International Symposium on Nonlinear Theory and Its Applications (NOLTA '95)
NTA Research Society of IEICE, Tokyo, Japan, 1995, pp.61–65

Multi–Layer Neural Networks with a Local Adaptive Learning Rule for Blind Separation of Source Signals

Andrzej Cichocki [1] Włodzimierz Kasprzak Shun–ichi Amari

FRP RIKEN (Institute of Physical and Chemical Research)
Laboratories for Artificial Brain Systems and Information Representation
Hirosawa 2–1, Saitama 351–01, Wako–shi, JAPAN
E–mail: cia@kamo.riken.go.jp

ABSTRACT

In this contribution a class of simple local unsupervised learning algorithms is proposed for multi–layer neural network performing source signal separation from linear mixture of them (the *blind separation* problem). The main motivation for using a multi–layer network instead of a single layer one for the blind separation problem is to improve the performance and robustness of separation while applying local learning rules. These rules are biologically justified opposite to existing more complex global learning rules. The proposed algorithms allow the separation of badly scaled signals and in case of ill–conditioned problems (if very similar mixtures of sources are available only). The application of developed methods for image enhancement is demonstrated.

I. Introduction

Blind separation of sources is a new emerging field of research with many potential applications, e.g. signal/image processing, telecommunication, biomedical science. There are many applications in which multiple signal observations are available, where each observation is a linear superposition of independent signals from different sources. It is desired to process these observations in such a way that the outputs correspond to the primary source signals [1] – [9].

In this paper a neural network approach is considered, first developed by Jutten and Herault [6]. Recently robust methods with global learning rules have been proposed, that can handle even ill–conditioned mixtures of signals [1], [2], [4].

But these methods do not ensure simplicity and locality of learning algorithms. The other kind of methods are at the other hand neither robust nor efficient in case of badly scaled and ill–conditioned problems [3], [6], [8].

The main purpose of this paper is to propose multi–layer models and to develop a family of associated adaptive learning algorithms which have special advantages in the cases of badly scaled and/or ill–conditioned problems. (i.e. for which previous biologically plausible algorithms had difficulties) [6]. The proposed algorithms could be considered as generalizations of the *Hebbian rule* (section II). Starting from the modified Jutten – Herault recursive neural network model a class of local on–line adaptive learning algorithms is proposed for the multi–layer feed–forward and recurrent architectures (section III). Possible extensions of proposed algorithms are described in section IV. The stability and efficiency of this new approach is demonstrated in section V for the image enhancement task.

II. Neural Networks For Blind Separation of Sources

Let us assume several independent source signals $s_j(t)$ $(j = 1, 2, \ldots, n)$ are linearly combined via unknown mixing coefficients (parameters) a_{ij} to form observations

$$x_i(t) = \sum_{j=1}^{n} a_{ij} s_j(t), (\text{ i.e. } \boldsymbol{x}(t) = \boldsymbol{A}\boldsymbol{s}(t)) \qquad (1)$$

It is required to estimate the synaptic weights w_{ij}^l, of a multi–layer feed forward neural network, to combine observations in each layer to form optimal estimates of the sources (see Fig. 1(a))

$$\hat{s}_p^l(t) = y_p^l(t) = \sum_{i=1}^{n} w_{pi}^l y_i^{l-1}(t), \qquad (2)$$

$$(\text{ i.e. } \boldsymbol{y}^l(t) = \boldsymbol{W}^l(t)\boldsymbol{y}^{l-1}(t)), \qquad (3)$$

[1] On leave from Warsaw University of Technology, Department of Electrical Engineering, Warsaw, Poland.

