ON OPTIMAL AND UNIVERSAL NONLINEARITIES FOR BLIND SIGNAL SEPARATION

Heinz Mathis[1] *and Scott C. Douglas*[2]

[1] Signal and Information Processing Laboratory,
Swiss Federal Institute of Technology Zurich, Switzerland.
[2] Department of Electrical Engineering,
Southern Methodist University, Dallas, TX 75275 USA.

mathis@isi.ee.ethz.ch, *douglas@engr.smu.edu*

ABSTRACT

The search for universally applicable nonlinearities in blind signal separation has produced nonlinearities that are optimal for a given distribution, as well as nonlinearities that are most robust against model mismatch. This paper shows yet another justification for the score function, which is in some sense a very robust nonlinearity. It also shows that among the class of parameterizable nonlinearities, the threshold nonlinearity with the threshold as a parameter is able to separate any non-Gaussian distribution, a fact that is also proven in this paper.

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. Many approaches have resulted in similar update equations for the separation matrix. In particular, the maximum likelihood (ML) and the information maximization (InfoMax) criterion used in a stochastic-gradient algorithm under the natural gradient both yield [1]

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

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

$$\mathbf{u} = \mathbf{W}\mathbf{x} = \mathbf{W}\mathbf{A}\mathbf{s}. \qquad (2)$$

In Eq. (2), $\mathbf{s}$ and $\mathbf{x}$ denote the source and mixed signal vectors, respectively. For both the ML and InfoMax approaches, the nonlinearity $g(u)$ is given by the score function

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

2. OPTIMAL NONLINEARITIES

2.1. 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 [2], [3], [4] have resulted in the statement that for local stability around an equilibrium point, the signal must satisfy

$$E\left\{g'(U)\right\} E\left\{U^2\right\} - E\left\{g(U)U\right\} > 0. \qquad (4)$$

For monomial nonlinearities

$$g(u) = au^p \qquad (5)$$

the stability condition (4) 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

$$U = \sigma_U \breve{U}. \qquad (6)$$

Using (5) and (6) in (4) results in

$$p\sigma^{p+1} aE\left\{\breve{U}^{p-1}\right\} - \sigma^{p+1} aE\left\{\breve{U}^{p+1}\right\} > 0 \qquad (7)$$

which is written in terms of the nonlinearities as

$$E\left\{g'(\breve{U})\right\} - E\left\{g(\breve{U})\breve{U}\right\} > 0. \qquad (8)$$

Eq. (8) 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 the mode of a pdf is inversely proportional to its standard deviation $p_U(0) \sim 1/\sigma_U$, so the stability condition can be written as

$$\frac{E\left\{g'(U)\right\}\sigma_U^2}{E\left\{g(U)U\right\}} = \frac{2\sigma_U^2 p_U(0)}{E\left\{|U|\right\}} = \frac{2p_{\breve{U}}(0)}{E\left\{|\breve{U}|\right\}} > 1. \qquad (9)$$

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

$$E\left\{g(\breve{U})\breve{U}\right\} = 1 \qquad (10)$$

where $\breve{U}$ is a normalized random variable with the distribution of one particular source signal U but scaled to unit variance, then the stability condition (4) simplifies to

$$E\left\{g'(\breve{U})\right\} > 1. \qquad (11)$$

Note that Eq. (11) is conditioned on the scaling constraint (10). 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 (4) has been evaluated for frequently applied nonlinearities and the resulting stability regions are given in Table 1. 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. Also note that $U = \sigma_U \breve{U}$ and $\sigma_{\breve{U}} = 1$.

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

Table 1. Stability regions of some nonlinearities.

