Reprinted from

IJCNN-2000
Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Network,
Como, Italy, 24-27 July, 2000

LOW COMPLEXITY ADAPTIVE NON-LINEAR FUNCTION FOR BLIND SIGNAL SEPARATION

Andrea Pierani[1], *Francesco Piazza*[1], *Mirko Solazzi*[1] *and Aurelio Uncini*[2]

[1] Electronics and Automatics Dept. - University of Ancona
Via Brecce Bianche, 60131 Ancona-Italy
email: upf@eealab.unian.it - URL: http://nnsp.eealab.unian.it/

[2] INFOCOM Dept. - University of Rome "La Sapienza"
Via Eudossiana, 18, 00183 Roma-Italy
email: aurel@infocom.uniroma1.it - URL: http://infocom.uniroma1.it/aurel

LOW COMPLEXITY ADAPTIVE NON-LINEAR FUNCTION FOR BLIND SIGNAL SEPARATION

Andrea Pierani[1], *Francesco Piazza*[1] , *Mirko Solazzi*[1] *and Aurelio Uncini*[2]

[1] Electronics and Automatics Dept. - University of Ancona
Via Brecce Bianche, 60131 Ancona-Italy
email: upf@eealab.unian.it - URL: http://nnsp.eealab.unian.it/

[2] INFOCOM Dept. - University of Rome "La Sapienza"
Via Eudossiana, 18, 00183 Roma-Italy
email: aurel@infocom.uniroma1.it - URL: http://infocom.uniroma1.it/aurel

ABSTRACT

In this paper a new adaptive non linear function for blind signal separation is presented. It is based on a spline approximation whose control points are adaptively changed using information maximization techniques. The monotonously increasing characteristic is obtained using suitable B-spline functions imposing simple constraints on its control points.
In particular, the problem of adaptively maximizing the entropy of the output is considered in the context of blind separation of independent sources. We derive a simple form of the learning algorithm which allows not only to adapt the separation matrix coefficients but also the shape of the non linear functions. A comparison with the Mixture–Of-Densities approach is also presented on some experimental data that demonstrates the effectiveness and efficiency of the proposed method.

1. INTRODUCTION

In the last years a great interest is raised on blind signal processing. In particular, several researchers proposed information maximization techniques, implemented in neural like architecture, for the problem of blind source separation and deconvolution of independent sources.
Bell and Sejnowski in [1], propose the use of a one-layer neural network in order to separate linear mixtures of signals. The architecture is composed by an invertible linear transformation followed by a bounded, monotonously increasing, nonlinear function applied to all outputs separately. The adaptation (or learning in neural network context) is carried-out by maximizing the output entropy. In this case, if the pdf's of the sources are known, the fixed nonlinearities should be taken equal to the cumulative density functions of the sources. In [2] the authors, giving a new explanation of work [1], underline the relevance of the output information and reinterpret the Bell and Sejnowski approach in a more general context of probability density function estimation.
However the cumulative density functions of the sources are usually not known. Although simulations exhibit good results in some cases also for nonlinearities that don't exactly match the signals, in general it is important to better estimate the exact non-linear functions.
Several approaches have been proposed to get adaptive non-linearities (see for example [3,4]) for use in blind separation problem. These functions however can present some limitations due their representation capability and/or computational complexity. In [5], in order to flatting the pdf of a signal, the authors proposed a polynomial functional link approach. The monotonously increasing characteristic of the curve is ensured by certain polynomial constraints. The polynomial shape adaptation, as discussed in [6], suffers of the "forgetting problem": changing each polynomial's coefficient produces the modification of the entire curve shape and this can cancel the information of the previous training patterns.
Recently, in order to reduce the computational burden and improve the generalization capabilities, an adaptive spline neural network has been proposed [6-7]. This architecture is shown to be suitable for signal processing applications [8], being based on a flexible Catmul-Rom spline activation function.