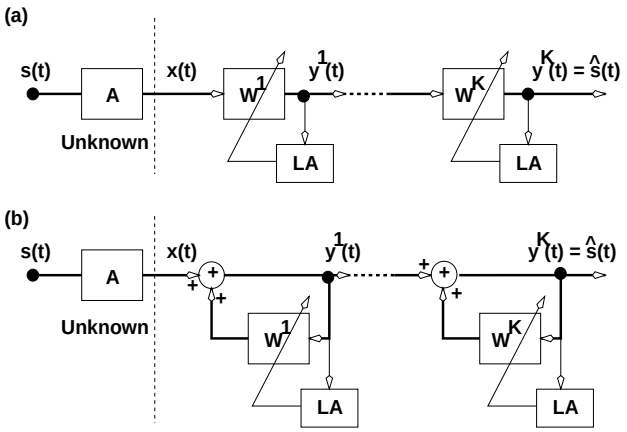


Figure 1: Feed-forward (a) and feedback (recurrent) (b) multi–layer neural networks with local learning algorithms (LA).

where $l = 1, 2, \ldots, K$ and $y_i^0(t) = x_i(t)$. The optimal weights correspond to the statistical independence of the output signals $y_p(t) = y_p^K(t)$ and they simultaneously ensure self–normalization of these signals.

It should be noted that the source separation is achieved as soon as the composite matrix

$$\boldsymbol{P}(t) = \boldsymbol{W}^K(t) \ldots \boldsymbol{W}^1(t)\boldsymbol{A} \tag{4}$$

has exactly only one nonzero element in each row and each column [6].

The above problem is designated as the *blind separation of sources* (BS). It is equivalent to the waveform–preserving *blind estimation of multiple independent sources*, where the problem is to estimate the multiple source signals from an array of sensors without knowing the mixing parameters $\{a_{ij}\}$ from the mixing matrix $\boldsymbol{A}$. No more is assumed to be known about the source signals besides that they are mutually independent and zero–mean signals, and that at most one source have a Gaussian distribution.

III. Local Adaptive Learning Algorithm

Let us consider a multi–layer feed–forward neural network described as

$$\boldsymbol{y}^l(t) = \boldsymbol{W}(t)\boldsymbol{y}^{l-1}(t); l = 1, 2, ..., K;\ \boldsymbol{y}^0(t) = \boldsymbol{x}(t) \tag{5}$$

Our objective is to propose an adaptive learning algorithm for synaptic weights $\{w_{ij}^l\}$. Starting from the independence and self–normalization criteria we have developed a local learning algorithm for such neural network:

$$\frac{dw_{ii}^l(t)}{dt} = -\mu(t)\left[f[y_i^l(t)]\ g[y_i^l(t)] - 1\right] \tag{6}$$

$$\frac{dw_{ij}^l(t)}{dt} = -\mu(t) f[y_i^l(t)]\ g[y_j^l(t)], \text{for } i \neq j \tag{7}$$

where $f(y)$ and $g(y)$ are different, odd functions (e.g. $f(y) = y$, $g(y) = \tanh(y)$ or $f(y) = y^3$ and $g(y) = y$) and $\mu(t)$ is a positive learning rate.

The learning algorithm can be written compactly in the matrix form as

$$\frac{d\boldsymbol{W}^l(t)}{dt} = \mu(t)\left[\boldsymbol{I} - \boldsymbol{f}[\boldsymbol{y}^l(t)]\boldsymbol{g}^T[\boldsymbol{y}^l(t)]\right] \tag{8}$$

The learning algorithm (8) can be easily transformed to (or approximated by) an iterative algorithm with discrete time $t = 0, 1, 2, ...$:

$$\boldsymbol{W}^l(t+1) = \boldsymbol{W}^l(t) + \eta(t)\{\boldsymbol{I} - \boldsymbol{f}[\boldsymbol{y}^l(t)]\ \boldsymbol{g}^T[\boldsymbol{y}^l(t)]\} \tag{9}$$

where $\eta(t) > 0$ is the learning rate (typically it is exponentially decreasing to zero during the learning process).

It is interesting to note that the same learning algorithm (8, 9) can be applied for multi-layer recurrent neural network as shown in Fig. 1(b).

