_k \left( \sigma_X^2 k E\left\{X^{k-1}\right\} - E\left\{X^{k+1}\right\} \right). \tag{27}$$

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

$$t_{K+2} < \frac{-\sum_{\substack{k=1 \\ k \text{ odd}}}^{K} t_k \left( \sigma_X^2 k E\left\{X^{k-1}\right\} - E\left\{X^{k+1}\right\} \right)}{\sigma_X^2(K+2)E\left\{X^{K+1}\right\} - E\left\{X^{K+3}\right\}}. \tag{28}$$

But for super-Gaussian signals $\sigma_X^2(K+2)E\left\{X^{K+1}\right\} - E\left\{X^{K+3}\right\} < 0$, so

$$t_{K+2} > \frac{-\sum_{\substack{k=1 \\ k \text{ odd}}}^{K} t_k \left( \sigma_X^2 k E\left\{X^{k-1}\right\} - E\left\{X^{k+1}\right\} \right)}{\sigma_X^2(K+2)E\left\{X^{K+1}\right\} - E\left\{X^{K+3}\right\}}. \tag{29}$$

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

$$g(X) = \sum_{\substack{k=1 \\ k \text{ odd}}}^{K} t_k X^k \tag{30}$$

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

$$g(X) = \sum_{\substack{k=1 \\ k \text{ odd}}}^{K+2} t_k X^k \tag{31}$$

will also only separate either sub- or super-Gaussian signals (but not necessarily the same distribution class as the original nonlinearity). By induction, this applies to any arbitrarily long odd power series. □

## 3. CONCLUSIONS

The single, universal nonlinearity that separates any mixtures of non-Gaussian distributions (possibly of different kurtoses signs), does not exist.

## 4. REFERENCES


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

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

[3] J.-F. Cardoso, "On the stability of source separation algorithms," in *Proc. NNSP*, Cambridge, UK, Sept. 1998, pp. 13–22.

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

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


## A. APPENDIX

### A.1. Proof of Eqs. (10) and (11)

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 $\lfloor \frac{m}{2} \rfloor = \lceil \frac{m}{2} \rceil = \frac{m}{2}$.

*Proof:* We can write the ratio of moments of the generalized Gaussian distribution as

$$\frac{E\left\{|\check{X}|^m\right\}}{E\left\{|\check{X}|^{m-2}\right\}} = \frac{\Gamma\left(\frac{m+1}{\alpha}\right)}{\Gamma\left(\frac{m-1}{\alpha}\right)} \cdot \frac{\Gamma\left(\frac{1}{\alpha}\right)}{\Gamma\left(\frac{3}{\alpha}\right)} = \frac{\Gamma\left(\frac{3}{\alpha} + \frac{m-2}{\alpha}\right)\Gamma\left(\frac{1}{\alpha}\right)}{\Gamma\left(\frac{1}{\alpha} + \frac{m-2}{\alpha}\right)\Gamma\left(\frac{3}{\alpha}\right)}. \tag{32}$$

Consider $\alpha < 2$ first. Using the recursive property $\Gamma(a+1) = a\Gamma(a)$ we can lower bound (32) by

$$\frac{\Gamma\left(\frac{3}{\alpha} + \frac{m-2}{\alpha}\right)\Gamma\left(\frac{1}{\alpha}\right)}{\Gamma\left(\frac{1}{\alpha} + \frac{m-2}{\alpha}\right)\Gamma\left(\frac{3}{\alpha}\right)} > \frac{\Gamma\left(\frac{3}{\alpha}\right) \cdot \frac{3}{\alpha}(\frac{3}{\alpha}+1)(\frac{3}{\alpha}+2)\cdot \ldots \cdot (\frac{3}{\alpha} + \lfloor \frac{m-2}{\alpha} \rfloor - 1) \cdot \Gamma\left(\frac{1}{\alpha}\right)}{\Gamma\left(\frac{1}{\alpha}\right) \cdot \frac{1}{\alpha}(\frac{1}{\alpha}+1)(\frac{1}{\alpha}+2)\cdot \ldots \cdot (\frac{1}{\alpha} + \lfloor \frac{m-2}{\alpha} \rfloor - 1) \cdot \Gamma\left(\frac{3}{\alpha}\right)} = \prod_{k=0}^{\lfloor \frac{m-2}{\alpha} \rfloor - 1} \frac{\frac{3}{\alpha} + k}{\frac{1}{\alpha} + k}. \tag{33}$$

