G. Ramponi et al. (eds.): *Signal Processing VIII: Theories and Applications.*
EURASIP/LINT Publ., Trieste, Italy, 1996, vol. I, pp. 579–582.

NEURAL NETWORK APPROACH TO BLIND SEPARATION AND ENHANCEMENT OF IMAGES

Andrzej CICHOCKI[1] *Włodzimierz KASPRZAK* *Shun–ichi AMARI*
RIKEN, Frontier Research Program, BIP Group
2–1 Hirosawa, Wako–shi, Saitama 351–01, JAPAN
Phone: +81 48 465 2645; Fax: +81 48 462 4633
e-mail: cia@kamo.riken.go.jp

ABSTRACT

In this contribution we propose a new solution for the problem of *blind separation of sources* (for one dimensional signals and images) in the case that not only the waveform of sources is unknown, but also their number. For this purpose multi–layer neural networks with associated adaptive learning algorithms are developed. The primary source signals can have any non–Gaussian distribution, i.e. they can be sub–Gaussian and/or super–Gaussian. Computer experiments are presented which demonstrate the validity and high performance of the proposed approach.

1 INTRODUCTION

Blind signal processing and especially blind separation of sources is a new emerging field of research with many potential applications in science and technology [1] - [8]. Let us assume a number of independent sources are linearly mixed by unknown mixing coefficients according to the matrix equation $\mathbf{x(t)} = \boldsymbol{As}(\mathbf{t})$, where $\boldsymbol{s}(t) = [s_1(t), ..., s_m(t)]^T$ is a vector of *unknown* zero–mean primary sources, $\boldsymbol{A} \in R^{n \times m}$ is an unknown $n \times m$ matrix ($m \leq n$) of full rank 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) and their number only on basis of the observation of sensor signals $\boldsymbol{x}(t)$. Most of the developed methods provides a solution when the number of sources is known a priori or fail successfully to separate signals when they are badly scaled or the problem is ill–posed (i.e. a mixing matrix $\boldsymbol{A}$ is for example near singular). In this paper we propose a solution which avoids these drawbacks.

2 PROPOSED SOLUTION

In order to solve the problem we apply a neural network approach. Generally a neural network consists of two or more linear layers (Fig. 1). The first sub–network (single or multi–layer) performs simultaneously *pre–whitening* and *separation* and the last (post–processing) layer eliminates redundant signals, i.e. it determines the number of active sources in the case if the number of sensor (mixed) signals n is greater than the number of the primary sources m.

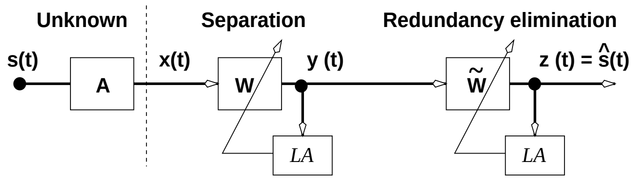


Figure 1: The scheme of proposed approach.

2.1 Separation of sources

The first layer performs the linear transformation $\boldsymbol{y}(t) = \boldsymbol{W}(t)\boldsymbol{x}(t)$, where $\boldsymbol{W}(t)$ is a $n \times n$ nonsingular matrix of synaptic weights updated according to an on–line learning rule [1-3]. Generally speaking, minimization of the cost function $J(\boldsymbol{y}) = ||\boldsymbol{I} - E\{\boldsymbol{f}(\boldsymbol{y})\boldsymbol{g}(\boldsymbol{y}^T)\}||_F$ leads us to following two learning rules:

- local (simplified) rule

$$\Delta \boldsymbol{W}(t) = \eta(t) \left\{ \boldsymbol{I} - \boldsymbol{f}[\boldsymbol{y}(t)]\boldsymbol{g}[\boldsymbol{y}^T(t)] \right\}, \qquad (1)$$