2.2. The 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 1.

An intuitive explanation of the appropriate form of the nonlinearity can be given as follows. If the nonlinearity is properly scaled, i.e. $E\left\{g(\breve{U})\breve{U}\right\} = 1$, the stability condition $E\left\{g'(\breve{U})\right\} > 1$ determines if the separating points of the nonlinearity are locally stable. To ensure stability we aim at making $E\left\{g'(\breve{U})\right\}$ 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, while sub-Gaussian distributions need nonlinearities of the form $g(u) = |u|^p u$ with $p > 0$, 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. $\text{sign}(.)$, $a\tanh(.)$, do not exhibit stability for the entire positive kurtosis plane, distributions might be constructed, for which both nonlinearities $g(u) = au^3$ and $g(u) = a\tanh(bu)$ fail [5].

2.3. Optimization of the nonlinearity

The fact that the stability of blind separation algorithms depend 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 Eq. (11). The scaling constraint

$$\int_{-\infty}^{\infty} g(u)p(u)u\,du = 1 \qquad (12)$$

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

$$\text{maximize} \int_{-\infty}^{\infty} g'(u)p(u)\,du \qquad (13)$$

$$\text{subject to} \int_{-\infty}^{\infty} g^2(u)p(u)\,du = c \qquad (14)$$

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

$$f = g'(u)p(u) + \lambda(g^2(u)p(u)) \qquad (15)$$

where λ is a Lagrange multiplier. To find the optimal $g(u)$, we have to solve the Euler-Lagrange equation [6]

$$\frac{\partial f}{\partial g} - \frac{\partial}{\partial u}\frac{\partial f}{\partial g'} - \frac{\partial}{\partial g}\frac{\partial f}{\partial g'}g' - \frac{\partial}{\partial g'}\frac{\partial f}{\partial g'}g'' = 0 \qquad (16)$$

where we abridged $p \triangleq p(u)$, $p' \triangleq p'(u) = \frac{\partial}{\partial u}p(u)$, $g \triangleq g(u)$, and $g' \triangleq g'(u) = \frac{\partial}{\partial u}g(u)$. Working out the different terms of (16) for (15) yields

$$\frac{\partial f}{\partial g} = 2\lambda gp \qquad (17)$$

$$\frac{\partial}{\partial u}\frac{\partial f}{\partial g'} = p' \qquad (18)$$

$$\frac{\partial}{\partial g}\frac{\partial f}{\partial g'}g' = 0 \qquad (19)$$

$$\frac{\partial}{\partial g'}\frac{\partial f}{\partial g'}g'' = 0. \qquad (20)$$

Using (17)–(20) in (16) results in

$$g = \frac{1}{2\lambda}\frac{p'}{p}. \qquad (21)$$

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 (21) into (12) gives us

$$\int_{-\infty}^{\infty} \frac{1}{2\lambda}\frac{p'(u)}{p(u)}up(u)\,du = \frac{1}{2\lambda}\int_{-\infty}^{\infty} up'(u)\,du = 1. \qquad (22)$$

Integrating by parts yields for the integral in (22)

$$\int_{-\infty}^{\infty} up'(u)\,du = p(u)u\Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} p(u)\,du = -1. \qquad (23)$$

The desired solution is thus $\lambda = -\frac{1}{2}$, leading to

$$g(u) = -\frac{p'(u)}{p(u)} \qquad (24)$$

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.

3. STABILIZATION OF MIXED DISTRIBUTIONS

3.1. Difficult distributions

From Table 1 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 [5]. 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. More of those challenging distributions can be constructed using quaternary symmetric signals and choosing $A_1, A_2, p_1 = \Pr(x = A_1)$, and $p_2 = \Pr(x = A_2)$ subject to the following constraints:

C1) distributional sum

$$p_1 + p_2 = \frac{1}{2} \tag{25}$$

C2) unit variance

$$p_1 A_1^2 + p_2 A_2^2 = \frac{1}{2} \tag{26}$$

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

$$p_1 A_1^4 + p_2 A_2^4 \geq \frac{3}{2} \tag{27}$$

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

$$1 - 2p_1 \tanh^2(A_1) - 2p_2 \tanh^2(A_2)$$
$$> 2p_1 A_1 \tanh(A_1) + 2p_2 A_2 \tanh(A_2). \tag{28}$$