The nonlinearities of activation functions $f(y)$ and $g(y)$ play essential role in the learning process. They are able to pick up higher–order moments of the input distribution and ensure mutual independence of output signals [3], [5], [6].

IV. Modifications And Extensions of Basic Learning

There are possible many extensions and modifications of the basic local learning algorithm (8), (9). First of all, in each layer different activation functions can be applied. Secondly the learning algorithm itself can be generalized as follows (the superscript l has been omitted for clarity reasons):

$$\frac{d\boldsymbol{W}}{dt} = \mu(t)\boldsymbol{G}(\boldsymbol{y}(t) \tag{10}$$

$$\text{or} \quad \boldsymbol{W}(t+1) = \boldsymbol{W}(t) + \mu(t)\boldsymbol{G}[\boldsymbol{y}(t)] \tag{11}$$

where matrix $\boldsymbol{G}(\boldsymbol{y}(t))$ can take one of the following form:

$$\begin{aligned}
\boldsymbol{G}_1[\boldsymbol{y}(t)] &= \boldsymbol{I} - \boldsymbol{f}[\boldsymbol{y}(t)]\ \boldsymbol{g}^T[\boldsymbol{y}(t)] \\
\boldsymbol{G}_2[\boldsymbol{y}(t)] &= \boldsymbol{I} - \boldsymbol{y}(t)\boldsymbol{y}^T(t) - \boldsymbol{f}[\boldsymbol{y}(t)]\boldsymbol{g}^T[\boldsymbol{y}(t)] \\
&\quad + \boldsymbol{g}[\boldsymbol{y}(t)]\boldsymbol{f}^T[\boldsymbol{y}(t)] \\
\boldsymbol{G}_3[\boldsymbol{y}(t)] &= \boldsymbol{I} - \boldsymbol{f}[\boldsymbol{y}(t)]\boldsymbol{g}^T[\boldsymbol{y}(t)] - \boldsymbol{y}(t)\boldsymbol{y}^T(t-T) \\
\boldsymbol{G}_4[\boldsymbol{y}(t)] &= \boldsymbol{I} - \sum_{k=0}^{m}[\boldsymbol{y}(t)\boldsymbol{y}^T(t-kT)]
\end{aligned}$$

and T is a suitable chosen delay time.

The choice of matrix $\boldsymbol{G}_i(\boldsymbol{y}), (i = 1, 2, 3, 4)$ depends on many factors, e.g. distribution of source signals, required convergence speed, desired accuracy (maximal allowed cross–talking effect). We found by computer experiments that the choice of matrix $\boldsymbol{G}(\boldsymbol{y})$ and nonlinearities of activation functions $f(y)$ and $g(y)$ are

essential condition for achieving high performance of separation (i.e. without a cross–talking effect). It should be noted also that the learning algorithm in (10) achieves convergence if the expectation value of every element from the matrix $\boldsymbol{G}(\boldsymbol{y})$ tends to zero with $t \to \infty$.

V. Experimental Results

Extensive computer experiments have shown that the performance of a multi–layer network, with the same local learning rule (8), (9) for each layer, is much higher than the performance of a pure single layer network. A considerable improvement for ill–conditioned problems (i.e. the problems in which the mixing matrices are near singular and/or if some source signals are very weak in comparison to others) can be observed.

We have simulated the proposed algorithms using both synthetic and real–world 2–D source signals like images. Due to limit of space two simulation results are presented only.

A. Well conditioned separation problem

Fig.2 shows the results of separating three images for a well–conditioned mixture matrix. The following mixing matrix has been used: $\{[3.0, 2.0, 1.0]; [0.22, 0.2, 0.2]; [0.5, 0.2, 0.05];\}$

Now only mixed images were available for the learning algorithm and there was no information about the mixing matrix neither the source images. The learning algorithm (8, 9) was able to separate images using one layer only, as shown in the last row in Fig. 2.