- global, robust rule

$$\Delta \boldsymbol{W}(t) = \eta(t) \left\{ \boldsymbol{I} - \boldsymbol{f}[\boldsymbol{y}(t)]\boldsymbol{g}[\boldsymbol{y}^T(t)] \right\} \boldsymbol{W}(t), \qquad (2)$$

where $\eta(t) > 0$ is the adaptive learning rate, $\boldsymbol{I}$ – the the $n \times n$ identity matrix, $\boldsymbol{f}(\boldsymbol{y}) = [f(y_1), ..., f(y_n)]^T$ and $\boldsymbol{g}(\boldsymbol{y}^T) = [g(y_1), ..., g(y_n)]$ are nonlinear activation functions, E is expectation and $||.||_F$ the Frobenius norm.

The choice of activation function depends on the *kurtosis* of source signals. If the source signals are expected to have negative kurtosis value, i.e. they are *sub–Gaussian* signals, we can choose for example $f(y_j) = y_j^3$ and $g(y_j) = \tanh(\gamma y_j)$. On the other hand, for *super–Gaussian* signals with positive kurtosis, in order to obtain successful separation, we can choose for example $f(y_j) = \tanh(\gamma y_j), g(y_j) = y_j^3$.

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

Image	m_1	μ_2	μ_3	$kurt_{\vert\vert}$
Girl (G)	0.4932	0.0723	-0.00159	-0.909
Sinus (S)	0.5019	0.1253	0.00001	-1.505
Noise (N)	0.5029	0.0424	-0.00005	-0.619
Miss (M)	0.4596	0.0667	-0.00250	-1.472

Table 1: Statistical characteristics of normalized images used in experiments m_1 – mean value, μ_2 – variance, μ_3 – skewness, $kurt_{||}(= \mu_4/\mu_2^2 - 3)$ – normalized kurtosis.

If sources may have both positive and negative kurtosis the local learning rule (1) can be generalized as:

$$\Delta w_{ij}(t) = \eta(t)\left\{\delta_{ij} - f_i[y_i(t)]g_j[y_j(t)]\right\}, \quad (3)$$

where δ_{ij} is Kronecker delta.

In this case the i–th neuron will extract a source signal with negative kurtosis if we select $f_i(y_i) = y_i^3$ and $g_j(y_j) = \tanh(\gamma y_j)$, and a source with positive kurtosis if we choose $f_i(y_i) = \tanh(\gamma y_i)$ and $g_j(y_j) = y_j^3$. As the sources are unknown the sign of kurtosis could be estimated in adaptive way in an on–line manner during a learning process as follows (for $p = 2, 4$):

$$\begin{aligned} \hat{\mu}_p[y_i(t)] &= [1 - \eta(t)]\hat{\mu}_p[y_i(t-1)] + \eta(t) y_i^p(t), \\ \widehat{kurt}[y_i(t)] &= \hat{\mu}_4[y_i(t)] - 3\hat{\mu}_2^2[y_i(t)], \end{aligned} \quad (4)$$

where $\eta > 0$ is a small constant.

Learning algorithms applying rules (1–3) perform simultaneously generalized *pre–whitening* (called also *sphering* or normalized orthogonalization) and *independent component analysis*, i.e. the separation of sources. The robust algorithm (2) possesses the *uniformity* or *equivariance* property [1, 4], but the simplified algorithm (1) does not have such property. Therefore, in order to improve the performance of separation, we could use a multilayer neural network [3].

2.2 Determination of source number

The output signals $y_j(t)$ contain some redundancy in the case when number of sources is smaller than number of sensors, i.e. some signals are repeated with different scale factors. In order to eliminate this redundancy we have applied one post–processing layer. Such a 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:

$$\begin{aligned} \widetilde{w}_{ii}(t) &= 1, \quad \forall t\ \forall i \\ \Delta\widetilde{w}_{ij}(t) &= -\eta(t) f[z_i(t)]g[z_j(t)], \quad i \neq j \end{aligned} \quad (5)$$

where $g(z)$ is a nonlinear odd activation function (e.g. $g(z) = \tanh(\gamma z)$) and $f(z)$ is either a linear or slightly non–linear odd function.