From (25) and (26) we can express p_1 and A_1 as a function of p_2 and A_2

$$p_1 = \frac{1}{2} - p_2 \tag{29}$$

$$A_1^2 = \frac{1 - 2p_2 A_2^2}{1 - 2p_2} \tag{30}$$

with the additional constraints

$$p_2 \leq \frac{1}{2}, \quad A_2 \geq 1. \tag{31}$$

From (27) and (31) we get a lower and an upper bound for p_2

$$p_2 \geq \frac{1}{A_2^4 - 4A_2^2 + 3} \tag{32}$$

$$p_2 \leq \frac{1}{2A_2^2}. \tag{33}$$

Invoking (28) we get an additional inequality for p_2 and A_2

$$1 - (1 - 2p_2) \tanh^2\left(\sqrt{\frac{1 - 2p_2 A_2^2}{1 - 2p_2}}\right) - 2p_2 \tanh^2(A_2)$$
$$> (1 - 2p_2)\sqrt{\frac{1 - 2p_2 A_2^2}{1 - 2p_2}} \tanh\left(\sqrt{\frac{1 - 2p_2 A_2^2}{1 - 2p_2}}\right)$$
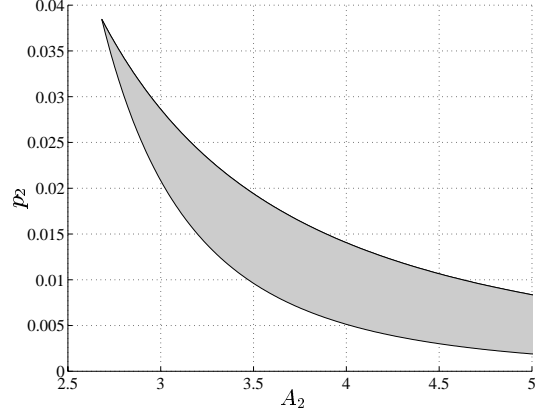$$+ 2p_2 A_2 \tanh(A_2). \tag{34}$$

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

The possible range of A_2 and p_2 is 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 $A_1 = 0.87$.

3.2. The threshold nonlinearity

The threshold nonlinearity [7]

$$g(u) = \begin{cases} 0, & |u| < \vartheta \\ a \operatorname{sign}(u), & |u| \geq \vartheta \end{cases} \tag{35}$$

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; see [7] for a definition of the fidelity criterion used. All but the threshold nonlinearity fail to separate the signals. This leads to the question if the threshold
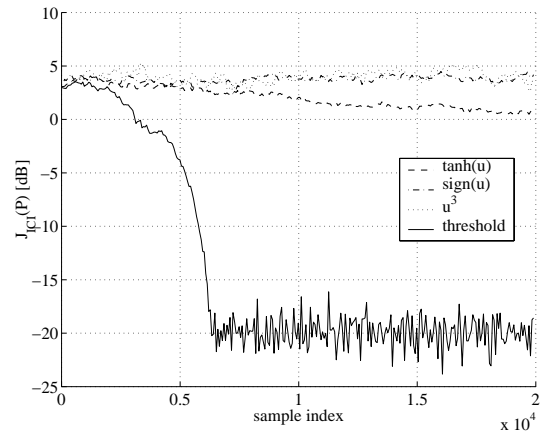

Fig. 2. Convergence of different nonlinearities for a mixture of signals exhibiting a "challenging" distribution.

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 (35) we omit scaling and obtain a more general case.

