Appeared in journal: **Neural Network World**
Vol.6, No.4, 1996, pp.515–523.
Published by: **IDG Co., Prague, Czech Republic.**

LOCAL ADAPTIVE LEARNING ALGORITHMS FOR BLIND SEPARATION OF NATURAL IMAGES

Andrzej CICHOCKI, Włodzimierz KASPRZAK *

Abstract: In this paper a neural network approach for reconstruction of natural highly correlated images from linear (additive) mixture of them is proposed. A multi–layer architecture with local on–line learning rules is developed to solve the problem of *blind separation* of sources. The main motivation for using a multi–layer network instead of a single–layer one is to improve the performance and robustness of separation, while applying a very simple local learning rule, which is biologically plausible. Moreover such architecture with on–chip learning is relatively easy implementable using VLSI electronic circuits. Furthermore it enables the extraction of source signals sequentially one after the other, starting from the strongest signal and finishing with the weakest one. The experimental part focuses on separating highly correlated human faces from mixture of them, with additive noise and under unknown number of sources.

1. Introduction

Blind signal processing and especially blind separation of sources [1, 2, 3, 4, 5, 6, 7], is a new emerging field of research with many potential applications in science and technology. Most of the developed methods provide large cross–talking or even fail to separate sources if many signals of similar distributions are mixed together or when the problem is ill–posed (i.e. the mixing matrix is near singular). In such a case even robust algorithms (e.g. [2, 3, 8]) may not be able to separate the sources with considerable quality.

The main assumption of this paper is that a single layer is not able optimally to solve this difficult problem and therefore a multi–layer architecture is required instead. We propose and test two types of learning algorithms for a multi–layer artificial neural network (ANN). The first algorithm is based on a recently developed adaptive local learning rule [9]. In this learning rule it is assumed that synaptic weights modification is performed on basis of locally available information only.

[1]Andrzej CICHOCKI, Włodzimierz KASPRZAK*
Institute of Chemical and Physical Research (RIKEN), Frontier Research Program, Brain Information Processing Group, 2–1 Hirosawa, Wako–shi, Saitama 351–01, JAPAN.

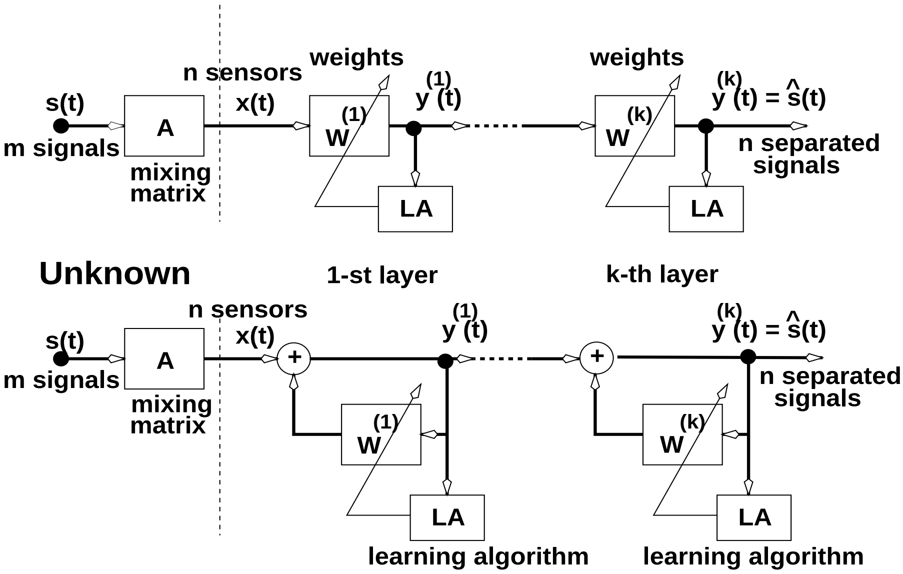

Fig. 1. *Two multi–layer neural network architectures for blind source separation: feed–forward and recurrent structure.*

An important question considered here is how optimally to select different non–linear activation functions in every consecutive layer from a set of reliable functions and how automatically to determine the learning rate, to ensure high or optimal convergence speed.