B. Ill–conditioned separation problem

In the second experiment the four digital images (primary sources), shown in the top row of Fig. 3, were mixed either by randomly chosen or by very ill–conditioned mixing matrix $\boldsymbol{A}$. The effect of mixture shown in the second row of Fig. 3 was achieved by applying the following mixing matrix: $\{[0.001, 3.0, 4.0, 0.02]; [0.0011, 3.1, 4.0, 0.024]; [0.0012, 3.0, 3.9, 0.026]; [0.0015, 3.2, 4.1, 0.019]\}$

Again only mixed images are available for the learning algorithm. In this example a three–layer neural network with $\boldsymbol{G}(\boldsymbol{y}) = \boldsymbol{G}_2(\boldsymbol{y})$ and $f(y) = y, g(y) = \tanh(10y)$ has been applied. Nevertheless how long the single layer is learning it is not able to separate the four images. But adding the second and third layer leads to a dramatic improvement of the separation. The separated images after the first, second and third layer are shown in consecutive rows in the Fig. 4. It is easy to note that the results are very good after the third layer.

VI. Conclusion

Biologically plausible and efficient local on–line learning algorithms for neural network based solution of the blind separation problem have been described. Their main advantage over a single layer learning is the handling of ill–conditioned signal mixtures. This can have direct impact for image sequence enhancement, where instead of different sensors a single camera may be applied. The validity and high performance of the proposed neural network have been illustrated by computer simulations. The proposed algorithm can be extended for complex–valued signals and convolution based mixing of independent signals.

References

[1] S. Amari, A. Cichocki, and H.H. Yang, *A new learning algorithm for blind signal separation*, **Proceedings of NIPS'95**, (Denver, USA, Nov.1995).

[2] S. Amari, A. Cichocki, and H.H. Yang, *Recurrent neural networks for blind separation of sources*, **Proceedings of NOLTA-95**, (Las Vegas, USA, Dec.1995), (this volume).

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

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

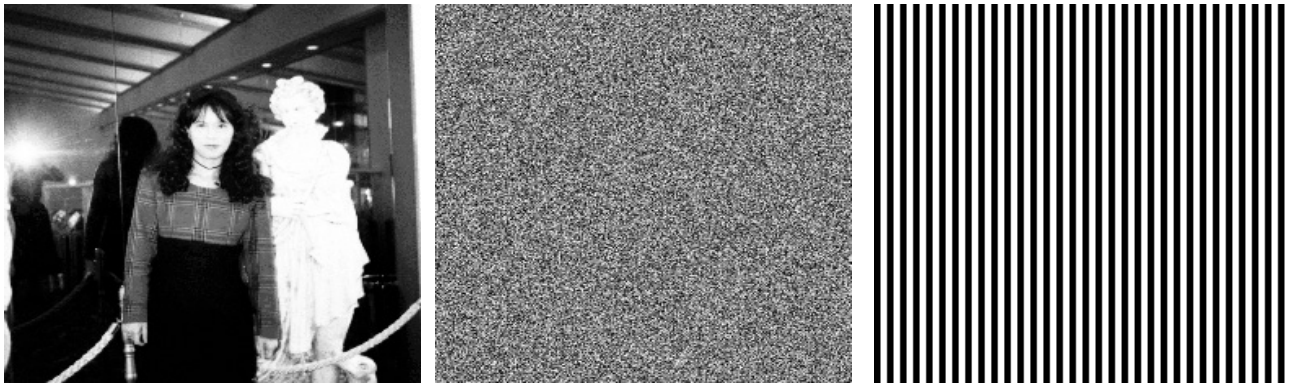
[5] A. Cichocki and R. Unbehauen, **Neural Network for Optimization and Signal Processing**, Teubner–Wiley, 1994.

[6] C. Jutten and J. Herault, *Blind separation of sources, Part I: An adaptive algorithm based on neuromimetic architecture*, **Signal Processing**, vol. 24, 1991, 1–20.