Using

$$\frac{\frac{3}{\alpha} + k}{\frac{1}{\alpha} + k} \geq \frac{\frac{3}{2} + k}{\frac{1}{2} + k} = \frac{3 + 2k}{1 + 2k}, \qquad \alpha < 2, k \geq 0 \tag{34}$$

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

$$\left\lfloor \frac{m-2}{\alpha} \right\rfloor \geq \left\lfloor \frac{m}{2} - 1 \right\rfloor = \frac{m}{2} - 1, \qquad \alpha < 2 \tag{35}$$

we can write

$$\frac{E\left\{|\check{X}|^m\right\}}{E\left\{|\check{X}|^{m-2}\right\}} > \prod_{k=0}^{\lfloor \frac{m-2}{\alpha} \rfloor - 1} \frac{\frac{3}{\alpha} + k}{\frac{1}{\alpha} + k} \geq \prod_{k=0}^{\lfloor \frac{m}{2} \rfloor - 2} \frac{3 + 2k}{1 + 2k} = 3 + 2\left(\left\lfloor \frac{m}{2} \right\rfloor - 2\right) = m - 1 \tag{36}$$

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 Eq. (10).

Now we consider $\alpha > 2$. Similarly to (33) we can now upper bound (32) by

$$\frac{\Gamma\left(\frac{3}{\alpha} + \frac{m-2}{\alpha}\right)\Gamma\left(\frac{1}{\alpha}\right)}{\Gamma\left(\frac{1}{\alpha} + \frac{m-2}{\alpha}\right)\Gamma\left(\frac{3}{\alpha}\right)} < \frac{\frac{3}{\alpha}(\frac{3}{\alpha}+1)(\frac{3}{\alpha}+2)\cdot \ldots \cdot (\frac{3}{\alpha} + \lceil \frac{m-2}{\alpha} \rceil - 1)}{\frac{1}{\alpha}(\frac{1}{\alpha}+1)(\frac{1}{\alpha}+2)\cdot \ldots \cdot (\frac{1}{\alpha} + \lceil \frac{m-2}{\alpha} \rceil - 1)} = \prod_{k=0}^{\lceil \frac{m-2}{\alpha} \rceil - 1} \frac{\frac{3}{\alpha} + k}{\frac{1}{\alpha} + k}. \tag{37}$$

Using

$$\frac{\frac{3}{\alpha} + k}{\frac{1}{\alpha} + k} \leq \frac{\frac{3}{2} + k}{\frac{1}{2} + k} = \frac{3 + 2k}{1 + 2k}, \qquad \alpha > 2, k \geq 0 \tag{38}$$

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

$$\left\lceil \frac{m-2}{\alpha} \right\rceil \leq \left\lceil \frac{m}{2} - 1 \right\rceil = \frac{m}{2} - 1, \qquad \alpha > 2 \tag{39}$$

we can write

$$\frac{E\left\{|\check{X}|^m\right\}}{E\left\{|\check{X}|^{m-2}\right\}} \leq \prod_{k=0}^{\lceil \frac{m-2}{\alpha} \rceil - 1} \frac{\frac{3}{\alpha} + k}{\frac{1}{\alpha} + k} < \prod_{k=0}^{\lceil \frac{m}{2} \rceil - 2} \frac{3 + 2k}{1 + 2k} = m - 1. \tag{40}$$

This completes the proof of Eq. (11). □