The second algorithm is applied in a post–processing layer. Its main task is determination of source number, i.e. the elimination of redundant signals in the case when the number of sensors is greater than the (unknown) number of sources. An auxiliary task of the post–processing layer if further to improve the performance of separation, i.e. to reduce cross–talking.

2. Problem formulation and main motivation

The problem of blind separation of sources in its simplest form can be formulated as follows. Let us assume an unknown number of independent sources are linearly mixed by unknown mixing coefficients according to the matrix equation $\mathbf{x}(\mathbf{t}) = \boldsymbol{A}\boldsymbol{s}(\mathbf{t})$, where $\boldsymbol{s}(t) = [s_1(t), ..., s_m(t)]^T$ is a vector of *unknown* primary sources, $\boldsymbol{A} \in R^{n\times m}$ is an unknown $n \times m$ full rank mixing matrix and $\boldsymbol{x}(t) = [x_1(t), ..., x_n(t)]^T$ is the *observed (measured)* vector of mixed (sensor) signals. It is desired to recover the primary source signals (or strictly speaking their waveforms or shapes) on the basis of sensor signals $\boldsymbol{x}(t)$.

Let us first assume that the number of sensors is equal to the number of stochastically independent sources, i.e $m = n$. The main design criteria are: type of neural networks (feed-forward, recurrent, hybrid) (see Fig. 1), number of layers and associated learning algorithms.

Here we propose the use of a multi–layer network for the blind separation problem (Fig. 1). Several facts strongly support the idea of multi–layer vs. one–layer neural networks:

1. For a multi–layer architecture a simplified local learning rule can be applied, which is biologically justified, whereas this kind of rule is usually not sufficient for a single–layer architecture.

2. While applying several layers different activation functions can be used in each layer, thus we have more flexibility and better ability to cancel higher order moments or cross–cumulants.

3. The multi–layer with local learning allows an ordered separation of sources from their mixture, i.e. in the first layer the signal with strongest energy in the mixture is completely extracted and in the last one – the weakest energy signal.

In this paper also a more general case is considered, where there are more sensors (mixtures) than measured sources, i.e. $\boldsymbol{A} \in R^{n\times m}$ and number of primary sources m is unknown. In this case the standard neural network will generate a redundant output signal set. For highly correlated sources even large cross–talking effect in output signals could appear. Hence a simple solution of redundancy elimination like pair–wise similarity tests between output signals would not work. For this task we propose an additional post–processing layer with a suitable local learning rule.

3. Multi–Layer Neural Network

3.1 Learning rules for independent component analysis

It is well known that *pre–whitening*, i.e. preprocessing or decorrelation, is usually necessary for any nonlinear PCA or PSA based separation [10, 11]. In this section we consider two different learning algorithms where each of them is at the same time performing a generalized (nonlinear) pre–whitening (called also *sphering* or *normalized orthogonalization*) and the required blind separation. Depending on condition number of the mixing matrix $\boldsymbol{A}$, the number of mixed sources and on their scaling factors, one or more layers of a neural network have to be applied.

A multi–layer feed–forward neural network performs the linear transformation $\boldsymbol{y}(t) = \boldsymbol{W}^{(k)}(t)...\boldsymbol{W}^{(1)}(t)\boldsymbol{x}(t)$, where $\boldsymbol{W}^{(l)}(t), (l = 1,...,k)$ is a $n \times n$ matrix of synaptic weights of the l–th layer (see Fig. 1 (a)). The synaptic weight matrix in each layer can be updated according to the on–line nonlinear global learning rule [1, 2, 8]:

$$\boldsymbol{W}(t+1) = \boldsymbol{W}(t) + \eta(t)[\boldsymbol{I} - \boldsymbol{f}(\boldsymbol{y}(t))\boldsymbol{g}^T(\boldsymbol{y}(t))]\boldsymbol{W}(t) \tag{1}$$

