# NEURAL BLIND SEPARATION OF COMPLEX SOURCES BY EXTENDED APEX ALGORITHM (EAPEX)

*Simone Fiori, Aurelio Uncini, and Francesco Piazza*

Dept. Electronics and Automatics – University of Ancona (Italy)
E-mail:simone@eealab.unian.it

**ABSTRACT**

Blind Source Separation by non-classical (non-quadratic) neural Principal Component Analysis has been investigated by several papers over the recent years, even if particular attention has been paid to the real-valued sources case. The aim of this work is to present an extension of the Kung-Diamantaras' APEX learning rule to non-quadratic complex optimization, and to show the new approach allows blind separation of complex-valued source signals from their linear mixtures.

## 1. INTRODUCTION

Blind Source Separation (BSS) by Independent Component Analysis (ICA) of complex-valued data [2, 4, 10] is a meaningful problem that has been investigated in a very few papers, while much more attention has been paid to develop several different algorithms for performing BSS-ICA of real-valued data. Among others, those methods based on non-linear extensions of Principal Component Analysis (PCA) have raised much interest in the Neural Network community (see for example [6, 9] and references therein). It has been proven by many papers that adding non-linearity to linear PCA neural networks makes them able to improve the independence of their outputs so as to allow blind separation of independent sources [6, 9]. Recently, some attempts have been made in order to extend the best known PCA algorithms to the complex case. In [1] Chen and Hou presented an heuristic complex version of the well-known APEX algorithm [5], in [3] De Castro *et al.* gave an heuristic complex extension of the GHA algorithm [11] successively generalized to the non-linear case by Fiori and Piazza [8], while in [7] Fiori and Uncini proposed a formal derivation of a large class of complex PCA neural algorithms containing, as a special case, the one found in [1].

In this paper we formally derive a new learning algorithm as a non-linear complex generalization of APEX, and discuss the choice of the non-linearity under the theoretical framework proposed by Sudjianto and Hassoun [12] extended to the complex case. Then we show how a particular non-linearity, called Rayleigh function, allows the neural network to separate out mixed independent complex-valued source signals.

This research was supported by the Italian MURST.

## 2. EXTENDED APEX ALGORITHM (EAPEX)

Consider the complex-weighted neural network with input $\mathbf{x} \in \mathcal{C}^p$ and output $\mathbf{y} \in \mathcal{C}^m$, with $m \leq p$, described by the following relationships:

$$z_k = \mathbf{w}_k^{\#}\mathbf{x} \; , \; y_k = z_k + \mathbf{h}_k^{\#}\mathbf{y} \; , \tag{1}$$

where $\mathbf{z} \in \mathcal{C}^m$, vectors $\mathbf{w}_k \in \mathcal{C}^p$ represent the network's direct connections, vectors $\mathbf{h}_k \in \mathcal{C}^p$ represent the network's lateral connections, and superscript $^{\#}$ denotes conjugate transpose. In the following $E_{\mathbf{x}}[\cdot]$ denote mathematical expectation with respect to $\mathbf{x}$.

In [7] it has been shown that it is possible to define a pair $(J, C)$ of objective functions as follows:

$$\begin{aligned} J(\mathbf{w}_k) &\stackrel{\text{def}}{=} \sum_{k=1}^{m} P_k(\mathbf{w}_k) + \frac{1}{2}\sum_{k=1}^{m}\lambda_k(1-\mathbf{w}_k^{\#}\mathbf{w}_k) \; , \\ C(\mathbf{h}_k) &\stackrel{\text{def}}{=} \frac{1}{2}\sum_{k=1}^{m} E_{\mathbf{x}}[|y_k|^2] + \sum_{k=1}^{m} E_{\mathbf{x}}[\psi_k](\mathbf{h}_k^{\#}\mathbf{h}_k) \; , \end{aligned}$$

where $P_k = E_{\mathbf{x}}[|y_k|^2]$ and $\lambda_k$ and $E_{\mathbf{x}}[\psi_k]$ are Lagrange multipliers, such that maximizing $J$ with respect to each $\mathbf{w}_k$ only and simultaneously minimizing $C$ with respect to each $\mathbf{h}_k$ only, gives a set of learning equations allowing the network to perform Principal Component Analysis of incoming complex-valued data.