Lemma: The threshold nonlinearity given by

$$g(u) = \begin{cases} 0, & |u| < \vartheta \\ \text{sign}(u), & |u| \geq \vartheta \end{cases} \qquad (36)$$

satisfies the local stability condition

$$\sigma_X^2 p_X(\vartheta) - \int_\vartheta^\infty p_X(x)x\, dx > 0 \qquad (37)$$

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

$$\sigma_N^2 p_N(\vartheta) - \int_\vartheta^\infty p_N(x)x\, dx \equiv 0, \quad \forall \vartheta \in \mathbb{R}_0^+ \qquad (38)$$

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 (1) 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 approximating discrete distributions by low-variance Gaussian kernels. We have to show that to satisfy the stability condition (4), the inequality

$$\sigma_X^2 p_X(\vartheta) > \int_\vartheta^\infty p_X(x)x\, dx \qquad (39)$$

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 (39), so

$$\sigma_X^2 p_X(\vartheta) \leq \int_\vartheta^\infty p_X(x)x\, dx, \quad \forall \vartheta \in \mathbb{R}_0^+ \qquad (40)$$

and lead the proof by contradiction.

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

$$\sigma_N^2 p_N(\vartheta) \equiv \int_\vartheta^\infty p_N(x)x\, dx, \quad \forall \vartheta \in \mathbb{R}_0^+. \qquad (41)$$

To this end we assume that

$$\sigma_X^2 p_X(\vartheta) - \int_\vartheta^\infty p_X(x)x\, dx = c \qquad (42)$$

for some non-positive constant c. Taking derivatives of both sides of (42) with respect to ϑ gives the differential equation

$$\sigma_X^2 \frac{dp_X(\vartheta)}{d\vartheta} + \vartheta p_X(\vartheta) = 0. \qquad (43)$$

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

$$p_X(\vartheta) = K \exp\left(-\frac{\vartheta^2}{2\sigma_X^2}\right), \quad K \geq 0. \qquad (44)$$

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 LHS of the stability condition inequality. All other continuously valued and differentiable

distributions must therefore satisfy the inequality. By taking ϑ as the last (right-most) crossing point of the distribution under consideration and the normal distribution, we have either (39), which is already in contradiction to (40), or

$$\sigma_X^2 p_X(\vartheta) < \int_\vartheta^\infty p_X(x)x\, dx \qquad (45)$$

for some region around that particular ϑ. By integrating both sides of Eq. (40) over $\mathbb{R}_0^+$, we get

$$\sigma_X^2 \int_0^\infty p_X(\vartheta)\, d\vartheta = \frac{\sigma_X^2}{2} < \int_0^\infty \int_\vartheta^\infty p_X(x)\, x\, dx\, d\vartheta \qquad (46)$$

where the strict inequality results from the region where (45) is valid. The right-hand side of (46) can be solved by exchanging the integrals

$$\int_0^\infty \int_\vartheta^\infty p_X(x)\, x\, dx\, d\vartheta = \int_0^\infty \int_0^x d\vartheta\, p_X(x)\, x\, dx$$
$$= \int_0^\infty p_X(x)\, x^2\, dx = \frac{\sigma_X^2}{2}. \qquad (47)$$

Eq. (47) is a contradiction to (46). This means that if there are values of ϑ satisfying (45), due to (47) there must also be values satisfying (39) and vice versa, which is in contradiction to (40). $\square$

4. CONCLUSIONS

The score function is a robust choice for model mismatch as long as the kurtosis sign does not change. There are special distributions (with positive kurtosis), which 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.

5. REFERENCES

[1] J.-F. Cardoso, "Infomax and maximum likelihood for source separation," *IEEE Signal Processing Lett.*, vol. 4, no. 4, pp. 112–114, Apr. 1997.

[2] 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.

[3] T. P. von Hoff, A. G. Lindgren, and A. N. Kaelin, "Transpose properties in the stability and performance of the classic adaptive algorithms for blind source separation and deconvolution," *Signal Processing*, vol. 80, no. 9, pp. 1807–1822, Aug. 2000.

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

[5] S. C. Douglas, "Self-stabilized gradient algorithms for blind source separation with orthogonality constraints," *IEEE Trans. Neural Networks*, vol. 11, no. 6, pp. 1490–1497, Nov. 2000.

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

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