or alternatively by the simplified local learning rule [9]:

$$\boldsymbol{W}(t+1) = \boldsymbol{W}(t) + \eta(t)[\boldsymbol{I} - \boldsymbol{f}(\boldsymbol{y}(t))\boldsymbol{g}^T(\boldsymbol{y}(t))], \tag{2}$$

which can be written in an equivalent scalar form as:

$$w_{ij}(t+1) = w_{ij}(t) + \eta(t)[\delta_{ij} - f(y_i(t))g(y_j(t))]. \tag{3}$$

In these formulas $\eta(t)$ is the learning rate ($\eta(t) > 0$), $\boldsymbol{I}$ means the $n \times n$ identity matrix, $\boldsymbol{f}(\boldsymbol{y}) = [f(y_1),...,f(y_n)]^T$ and $\boldsymbol{g}(\boldsymbol{y}) = [g(y_1),...,g(y_n)]^T$ are vectors of non-linear functions (for example $f(y_j) = y^3$, $g(y_j) = tanh(y_j)$), and T is the transpose of a vector. Furthermore, for simplicity the superscript (l) indicating the layer was omitted. It can be proved that these learning algorithms perform simultaneously *whitening* and *independent component analysis*, i.e. the separation of sources.

Let us also notice that a standard linear pre–whitening adaptive rule takes the form :

$$\boldsymbol{W}(t+1) = \boldsymbol{W}(t) + \eta(t)[\boldsymbol{I} - \boldsymbol{y}(t)\boldsymbol{y}^T(t)]\boldsymbol{W}(t). \tag{4}$$

Thus in our generalization of the pre–whitening rule nonlinear functions $\boldsymbol{f}(\boldsymbol{y})$ and $\boldsymbol{g}(\boldsymbol{y})$ are used. Moreover, a simplification to the form of local learning rule (eq. 2)) is proposed. However, this biologically more justified rule than the global one (eq. 1) is less robust. Usually we are not able to separate a mixture of several signals in one single layer, even after a very long learning time.

It is interesting to note that the local learning rule (2) is valid for feed–forward, recurrent and hybrid (i.e. feed–forward and recurrent layers) architectures (Fig. 1).

3.2 Multi–layer networks with various functions and learning rates

But even with the global learning rule (1) a single layer does not ensure complete separation in the case of highly correlated sources. Typically a relatively large cross–talking between separated signals may be observed. This effect grows, if the number of mixed sources (and the number of mixtures) is steadily growing. In order to improve the performance of separation (reduction of cross–talking and noise) we iterate the separation, but applying different activation functions in next steps than in the first (multi–)layer. Typical *activation functions* are [1, 4, 12]:

1. for separation of signals with negative kurtosis:

$$\begin{array}{ll} f(y) = y^3 & g(y) = y, \\ f(y) = y & g(y) = \tanh(10y), \\ f(y) = y^3 & g(y) = \tanh(10y), \\ f(y) = 1/6y^7 - 0.5y^5 - 0.5y^3 & g(y) = y. \end{array} \tag{5}$$

2. for separation of signals with positive kurtosis:

$$f(y) = \tanh(10y) \qquad g(y) = y^3. \tag{6}$$

The nonlinear activation functions $f(y_j)$ should be chosen appropriately in order to minimize cross–talking and ensure convergence of the algorithm. On basis of signals extracted in the first layer one usually has some rough information about source signals. By estimating probability density functions (pdf) of the output signals one could choose suboptimal nonlinearities for use in the next layer [3, 5, 6]. Due to limit of space details are out of scope of this paper.

The time of learning is steadily decreasing from layer to layer. Actually the same effect can be achieved if the learning rate is faster decreasing to zero from layer to layer. In our computer experiments we start with 5–10 epochs of image data and we steadily reduce the number of epochs to 2 or 1 (in the final layer).