In this paper we generalize the criterion $J(\mathbf{w}_k)$ by defining instead:

$$P_k \stackrel{\text{def}}{=} E_{\mathbf{x}}[f(|y_k|)] \; , \tag{2}$$

where $f(u) : \mathcal{R}_0^+ \to \mathcal{R}_0^+$ is a function continuously differentiable almost everywhere, non-decreasing with a unique minimum in $u = 0$. In this case we thus have:

$$\frac{\partial J(\mathbf{w}_k)}{\partial \mathbf{w}_k} = E_{\mathbf{x}}\left[f'(|y_k|)\frac{\partial |y_k|}{\partial \mathbf{w}_k}\right] \; , \tag{3}$$



0-7803-5474-5/99/$10.00(C)1999 IEEE

while from [7] we know that:

$$\frac{\partial |y_k|^2}{\partial \mathbf{w}_k} = 2\mathbf{x}y_k^\star = 2|y_k|\frac{\partial |y_k|}{\partial \mathbf{w}_k} , \qquad (4)$$

where superscript $^\star$ denotes conjugation. Moreover, from [7] we also know that:

$$\frac{\partial |y_k|^2}{\partial \mathbf{h}_k} = 2\mathbf{y}_{[k]}y_k^\star , \qquad (5)$$

where $\mathbf{y}_{[k]} \stackrel{\text{def}}{=} [y_1\ y_2\ \cdots\ y_{k-1}\ 0\ \cdots\ 0]^T$ with $k > 1$, and $\mathbf{y}_{[1]} \stackrel{\text{def}}{=} [0\ 0\ \cdots\ 0\ 0]^T$. Define now $F(u) \stackrel{\text{def}}{=} \frac{f'(u)}{u}$; then we may write:

$$\begin{aligned} \frac{\partial J(\mathbf{w}_k)}{\partial \mathbf{w}_k} &= E_{\mathbf{x}}[F(|y_k|)y_k^\star \mathbf{x}] - \lambda_k \mathbf{w}_k , \\ \frac{\partial C(\mathbf{h}_k)}{\partial \mathbf{h}_k} &= E_{\mathbf{x}}[y_k^\star \mathbf{y}_{[k]}] + E_{\mathbf{x}}[\psi_k]\mathbf{h}_k . \end{aligned}$$

From optimization theory we know the optimal multipliers $\lambda_k$ can be found by solving the following equations:

$$\mathbf{w}_k^\# \frac{\partial J}{\partial \mathbf{w}_k} = E_{\mathbf{x}}[F(|y_k|)]y_k^\star z_k] - \lambda_k = 0 ,$$

whose solutions give the optimal gradients:

$$\left(\frac{\partial J(\mathbf{w}_k)}{\partial \mathbf{w}_k}\right)^{\text{opt}} = E_{\mathbf{x}}[F(|y_k|)y_k^\star(\mathbf{x} - z_k\mathbf{w}_k)] . \qquad (6)$$

Standard Kuhn-Tucker theory shows that it is not possible to find optimal multipliers $E_{\mathbf{x}}[\psi_k]$, instead. Then, by letting weight vectors adapt by means of gradient steepest ascent/descent algorithms, we find the new learning rule:

$$\begin{aligned} \Delta \mathbf{w}_k &= \eta \frac{\partial J}{\partial \mathbf{w}_k} = \eta E_{\mathbf{x}}[F(|y_k|)y_k^\star(\mathbf{x} - z_k\mathbf{w}_k)] , \qquad (7) \\ \Delta \mathbf{h}_k &= -\eta \frac{\partial C}{\partial \mathbf{h}_k} = -\eta E_{\mathbf{x}}[y_k^\star \mathbf{y}_{[k]}] - \eta E_{\mathbf{x}}[\psi_k]\mathbf{h}_k , \qquad (8) \\ k &= 1, 2, \ldots, m , \ \eta > 0 , \end{aligned}$$

that is referred to as *Extended APEX Algorithm* (EAPEX) in that for $f(u) = \frac{1}{2}u^2$, $\psi_k = |y_k|^2$, as long as $y_k \cong z_k$, and in presence of real valued data, it resembles the APEX learning algorithm by Kung and Diamantaras [5].

In the next Section we discuss a choice of $F(\cdot)$ arising from an interpretation of non-quadratic Hebbian learning due to Sudjianto and Hassoun [12] extended to the complex case.