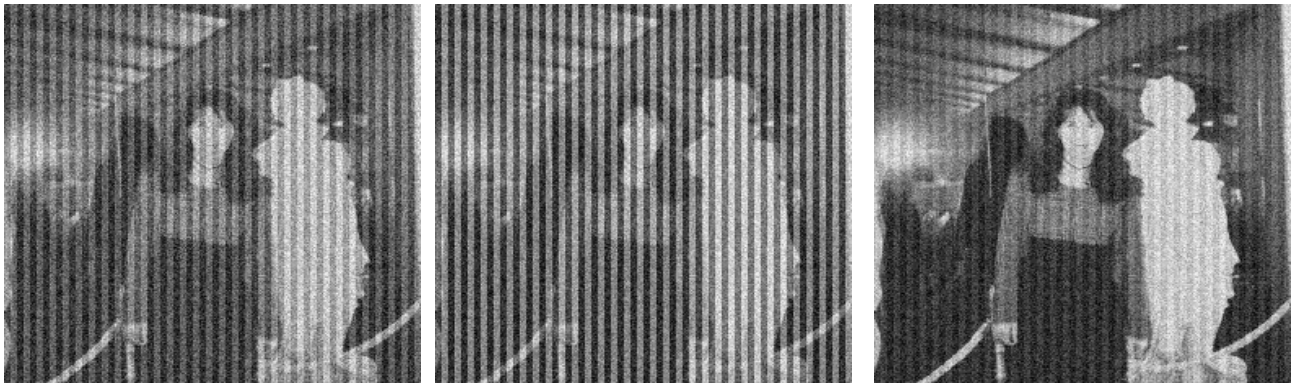
[7] J. Karhunen and J. Joutsensalo, *Representation and separation of signals using nonlinear PCA type learning*, **Neural Networks**, vol. 7, No.1, 1994, 113–127.

[8] E. Oja and J. Karhunen, *Signal separation by nonlinear hebbian learning*, **Proceedings ICNN-95**, (Perth, Australia, Dec.1995).

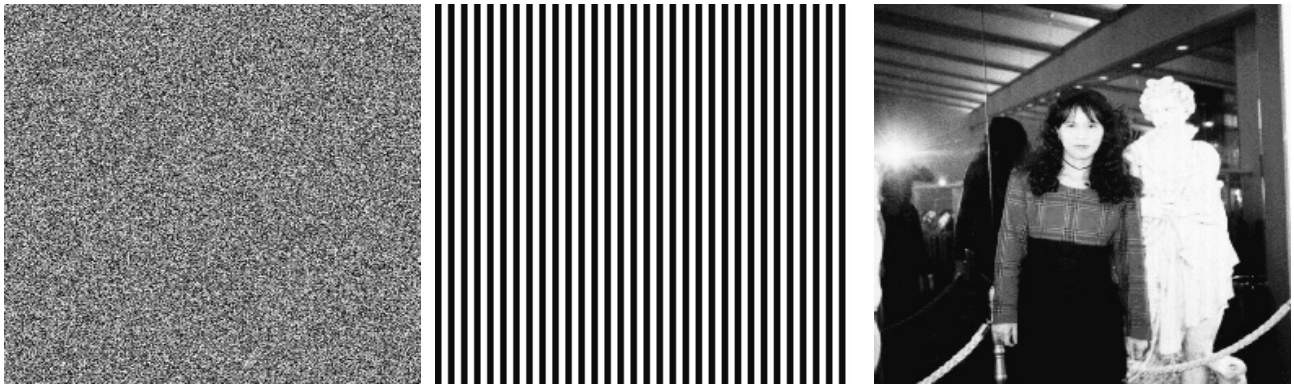
[9] E. Moreau and O. Macchi, *A complex self–adaptive algorithms for source separation based on high order contrasts*, **Proceedings. EUSIPCO-94**, 1157–1160.