3.3 Learning rule for redundancy elimination

In order automatically to eliminate redundant output signals under existing cross–talking effects we apply an additional post–processing layer. This layer is described by the linear transformation $\boldsymbol{z}(t) = \widetilde{\boldsymbol{W}}(t)\boldsymbol{y}(t)$, where synaptic weights are updated using the following adaptive local learning algorithm:

$$\widetilde{w}_{ii} = 1, \quad \widetilde{w}_{ij}(t+1) = \widetilde{w}_{ij}(t) - \eta(t) z_i(t) g[z_j(t)], \tag{7}$$

where $g(z)$ is a nonlinear odd activation function (e.g. $g(z) = \tanh(10z)$). This layer not only eliminates (suppresses) redundant signals but also slightly improves the separation performance.

4. Computer Simulation Experiments

The proposed approach has been tested by computer simulation experiments on natural images. As most natural images have the same or similar first principal component, whereas the synthetic images have not this property, the separation of natural images is much more difficult than the separation of synthetic ones. In this section several examples of face image separation are presented that demonstrate the validity and robustness of proposed method.

4.1 Multi–layer separation

In the first experiment we have mixed four face images and a Gaussian noise (see Fig.2). By this example we observe the progress of a multi–layer separation with different activation functions. Four face images have been successfully separated in first three layers with a highly nonlinear activation function. In the next two layers the two activation functions $\boldsymbol{f}, \boldsymbol{g}$ have been replaced by more linear ones. This allows an improvement of the performance index PI [2, 6]. Already after two such layers the best possible separation is nearly reached.

In the second experiment four original images *Susie, Lenna, Ali, Cindy*, were mixed by ill–conditioned matrices $\boldsymbol{A}$, which are assumed to be unknown. If the images are mixed with similar energies already two layers with the local learning rule (eq. 2) were sufficient for separation.

In case of different energy mixtures four layers with the local learning rule (eq. 2) have been applied. The strongest image is already visible in all mixtures. In this case the remaining three signals are separated one after the other starting with the second strongest signal first and finishing with the weakest signal contained in the mixture set.

We compare the achieved separation performance with the results of applying one layer that performs either the global rule learning (eq. 1) or the standard pre–whitening rule only (eq. 4). The appropriate quantitative results of these three approaches are provided in Tab. 1.

Method	Susie	Ali	Lenna	Cindy
	SNR[dB]	SNR[dB]	SNR[dB]	SNR[dB]
pre–whitening (eq. 4)	10.99	5.82	18.37	3.27
global rule (eq. 1)	18.20	19.82	20.88	14.11
local rule (eq. 2)	18.16	20.60	21.04	12.50

Tab. 1. *Performance factors of applied multi–layer network with local learning rule versus a single layer with global learning rule or a standard pre–whitening layer, in case of a complete signal mixture.*

$NMSE: 1.84\ \%$	$3.47\ \%$	$0.12\ \%$	$0.98\ \%$	$2.73\ \%$
$SNR: 17.35\ dB$	$14.60\ dB$	$29.15\ dB$	$20.09\ dB$	$15.63\ dB$

Fig. 2. *A mixing and separation of natural face images and Gaussian noise: (top row) five original images (but unknown to the neural net), (second row) five input images for separation, (third row) separation results after three layers, (bottom row) five separated images after five layers ($PI = 0.144$).*