## 3. THE SUDJIANTO-HASSOUN INTERPRETATION OF NON-QUADRATIC HEBBIAN LEARNING

In [12], Sudjianto and Hassoun considered the problem of maximizing a criterion $P(\mathbf{w}) \stackrel{\text{def}}{=} E[S^2(\mathbf{w}^T\mathbf{x})]$ subject to the restriction $\mathbf{w}^T\mathbf{w} = 1$, where $y = \mathbf{w}^T\mathbf{x}$ is the output of a single-unit real-weighted neural network and $S(\cdot)$ is a generic saturating sigmoidal function, for instance such that $S(\cdot) \in [-1, +1]$, that is the same problem addressed in Section 2 particularized to $m = 1$. They noted that maximizing the variance of a saturating function of $y$ leads the neuron to prefer configurations $\mathbf{w}$ corresponding to values of $S(y)$ concentrated near the extremes $-1$ and $+1$. If the quantity $z = S(y)$ is perceived as a new random variable with probability density function $q_Z(z|\mathbf{w})$, this makes U–shaped the distribution $q_Z$ [12]. The gradient steepest ascent learning rule for the neuron is:

$$\frac{d\mathbf{w}}{dt} = \frac{\partial J}{\partial \mathbf{w}} = (\mathbf{I} - \mathbf{w}\mathbf{w}^T)E_{\mathbf{x}}[\ell(y)\mathbf{x}] , \qquad (9)$$

where $\ell(u) \stackrel{\text{def}}{=} 2S'(u)S(u)$. Denote now by $q_Y(y|\bar{\mathbf{w}})$ the probability density function of the random variable $y$ due to a configuration $\bar{\mathbf{w}}$, and with $Q_Y(y|\bar{\mathbf{w}})$ its cumulative distribution function, namely:

$$Q_Y(y|\bar{\mathbf{w}}) \stackrel{\text{def}}{=} \int_{-\infty}^{y} q_Y(\eta|\bar{\mathbf{w}})d\eta .$$

Assume then $S(y) = 2Q_Y(y|\bar{\mathbf{w}}) - 1$. In this case it is well known [12] that $z$ will be uniformly distributed within $[-1, +1]$. The central idea developed by Sudjianto and Hassoun is that the learning rule (9) will converge to a weight vector surely different from $\bar{\mathbf{w}}$, since the rule seeks a U–shaped distribution of $z$, that is, a distribution that deviates away from a uniform one. If, for instance, $\mathbf{x} \in \mathcal{R}^2$, $x_1$ has a pdf $q_1(x_1)$ and $x_2$ has a pdf $q_2(x_2)$ different from $q_1$, then choosing $S(y) = 2Q_1(y) - 1$ makes the rule (9) able to filter the signal $x_1$ allowing for $y = x_2$ at convergence, and vice-versa. In other words, the rule (9) behaves as a *probabilistic filter*.

Consider now the extension of the previous theory to the complex case. Define the cost function:

$$P(\mathbf{w}) \stackrel{\text{def}}{=} E_{\mathbf{x}}[S^2(|y|)] . \qquad (10)$$

In our case we assume:

$$S(|y|) = Q_{|Y|}(|y|) \stackrel{\text{def}}{=} \int_0^{|y|} q(\eta)d\eta ,$$

where $q(\cdot)$ represents a generic probability density function. Clearly this implies $f(u) = S^2(u)$, hence:

$$f'(u) = 2q(u)\int_0^u q(\eta)d\eta . \qquad (11)$$

Ultimately it is clear that training each neuron of a linear complex-weighted neural network by means of the learning rule EAPEX with the non-linearity (11) causes the network to learn connection strengths that filter the outputs so that



0-7803-5474-5/99/$10.00(C)1999 IEEE

the probability density function of the output moduli $|y_k|$ deviates away from $q(\cdot)$. In the next section it will be made clear how this principle could be employed for separating out independent complex signals from their linear mixtures.

## 4. APPLICATION TO BLIND COMPLEX SOURCE SEPARATION

Suppose input $\mathbf{x}$ contains a complex linear mixture of statistically independent signals [10], and that one of these signals is a Gaussian noise of the form $v = r + is$, where both $r$ and $s$ are zero-mean Gaussian random variables of variance $\sigma^2$. Then it is known that the modulus $|v|$ follows the Rayleigh distribution:

$$q_{\mathrm{R}}(|v|) = \frac{|v|}{\sigma^2} \exp\left(-\frac{|v|^2}{2\sigma^2}\right) .$$

Then by formula (11) we find:

$$f'_{\mathrm{R}}(u) = \frac{2u}{\sigma^2}\left[\exp\left(-\frac{u^2}{2\sigma^2}\right) - \exp\left(-\frac{u^2}{\sigma^2}\right)\right]\Gamma(u) ,$$

where $\Gamma(u)$ is the unitary step. Figure 1 depicts the Rayleigh non-linearity $F_{\mathrm{R}}(u) = f'_{\mathrm{R}}(u)/u$ for a unitary noise power. In this case it is possible to express the cumulative distribution function in closed form simply as:

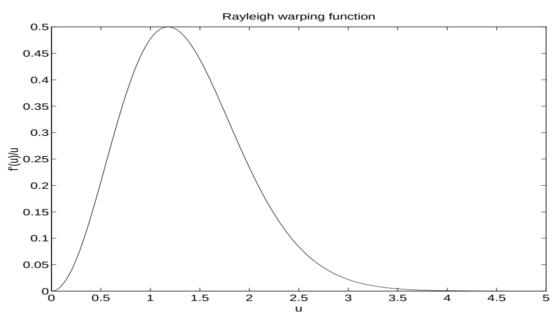


Figure 1: Rayleigh warping function for $\sigma^2 = 1$.

$$Q_{\mathrm{R}}(u) = 1 - \exp\left(-\frac{u^2}{2\sigma^2}\right) .$$

By using in the EAPEX algorithm the function $F_{\mathrm{R}}(u)$ it is then possible to separate out independent complex-valued signals mixed by a unitary operator. The general problem where generic linear mixtures are concerned can be solved by pre-whitening the data [2, 10].

## 5. EXPERIMENTAL RESULTS

As a numerical example, suppose input $\mathbf{x} \in \mathcal{C}^4$ is formed by a linear mixture of four independent signals arranged in a vector $\mathbf{s} \in \mathcal{C}^4$. Signal $s_1$ is QAM4 and $s_2$ is QAM16, both with small Gaussian phase deviation; signal $s_3$ is PSK, and $s_4$ is a Gaussian noise of variance $\sigma^2 = 0.5$. The mixture is computed as $\mathbf{x} = \mathbf{M}\mathbf{s}$, where $\mathbf{M}$ is a randomly generated $4 \times 4$ complex matrix. The first row of Figure 2 depicts the independent signals while second row shows the obtained four mixtures.

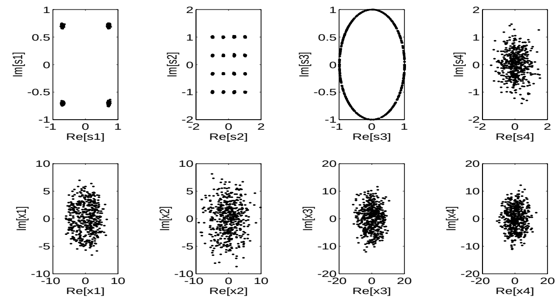


Figure 2: The four independent signals and the four mixtures of them.

By means of the Sudjianto-Hassoun principle, a linear neural network with four inputs and four outputs, trained by the learning rule EAPEX with the Rayleigh non-linearity, should be able to recover the independent signals except for a phase shift and a random permutation [2] after mixture prewhitening. Simulation results are shown in Figure 3: The first row depicts whitened mixtures obtained by means of the well-known Laheld-Cardoso's standardizing algorithm [10] in its stabilized version:

$$\Delta\mathbf{U} = \gamma\frac{\mathbf{U}(\mathbf{I} - \mathbf{v}\mathbf{v}^{\#})}{|1 + \gamma\mathbf{v}^{\#}\mathbf{v}|} \ , \ \mathbf{v} = \mathbf{U}^{\#}\mathbf{y} \ ,$$

with $\gamma = 0.0005$. The second row shows the last 100 outputs of the network trained by EAPEX on the prewhitened data $\mathbf{v}$, with the choice $\psi_k \stackrel{\mathrm{def}}{=} |y_k|$ (for a discussion on the possible choices of the $\psi_k$'s see [7]), $\eta = 0.005$, $\sigma^2 = 0.5$ and $\mathbf{W}(0) = \mathbf{I}$. Figure 4 shows the histograms of the

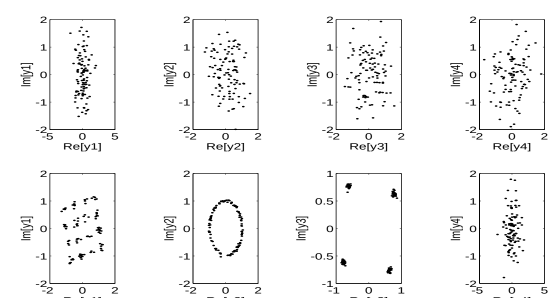


Figure 3: Network's output after learning by rule EAPEX.