0-7695-0619-4/00/$10.00 (C) 2000 IEEE

The basic idea of deriving adjustable non-linear functions by using a spline approximation whose control points are adaptively changed, can be usefully applied also in the blind signal processing context, particularly using information maximization techniques. In this paper a new non-linear architecture is proposed and applied to the problem of blind signal separation. It is based on B-spline functions that allow to have only simply constraints on the control parameters in order to ensure the needed monotonously increasing characteristics.
In Section 2 the new architecture is presented while in Section 3 the derivation of the learning algorithm is reported. Some experimental results are also shown in Section 4.

2. THE B-SPLINE NON-LINEARITY

The spline activation functions (see [6]) are smooth parametric curves, divided in multiple tracts (spans) each controlled by four control points. Let y=h(u) be the non-linear function to reproduce, then the spline activation function can be expressed as:

$$y = h(u) = \bar{h}(v,i) = \sum_{i=1}^{N-3} h_i(v) \tag{1}$$

i.e. as a composition of (N-2) spans $h_i(v)$ i=1,…,N-3 (where N is the total number of the control points Q_i i=1,…,N) each depending from a local variable $v \in [0,1)$ and controlled by the $Q_i, Q_{i+1}, Q_{i+2}, Q_{i+3}$ control points. The two parameters i, v can be derived by an internal variable z

$$z = \frac{u}{\Delta u} + \frac{N-1}{2} \tag{2}$$

$$z = \begin{cases} 1 & \text{if } z < 1 \\ z & \text{if } 1 \le z \le N-3 \\ N-3 & \text{if } z > N-3 \end{cases} \tag{3}$$

where Δu is the fixed distance between two adjacent control points; the constraints imposed by equation (3) are necessary to keep the input within the active region that encloses the control points. Separating z into integer and fractional parts using the floor operator $\lfloor . \rfloor$ finally we get

$$\begin{aligned} i &= \lfloor z \rfloor \\ v &= z - i \end{aligned} \tag{4}$$

In matrix form the output can be expressed as

$$y = \bar{h}(v,i) = \mathbf{T}_v \cdot \mathbf{M} \cdot \mathbf{Q}_i \tag{5}$$

where:

$$\mathbf{T}_v = \begin{bmatrix} v^3 & v^2 & v & 1 \end{bmatrix} \qquad \mathbf{M} = \frac{1}{6}\begin{bmatrix} -1 & 3 & -3 & 1 \\ 3 & -6 & 3 & 0 \\ -3 & 0 & 3 & 0 \\ 1 & 4 & 1 & 0 \end{bmatrix} \tag{6}$$

$$\mathbf{Q}_i = \begin{bmatrix} Q_i & Q_{i+1} & Q_{i+2} & Q_{i+3} \end{bmatrix}^T \tag{7}$$

with $0 \le v < 1$ and **M** is the coefficient matrix of the B-spline version of [6]. In fact, using the B-spline matrix instead of the original Catmull-Rom spline allows to ensure the monotonously increasing characteristic of the overall function only with the following additional constraint on the control points:

$$Q_1 < Q_2 < \ldots < Q_N \qquad (8)$$

3. ADAPTATION ALGORITHM

Following an information theoretic approach, the entropy of the output ***y*** has to be maximized with respect to the control points. As an exemplary case we analyze the two input - two output separation problem where the network is composed by only 2 neurons denoted by the index *j* (*j*=1,2). Let x_j be the input of the network, u_j, y_j respectively the input and output of the non linearities, w_{jk} the weight between the k-th input and the activation value u_j of the j-th neuron. The entropy is defined as:

$$H(\mathbf{y}) = -E\{\ln p_{\mathrm{y}}(\mathbf{y})\} = E\{\ln|J|\} + H(\mathbf{x}) \qquad (9)$$

where $p_{\mathrm{y}}(\mathbf{y})$ is the pdf of the output vector **y** and J represent the Jacobian.