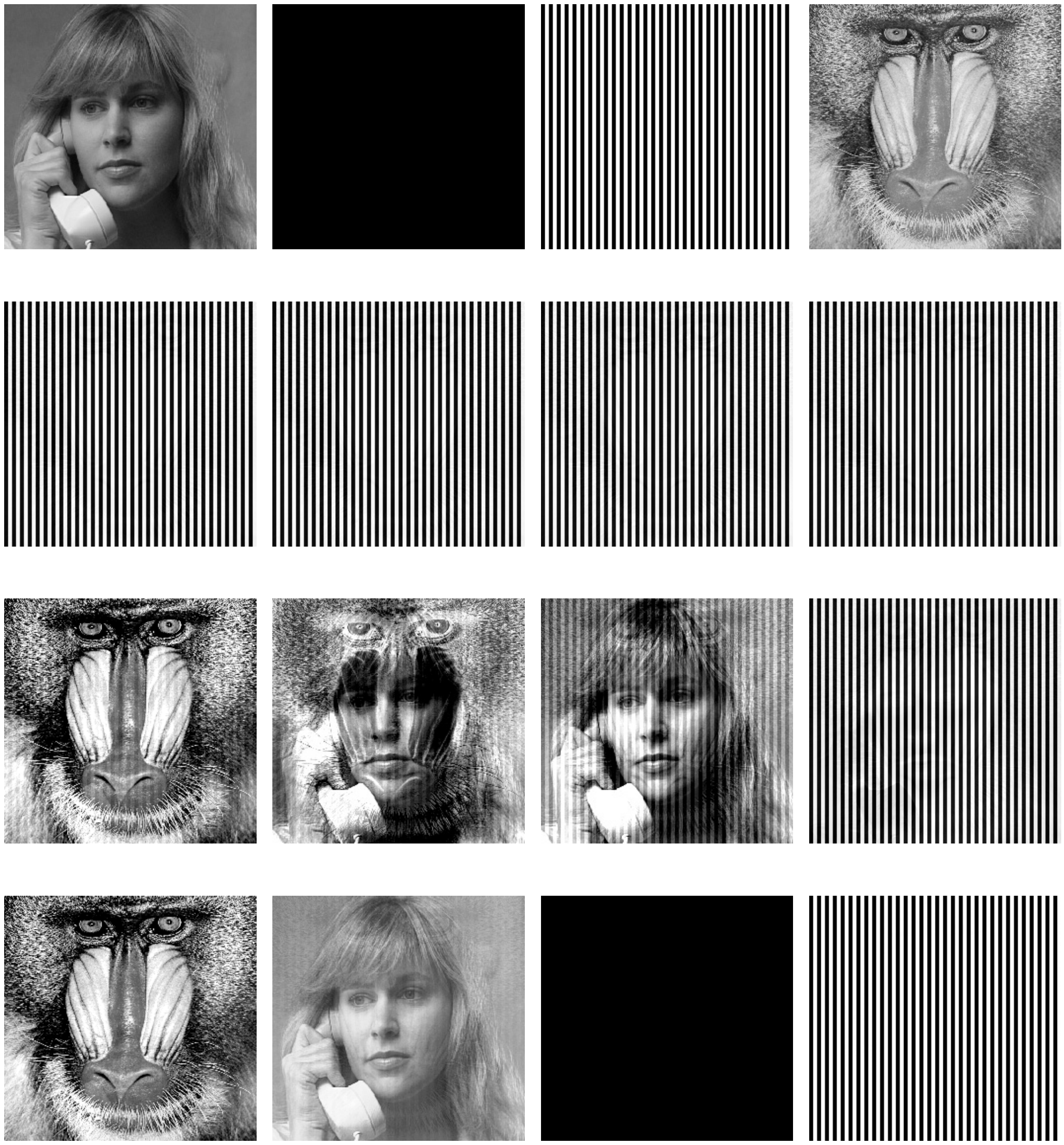

Fig. 3. *Blind separation in the case when the number of sources is smaller than the number of mixed images. The images Susie and Baboon are mixed with factors approximately 200 and 20 times weaker, respectively, than the image Stripes. After the first processing stage there are two images Susie detected with large cross–talking with the Baboon and Stripes images. After the second processing stage one signal is suppressed, indicating that only 3 sources were mixed. Also the cross–talking effects are almost completely eliminated.*

4.2 Redundancy reduction

An experiment with more sensors than sources is illustrated in Fig. 3. Four original images – *Susie, Gaussian noise, Stripes, Baboon*, given in top row of Fig. 3 – were mixed by ill–conditioned matrices $\boldsymbol{A}$ which are assumed to be unknown.