Three original images

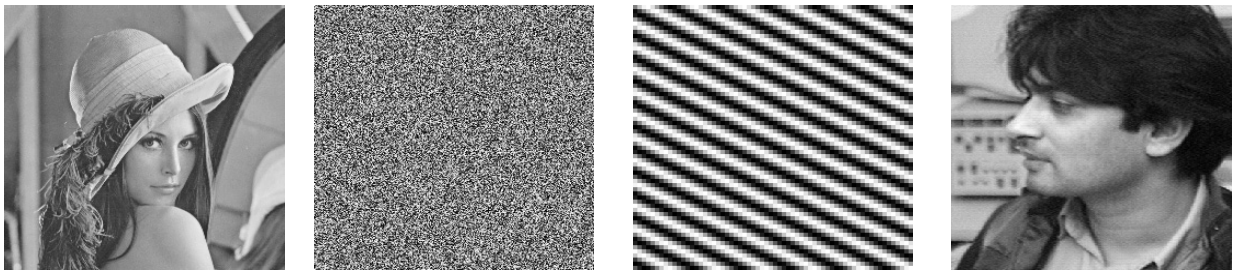


Three mixed images

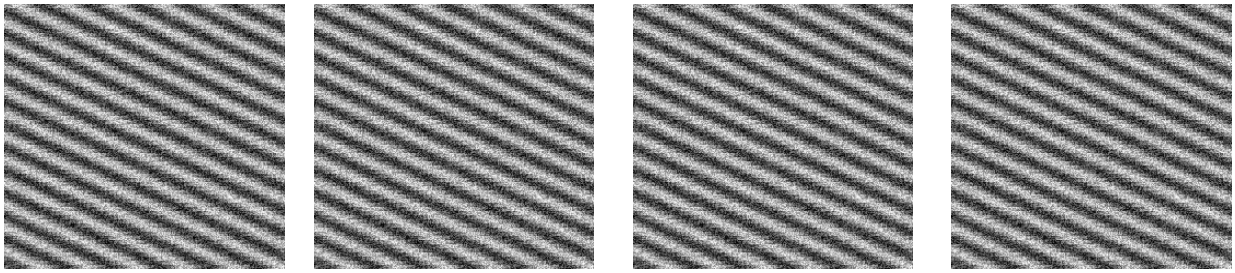


Separated images already after the first layer

Figure 2: Results of blind separation of three images, mixed by a well–conditioned matrix. Separation was possible already after the first layer.

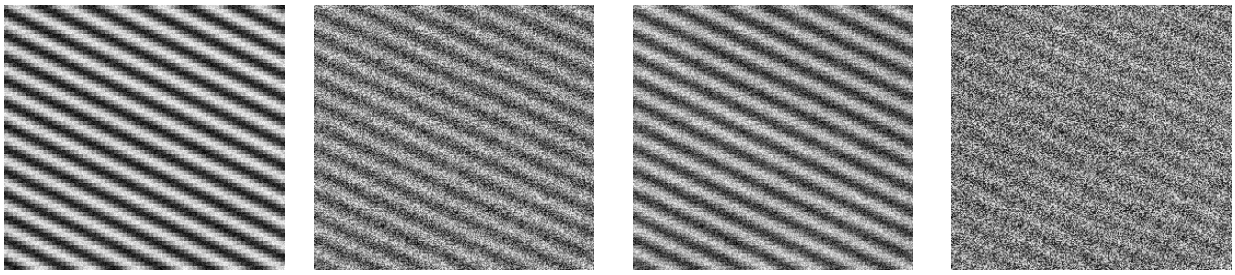


Four original images



Four mixed images

Figure 3: The mixed images are obtained from source images by applying an ill–conditioned mixing matrix. The images *Lenna* and *Ali* are app. 3000 and 150 times weaker than *noise* and *texture*.



Separated images after the first layer

Separated images after the second layer

Separated images after the third layer

Figure 4: The results of blind separation after the first, second and third layer.