Since $\frac{\partial H[x_j]}{\partial Q_{j,i+m}} = 0$, posing $\frac{\partial E\{\ln|J|\}}{\partial Q_{j,i+m}} \approx \frac{\partial \ln|J|}{\partial Q_{j,i+m}}$ we get:

$$\Delta Q_{j,i+m} = \eta \frac{\partial H(y)}{\partial Q_{j,i+m}} = \eta \frac{\partial \ln|J|}{\partial Q_{j,i+m}} \qquad (10)$$

where η is the adaptation rate constant, *j* is a further index representing the neuron and *i* is the index of the span. It results

$$\ln|J| = \ln(y_1') + \ln(y_2') + \ln(D) \qquad (11)$$

with

$$D = \left(\frac{\partial u_1}{\partial x_1}\frac{\partial u_2}{\partial x_2} - \frac{\partial u_1}{\partial x_2}\frac{\partial u_2}{\partial x_1} \right) = w_{11} \cdot w_{22} - w_{12} \cdot w_{21} \qquad y_1' = \frac{\partial y_1}{\partial u_1} \qquad y_2' = \frac{\partial y_2}{\partial u_2} \qquad (12)$$

Being $\frac{\partial y_2'}{\partial Q_{1,i+m}} = 0$ and $\frac{\partial D}{\partial Q_{1,i+m}} = 0$, for the first B-spline neuron it holds:

$$\Delta Q_{1,i+m} = \eta \frac{\partial \ln|J|}{\partial Q_{1,i+m}} = \frac{1}{y_1'} \frac{\partial y_1'}{\partial Q_{1,i+m}} \qquad (13)$$

where

$$y_1' = \frac{\partial y_1}{\partial u_1} = \frac{\partial v_1}{\partial u_1}\frac{\partial y_1}{\partial v_1} = \frac{1}{\Delta u} \frac{\partial [\mathbf{T}_{1,v} \cdot \mathbf{M} \cdot \mathbf{Q}_{1,i}]}{\partial v_1} = \frac{1}{\Delta u} \dot{\mathbf{T}}_{1,v} \cdot \mathbf{M} \cdot \mathbf{Q}_{1,i} \qquad (14)$$

with

$$\dot{\mathbf{T}}_{1,v} = \frac{d\,\mathbf{T}_{1,v}}{dv} = \left[3v^2 \quad 2v \;\; 1 \;\; 0\right].$$

From (14) it results

$$\frac{\partial y_1^{'}}{\partial Q_{1,i+m}} = \frac{1}{\Delta u}\dot{\mathbf{T}}_{1,v} \cdot \mathbf{M}_m \tag{15}$$

where $\mathbf{M}_m$ represents the (m+1)-th column of matrix $\mathbf{M}$. The final adaptation formula for the control points of the first neuron is therefore

$$\Delta Q_{1,i+m} = \eta \frac{\dot{\mathbf{T}}_{1,v} \cdot \mathbf{M}_m}{\dot{\mathbf{T}}_{1,v} \cdot \mathbf{M} \cdot \mathbf{Q}_{1,i}} \tag{16}$$

similarly for the second neuron control points we get:

$$\Delta Q_{2,i+m} = \eta \frac{\dot{\mathbf{T}}_{2,v} \cdot \mathbf{M}_m}{\dot{\mathbf{T}}_{2,v} \cdot \mathbf{M} \cdot \mathbf{Q}_{2,i}} \tag{17}$$

The weights of the network are adapted following [4]. Let be **W** the weight matrix and *F* the following cost fucntion:

$$F(\mathbf{W}) = \int_{\mathbf{u}} p_u(\mathbf{u}) \;\; \log \frac{p_u(\mathbf{u})}{\prod_{i=1}^{n} g_i(u_i)} d\mathbf{u} \tag{18}$$

where $g_i(u_i) = f_i^{'}(u_i)$ and f_i are the non linear functions and n is 2 for our exemplary case. Since

$$p_u(\mathbf{u}) = p_y(\mathbf{y}) \;\; \prod_{i=1}^{n} g_i(u_i) \tag{19}$$

we get

$$F(\mathbf{W}) = \int_{\mathbf{y}} p_y(\mathbf{y}) \;\; \log p_y(\mathbf{y}) d\mathbf{y} = -H(\mathbf{y}) \tag{20}$$