Quantitative results are provided in Tab. 2. It contains performance factors for signal separation if a multi–layer local rule learning method [9] with a post–processing redundancy elimination layer was applied.

After	PI	Susie		Stripes		Baboon	
layer	[]	NMSE	PSNR	NMSE	PSNR	NMSE	PSNR
L1	3.194	-	-	0.2120	6.73	0.0385	23.50
L2	0.733	0.0220	27.42	0.0001	42.45	0.0120	18.56
L3	0.532	0.0116	30.18	0.0016	37.20	0.0171	17.68
P	0.217	0.0153	29.00	0.0002	37.22	0.0066	31.08

Tab. 2. *Performance factors for the incomplete source case of a multi–layer local separation (L1–L3) with final redundancy reduction layer (P) (see Fig. 3).*

The post–processing layer is not disturbing already clear separated images. Even if instead of mixtures the original sources are given to this layer the results are of good quality.

5. Summary

In this paper we have proposed and tested a multi–layer neural network with different associated learning algorithms for the problem of blind separation of several highly correlated sources like natural face images. The perspective application of the blind separation method for image restoration and enhancement was demonstrated.

A local learning algorithm is biologically plausible and allows an ordered separation of sources, starting from the signals of highest energy in the mixtures and ending with the smallest energy source. We have used many successive layers with different nonlinear activation functions in an efficient way. A fuzzy–function for selection of a proper activation function improves the quality of separation and highly reduces the cross–talking effects.

An additional layer with associated learning rule has been proposed in order automatically to eliminate redundancy of the separated signal set. This occurs if the number of sources is less than the number of sensors.

References

[1] S.-I. Amari, A. Cichocki, and H.H. Yang. A new learning algorithm for blind signal separation. In *NIPS'95*, vol. 8, 1996, MIT Press, Cambridge, Mass., (in print).

[2] S.-I. Amari, A. Cichocki, and H.H. Yang. Recurrent neural networks for blind separation of sources. In *Proceedings. NOLTA'95*, NTA Research Society of IEICE, Tokyo, Japan, 1995, 37–42.

[3] J.F. Cardoso, A. Belouchrani, and B. Laheld. A new composite criterion for adaptive and iterative blind source separation. In *Proceedings ICASSP–94*, vol.4, 1994, 273–276.

[4] Ch. Jutten and J. Herault. Blind separation of sources. An adaptive algorithm based on neuromimetic architecture. *Signal Processing*, 24(1991), No. 1, 1–31.

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

[6] E. Moreau and O. Macchi. A complex self–adaptive algorithm for source separation based on high order contrasts. In *Proceedings. EUSIPCO–94*, Lausanne, EUSIP Association, 1994, 1157–1160.

[7] A. Cichocki and R. Unbehauen. *Neural Networks for Optimization and Signal Processing.* John Wiley & Sons, New York, 1994.

[8] A. Cichocki, R. Unbehauen, L. Moszczynski, and E. Rummert. A new on–line adaptive learning algorithm for blind separation of source signals. In *Proceedings. ISANN–94*, Taiwan, 1994, 406–411.

[9] A. Cichocki, W. Kasprzak, and S. Amari. Multi–layer neural networks with a local adaptive learning rule for blind separation of source signals. In *Proceedings. NOLTA'95*, NTA Research Society of IEICE, Tokyo, Japan, 1995, 61–65.

[10] J. Karhunen and J. Joutsensalo. Representation and separation of signals using nonlinear pca type learning. *Neural Networks*, 7(1994), No. 1, 113–127.

[11] E. Oja and J. Karhunen. Signal separation by nonlinear hebbian learning. In *Proceedings ICNN–95*, Perth, Australia, 1995, 417–421.

[12] A. Cichocki, R. Bogner, and L. Moszczynski. Improved adaptive algorithms for blind separation of sources. In *Proceedings of 18–th Conference on Circuit Theory and Electronic Circuits (KKTOiUE–95)*, Polana Zgorzelisko, Poland, 1995, 648–652.