Alternative solution is to apply a lower or upper triangular matrix $\tilde{\boldsymbol{W}}(t)$. The learning rule takes then the following form:

$$\begin{aligned} \widetilde{w}_{ii}(t) &= 1, \quad \forall t\ \forall i \\ \Delta\widetilde{w}_{ij}(t) &= -\eta(t) f[z_i(t)]g[z_j(t)], \quad i < j \text{ (or } i > j) \\ \widetilde{w}_{ij}(t) &= 0, \quad \forall t, \quad \text{for } i > j \quad \text{(or resp. } i < j) \end{aligned} \quad (6)$$

Constraints imposed on matrix $\tilde{\boldsymbol{W}}(t)$ ensure mutual decorrelation of output signals, what eliminates the information redundancy of the output signals and may also improve the performance of separated signals. However, we found this performance strongly to depend on statistical distribution of sources. If the source signals does not satisfy fully the independency requirement, e.g. mixing of two or more natural images, the post–processing layer should be used rather for redundancy elimination only.

The above learning rules can be derived using the same optimization criterion as was used for the separation of sources (eqs. 1, 2), but with some constraints for synaptic weights of the matrix $\tilde{\boldsymbol{W}}(t)$, e.g. $\tilde{w}_{ii}(t) = 1, \forall i$. It should be noted that the rule (5) is similar to the well–known Herault–Jutten rule [6], but applied to a feed–forward network and with different activation functions.

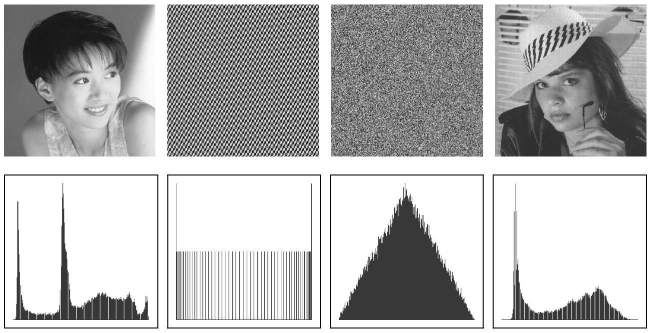


Figure 2: Four original images (assumed to be completely unknown) (top row) and their histograms (bottom row).

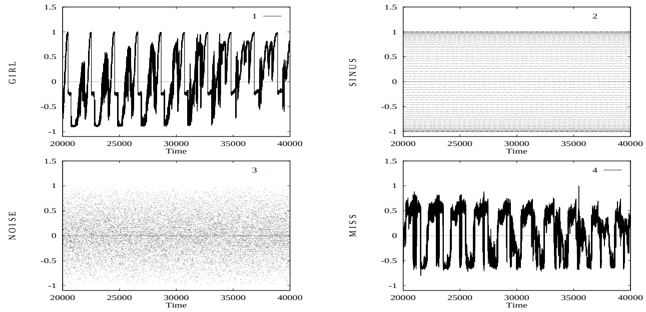
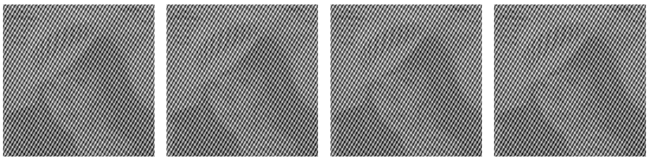


Figure 3: Image sources after normalization and transformation to zero–mean 1–D signals.