Minimizing $F(\mathbf{W})$ is equivalent to maximizing the entropy $H(\mathbf{y})$:

$$\nabla_W F(\mathbf{W}) = -E\left\{\left[\mathbf{W}^T\right]^{-1} + \mathbf{h}(\mathbf{u})\mathbf{x}^T\right\} \tag{21}$$

where

$$\mathbf{h}(\mathbf{u}) = \left[h_1(u_1), \ldots, h_n(u_n)\right]^T \qquad\qquad h_i(u_i) = \frac{g_i^{'}(u_i)}{g_i(u_i)}.$$

Using the natural gradient as in [9], the final weight update rule is therefore:

$$\Delta\mathbf{W} = \mu \;\; [\mathbf{I} + \mathbf{h}(\mathbf{u})\mathbf{u}^T] \;\; \mathbf{W}$$

3. EXPERIMENTAL RESULTS

Several different experiments have been carried out on input signals $x(t)$ sampled at 16kHz with various $p_x(x)$. For all the experiments an adaptive spline structure with 24 control points is used, while the initial shape of the function is a typical hyperbolic tangent. Three different experiments of signal separation are reported here (see fig. 1): (a) male and female speech, (b) female speech and uniform noise, (c) female speech and bi-constant noise. The learning has been performed on 60'000 samples repeated 5 times for a total of 300'000 adaptation steps. The obtained performance has been compared with those of the Mixture-Of-Densities method [4] and the method reported in [5]. Although the performance are quite similar, the required computational complexities are different; as reported in tables 1 and 2 the proposed method shows a superior performance in terms of complexity both in the forward and learning phase.

	Multiply	Division	Non-Linearity
Adapt. BSF [5]	$3m$-2	1	1 tanh
MOD [4]	$2m$	m	m exp
Proposed	18	1	-

Table 1 - Computational complexity of forward phase (output computation).

	Multiply	Division	Non-Linearity
Adapt. BSF [5]	$12m$-2	m	1 tanh
MOD [4]	m^2+14m	$2m$	$2m$ exp
Proposed	13	1	-

Table 2 - Computational complexity of the adaptation algorithm

5. REFERENCES

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

[2] J.Dehaene and N. Twun-Danso, "Local Adaptive Algorithm for Information Maximization in Neural Networks, and Application to Source Separation", *Proc. of International Conference on Acoustic, Speech and Signal Processing (ICASSP97)*, pp. 59-62, Munich, DE, April , 1997.

[3] C.C. Cheung, L. Xu, "Separation of two independent sources by the information-theoretic approach with cubic non-linearity", *Proc. of IEEE ICNN-97*, *Houston TX, USA*, vol. 4, pp. 2239-2244, June 1997.

[4] L. Xu, C.C. Cheung, H.H. Yang, S. Amari, "Independent component analysis by the information theoretic approach with mixture of densities", *Proc. of IEEE ICNN-97*, Houston TX, USA, vol. 3, pp. 1821-1826, June 1997.

[5] S. Fiori, P. Bucciarelli, F. Piazza, "Blind Signal Flatting Using Warping Neural Models", *Proc. of IJCNN 98,* vol.3,pp. 2312-2317, May 1998.

[6] S. Guarnieri, F. Piazza and A. Uncini, "Multilayer Feedforward Networks with Adaptive Spline Activation Function", *IEEE Trans. on Neural Network*, Vol. 10, No. 3, pp.672-683, May 1999

[7] L. Vecci, F. Piazza, A. Uncini, "Learning and approximation Capabilities of Adaptive Spline Activation Function Neural Networks", *Neural Networks*, Vol. 11, No. 2, pp.259-270, March 1998.

[8] A. Uncini, L. Vecci, P. Campolucci and F. Piazza, "Complex-valued Neural Networks with Adaptive Spline Activation Function for Digital Radio Links Nonlinear Equalization", IEEE Trans. on Signal Processing, Vol. 47, No. 2, February 1999.