0-7803-5474-5/99/$10.00(C)1999 IEEE

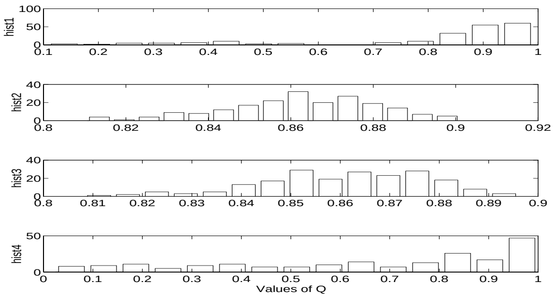


Figure 4: Functions $Q_{\mathrm{R}}(\cdot)$ histograms.

last 200 samples of $Q_{\mathrm{R}}(|y_1|), \ldots, Q_{\mathrm{R}}(|y_4|)$, while Figure 5 depicts the residual interference defined on the basis of the separation product $\mathbf{\Pi} \stackrel{\mathrm{def}}{=} \mathbf{W}^{\#}\mathbf{U}^{\#}\mathbf{M}$ as the sum of the 12 smallest numbers among $|\Pi_{ij}|^2$ divided by the sum of the 4 largest numbers among $|\Pi_{ij}|^2$.

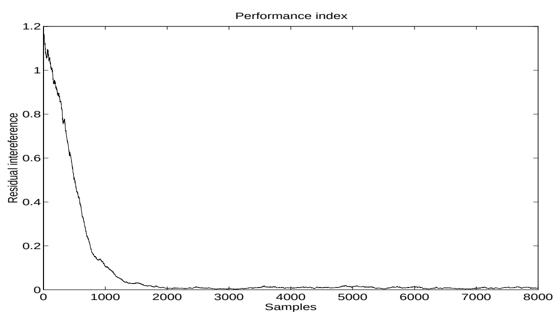


Figure 5: Residual intereference.

Simulation results show that the network is able to recover the independent signals. The histograms $Q_{\mathrm{R}}(|y_1|)$, $Q_{\mathrm{R}}(|y_2|)$ and $Q_{\mathrm{R}}(|y_3|)$ are in good accordance with the signals in Figure 3, while the presence of a peak in $+1$ on the histogram of $Q_{\mathrm{R}}(|y_4|)$ confirms that the fourth neuron cannot separate out the Gaussian noise and its output contains a mixture of the other source signals, as expected.

## 6. CONCLUSION

Blind separation by extended Hebbian algorithm has become a fruitful research field, as already pointed out also by Hyvärinen and Oja [6] who have recently proposed an algorithm for blind separation of real-valued signals from orthogonal mixtures via linear networks based on non-classical Hebbian learning. In this paper a new adapting rule for linear neural networks as generalization of APEX learning has been presented by using some results of an our previous work [7]. It provides a generalization in that it applies to complex-weighted neural networks and embeds non-linearity in the classical Hebbian learning. A particular choice of the non-linearity is discussed by recalling the Sudjianto-Hassoun interpretation of non-classical Hebbian learning extended to the complex case.

## 7. REFERENCES


[1] Y. Chen and C. Hou, *High resolution adaptive bearing estimation using a complex-weighted Neural Network*, Proc. of IEEE-ICASSP, Vol. II, pp. 317 – 320, 1992

[2] P. Comon, *Independent Component Analysis, a new concept ?*, Signal Processing, Vol. 36, pp. 287 – 314, 1994

[3] M.C.F. De Castro, F.C.C. De Castro, J.N. Amaral, and P.R.G. Franco, *A Complex Valued Hebbian Learning Algorithm*, Proc. of IEEE-ICNN, pp. 1235 – 1238, 1998

[4] G. Desodt and D. Muller, *Complex ICA applied to the separation of radar signals*, Proc. of EUSIPCO, Vol. I, pp. 665 – 668, 1990

[5] K.I. Diamantaras and S.-Y. Kung, *Principal Component Neural Networks: Theory and Applications*, J. Wiley & Sons, 1996

[6] A. Hyvärinen and E. Oja, *Independent Component Analysis by general non-linear Hebbian-like rules*, Signal Processing, Vol. 64, No. 3, pp. 301 – 313, 1998

[7] S. Fiori and A. Uncini, *A unified approach to laterally-connected neural networks*, Proc. EUSIPCO, Vol. I, pp. 379 – 382, 1998

[8] S. Fiori and F. Piazza, *Neural blind separation of complex sources by extended Hebbian learning (EGHA)*, These proceedings

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

[10] B. Laheld and J.F. Cardoso, *Adaptive source separation with uniform performance*, Proc. of EUSIPCO, Vol. 1, pp. 183 – 186, 1994

[11] T.D. Sanger, *Optimal unsupervised learning in a single-layer linear feedforward neural network*, Neural Networks, Vol. 2, pp. 459 – 473, 1989

[12] A. Sudjianto and M.H. Hassoun, *Nonlinear Hebbian rule: A statistical interpretation*, Proc. of IEEE-ICNN, Vol. 2, pp. 1247 – 1252, 1994




0-7803-5474-5/99/$10.00(C)1999 IEEE