3 COMPUTER SIMULATIONS RESULTS

Performance and validity of the proposed approach has been confirmed by extensive computer simulation experiments. Due to lack of space we present here only two



Mixed (sensor) images of four original sources
($n = m = 4$)

Separated images after three layers with local rule (1)

Separated images after one layer with global rule (2)

The redundancy elimination layer is not suppressing any separated signal in complete source case

Figure 4: Blind separation in the case when the number of sources is equal to the number of sensors (mixed signals). Notice that the mixed signals look almost identical because the mixing matrix is ill–conditioned.

examples for image restoration and enhancement. Four original images – *Girl, Sinus, Noise, Miss*, shown in Fig. 2 – were mixed by ill–conditioned matrices $\boldsymbol{A}_1 \in R^{4\times4}$ or $\boldsymbol{A}_2 \in R^{3\times5}$ ($s_3(t) = 0$) which are assumed to be completely unknown to the separation network. The results of processing the first source mixture set ($n = m = 4$, $\boldsymbol{A}_1$) are given in Fig. 4, whereas the results for the over–determined case (five mixtures of three sources, $\boldsymbol{A}_2$) are provided in Fig. 5. For the first experiment quantitative results are given in Tab. 3. The table contains performance factors for signal separation if a multi–layer with local rule learning (1) or a single layer with global rule learning (2) method were applied during the first step of initial signal separation. For the second experiment, quantitative results are given in Tab. 4. Here the two optional learning rules for separation have also been tested.

In order to estimate the separation quality we assume to know the original sources and the mixing matrix. Then the quality of obtained results is provided in two ways:

Signal	Absolute values of signal correlations (×100)					
Sources						
	G-S	G-M	G-N	S-M	S-N	M-N
s	0.004	**30.46**	0.261	0.041	0.178	0.304
Sensor signals						
	1-2	1-3	1-4	2-3	2-4	2-3
x	99.99	100.0	100.0	99.99	99.99	100.0
After local rule separation						
$y^{(1)}$	-	-	-	3.301	69.52	16.75
$y^{(2)}$	0.691	**1.292**	0.388	0.934	0.935	0.225
$y^{(3)}$	0.001	**0.569**	0.021	0.004	0.013	0.011
z	0.034	**16.26**	0.232	0.006	0.135	0.259
After global rule separation						
y	0.001	**15.34**	0.416	0.069	0.334	0.566
z	0.006	**8.817**	0.227	0.057	0.111	0.282

Table 2: Absolute values of source correlations, i.e. $|E\{s_i s_j\}|$, and signal correlations after mixing, separation and redundancy elimination. The correlation factors between natural images are highlighted.

Signal	*EI*	PSNR [dB]			
		Girl	*Sinus*	*Miss*	*Noise*
Multi–layer separation with local rule					
$\boldsymbol{y}^{(1)}$	1.970	-	12.62	31.60	-
$\boldsymbol{y}^{(2)}$	0.049	24.80	36.97	18.63	43.27
$\boldsymbol{y}^{(3)}$	0.047	24.58	48.33	18.96	47.86
Separation with global rule					
$\boldsymbol{y}$	0.021	37.61	61.18	21.56	44.39

Table 3: Error index and quality factors for multi–layer with local rule separation or single–layer with global rule separation.

1. individually for each source, by estimating a PSNR (peak signal to noise ratio) between the reconstructed source and the corresponding original source;

2. for the whole separated signal set, by calculation of a normalized error index EI, which is defined as:

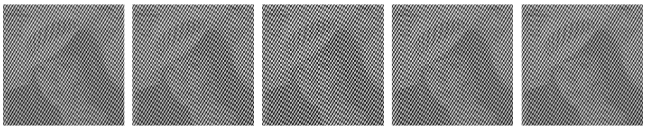
$$EI = \sum_{j=1}^{m} \left(\sum_{i=1}^{n} \frac{|p_{ij}|^2}{max_i |p_{ij}|^2} - 1 \right), \qquad (7)$$

The p_{ij}-s are entries of a normalized matrix $\boldsymbol{P}(t) \in R^{n\times m}$ is defined as