[9] S. Amari, A. Cichocki, H. H. Yang, "A New Learning Algorithm for Blind Signal Separation", Advances in Neural Information Processing Systems 8, Editors D. Touretzky, M. Mozer and M. Hasselmo, pp. 757-763, MIT Press, Cambridge MA, 1996.

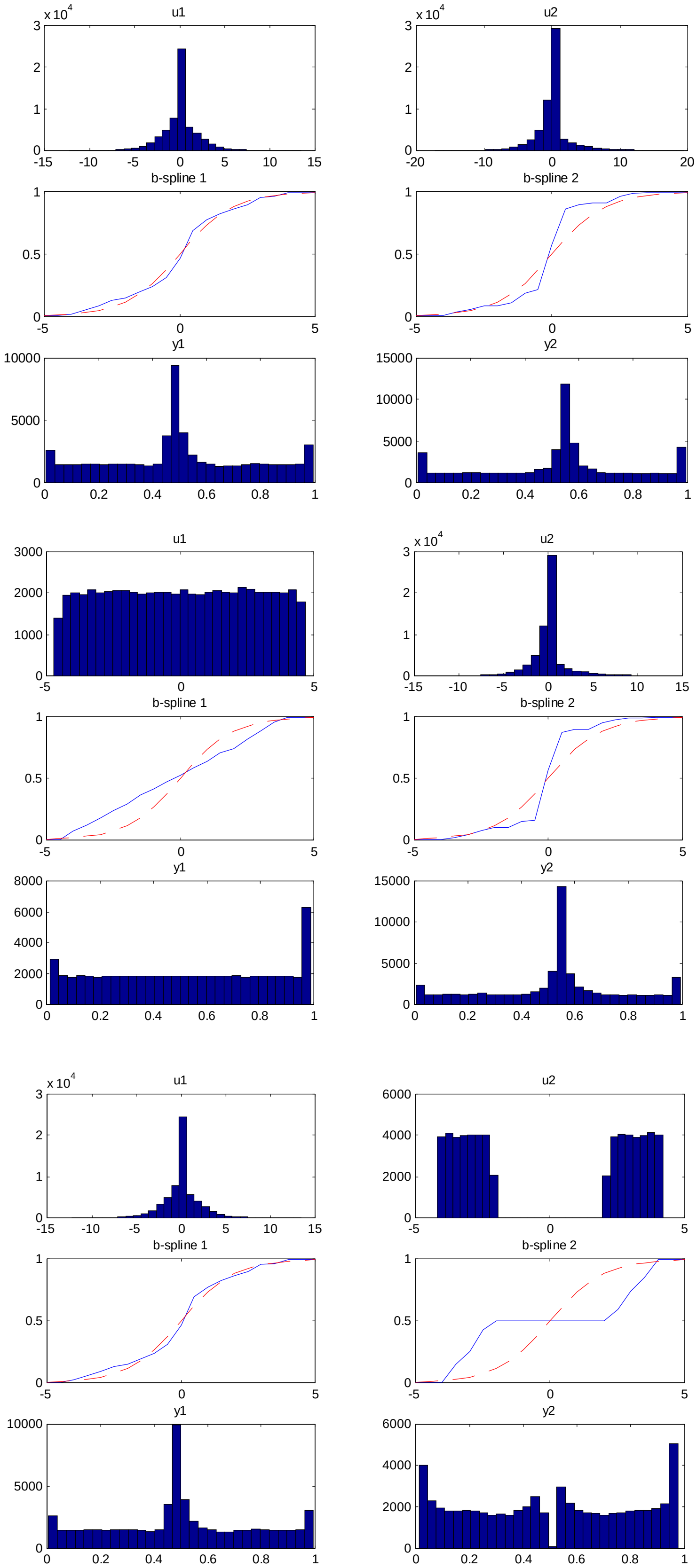

Figure 1 Input pdf, function shape and output pdf for the two channels after adaptation: (a) male and female speech, (b) female speech and uniform noise, (c) female speech and bi-constant noise.

(a)

(b)

(c)