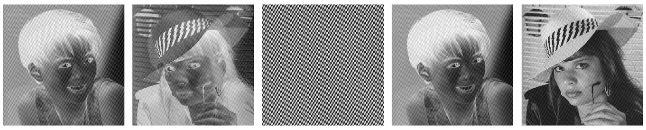
$$\boldsymbol{P} = \widetilde{\boldsymbol{W}} \boldsymbol{W}^{(k)} ... \boldsymbol{W}^{(1)} \boldsymbol{A}.$$

where every non–zero row of matrix $\boldsymbol{P}$ is normalized by $\sum_{j=1}^{m} |p_{ij}| = 1, \forall i$. This EI index is more general than the PI index defined in [8] as it is also valid for the overdetermined case.

Two main observations can be done. In the separation stage performed by the local learning layer, the



Five mixed images of three unknown sources



Separated images after four layers with local rule (1)

Separated images after single layer with global rule (2)

The redundancy elimination layer is suppressing redundant signals

Figure 5: Blind separation in the case when the number of sources is smaller than the number of mixed images. After separation there are two redundant images detected. After the redundancy elimination layer both these signals are suppressed.

signals are separated one after the other starting with the strongest signal first and finishing with the weakest signal contained in the mixture set. The global rule learning algorithm determines all sources at the same time using single layer sub–network. In case of redundant sensor number the separation quality is usually slightly worser than in complete source case. The final redundancy elimination layer suppresses redundant signals and is not switching between channel signals. However, if the separated images are already of high quality, the redundancy elimination layer may lead to slight deterioration of separation performance. Especially in the complete source case, it is often required to take as output the signals $y_i(t)$ from the separation layer, that correspond to non–suppressed channels.

4 CONCLUSIONS

Adaptive on–line learning algorithms have been developed for the problem of blind separation of unknown number of sources (only the maximum active source number is known in advance). In such a case a redundant number of sensors has usually to be used. The proposed solution was divided into two consecutive stages: the first stage performs separation as usual, whereas the second stage eliminates redundant signal separations.

Signal	EI	PSNR [dB]		
		Girl	*Sinus*	*Miss*
Multi–layer separation with local rule				
$\boldsymbol{y}^{(1)}$	2.028	-	18.89	20.30
$\boldsymbol{y}^{(2)}$	2.033	19.53	10.64	16.14
$\boldsymbol{y}^{(3)}$	2.025	15.83	21.39	14.53
$\boldsymbol{y}^{(4)}$	1.990	16.22	34.83	15.23
Redundancy elimination				
$\boldsymbol{z}$	0.080	18.04	43.66	19.63
Separation with global rule				
$\boldsymbol{y}$	1.977	45.09	43.32	21.97
Redundancy elimination				
$\boldsymbol{z}$	0.036	41.13	36.98	18.44

Table 4: Error index and performance factors for the over–determined case of either local rule or global rule separation with final redundancy elimination layer.

Presented computer experiments demonstrate the validity and high performance of proposed algorithms, especially for image sources.

References

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

[2] S. Amari, A. Cichocki, and H.H. Yang, *Recurrent Neural Networks for Blind Separation of Sources*, Proceedings of NOLTA'95, NTA Research Society of IEICE, Tokyo, Japan, 1995, 37–42.

[3] A. Cichocki, W. Kasprzak, and S. Amari, *Multi–Layer Neural Networks with a Local Adaptive Learning Rule for Blind Separation of Source Signals*, Proceedings of NOLTA'95, NTA Research Society of IEICE, Tokyo, Japan, 1995, 61–65.

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

[5] A. Cichocki, R. Unbehauen, L. Moszczynski, E. Rummert, *A new on-line adaptive learning algorithm for blind separation of sources*, Proceedings of ISANN-94, Taiwan, 1994, 406–411.

[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] E. Oja and J. Karhunen, *Signal separation by nonlinear Hebbian learning*, In: Computational Intelligence, IEEE Press, 1995, 83–96.

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