COMPLEX INDEPENDENT COMPONENT ANALYSIS BY NONLINEAR GENERALIZED HEBBIAN LEARNING WITH RAYLEIGH NONLINEARITY

Eraldo Pomponi, Simone Fiori and Francesco Piazza

Dept. Electronics and Automatics – University of Ancona
Via Brecce Bianche I, 60131 Ancona Italy
E-mail:simone@eealab.unian.it

ABSTRACT

The aim of this paper is to present a non-linear extension of the Sanger's Generalized Hebbian Algorithm to the processing of complex-valued data. A possible choice of the involved non-linearity is discussed recalling the Sudjianto-Hassoun interpretation of the nonlinear Hebbian learning. Extension of this interpretation to the complex case leads to a nonlinearity called Rayleigh function, which allows for separating mixed independent complex-valued source signals.

1. INTRODUCTION

Independent Component Analysis (ICA) of complex-valued data [2, 3, 8] 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 ICA of real-valued data. Among others, those methods based on non-linear extensions of Principal Component Analysis (PCA) have raised a lot of interest in the Neural Network community (see for example [5, 7] and references therein). It has been proved by many papers that adding nonlinearity to linear PCA neural makes them able to improve the independence of their outputs so as to allow blind separation of independent sources [5, 7]. 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 [4], while in [6] 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 GHA, and discuss the choice of the non-linearity under the theoretical framework proposed by Sudjianto and Hassoun [10] extended to the complex case. Then we show how a particular non-linearity, called Rayleigh function, allows the network to separate out mixed independent complex-valued source signals.

This research was supported by the Italian MURST.

2. COMPLEX NONLINEAR GENERALIZED HEBBIAN LEARNING

We consider a complex-weighted single-layer neural network, formed by linear units, which performs non-classic Principal Component Analysis of complex-valued data. The network is described by the input vector $\mathbf{x} \in \mathcal{C}^p$, a set of weight-vectors $\mathbf{w}_k$ and outputs $y_k \stackrel{\text{def}}{=} \mathbf{w}_k^H \mathbf{x}$, where "$\cdot^H$" denotes conjugate transpose. The number of neurons of the network is denoted here with $m \leq p$.

Let the following criterion be defined:

$$J(\mathbf{w}_k) \stackrel{\text{def}}{=} U(\mathbf{w}_k) + L(\mathbf{w}_k) \ . \tag{1}$$

The criterion $U(\cdot)$ contains a nonlinear function of the k^{th} neuron's output and is defined as follows:

$$U(\mathbf{w}_k) \stackrel{\text{def}}{=} E_{\mathbf{x}}[f(\mathbf{w}_k^H \mathbf{x})|\mathbf{w}_k] \ , \tag{2}$$

where the symbol $E_{\mathbf{x}}[f|\mathbf{w}]$ denotes conditional expectation of f with respect to $\mathbf{x}$ subject to the hypothesis $\mathbf{w}$, hereafter simply written in short notation as $E[f]$. The $f(\cdot)$ is a real valued, positive function of complex-valued argument, and by definition is supposed of the form:

$$f(z) \stackrel{\text{def}}{=} g(|z|) \ , z \in \mathcal{C} \ , g : \mathcal{R}_0^+ \to \mathcal{R}_0^+ \ , \tag{3}$$

with $g(\cdot)$ being continuously differentiable almost everywhere, non-decreasing with a unique minimum in $|z| = 0$.

As far as PCA is concerned, the adjoint function $L(\cdot)$ is used for embedding on the criterion (1) the necessary constraints of orthonormality of the weight vectors, namely $\mathbf{w}_\alpha^H \mathbf{w}_\beta = 0$ if $\alpha \neq \beta$, and $\mathbf{w}_\alpha^H \mathbf{w}_\alpha = 1$. Note that orthogonality conditions can be rewritten more conveniently by observing that $\mathbf{w}_\alpha^H \mathbf{w}_\beta = 0$ if and only if $\mathrm{Re}\{\mathbf{w}_\alpha^H \mathbf{w}_\beta\} = 0$ and $\mathrm{Im}\{\mathbf{w}_\alpha^H \mathbf{w}_\beta\} = 0$, thus the function $L(\cdot)$ may be expressed as:

$$L(\mathbf{w}_k) \stackrel{\text{def}}{=} \sigma_{kk}(\mathbf{w}_k^H \mathbf{w}_k - 1) + \sum_{j=1}^{k-1} \mathrm{Re}\{\sigma_{kj}^\star \mathbf{w}_k^H \mathbf{w}_j\} \ , \quad (4)$$

where a set of complex Lagrange multipliers $\{\sigma_{kj}\}$ has been introduced, and superscript "$\star$" denotes conjugation.

To look for optimal weights $\mathbf{w}_k^{\text{opt}}$ maximizing the criterion (1), a Gradient Steepest Ascent (GSA) learning algorithm is employed here. By definition, the gradient of a real-valued function $F(\mathbf{w})$ with respect to a complex-valued vector $\mathbf{w}$ is intended as:

$$\frac{\partial F(\mathbf{w})}{\partial \mathbf{w}} \stackrel{\text{def}}{=} \frac{\partial F(\mathbf{u},\mathbf{v})}{\partial \mathbf{u}} + i\frac{\partial F(\mathbf{u},\mathbf{v})}{\partial \mathbf{v}} \ , \quad (5)$$

where $i \stackrel{\text{def}}{=} \sqrt{-1}$ and $\mathbf{u} + i\mathbf{v} = \mathbf{w}$. First, the aim is to evaluate:

$$\frac{\partial U(\mathbf{w}_k)}{\partial \mathbf{w}_k} = E\left[\frac{dg(|y_k|)}{d|y_k|}\frac{\partial |y_k|}{\partial \mathbf{w}_k}\right] = E\left[g'(|y_k|)\frac{\partial |y_k|}{\partial \mathbf{w}_k}\right] \ . \quad (6)$$

From definition (5) it follows that $|y_k|\frac{\partial |y_k|}{\partial \mathbf{w}_k} = y_k^\star \mathbf{x}$, thus the expression of the gradient of the objective function $U(\cdot)$ is:

$$\frac{\partial U(\mathbf{w}_k)}{\partial \mathbf{w}_k} = E\left[\frac{g'(|y_k|)}{|y_k|} y_k^\star \mathbf{x}\right] \ . \quad (7)$$

Moreover, the gradient of $L(\mathbf{w}_k)$ with respect to $\mathbf{w}_k$ is found to be:

$$\frac{\partial L(\mathbf{w}_k)}{\partial \mathbf{w}_k} = 2\sigma_{kk}\mathbf{w}_k + \sum_{j=1}^{k-1} \sigma_{kj}^\star \mathbf{w}_j \ . \quad (8)$$

Thus, by gathering equations (7) and (8) we obtain:

$$\frac{\partial J(\mathbf{w}_k)}{\partial \mathbf{w}_k} = E\left[\frac{g'(|y_k|)}{|y_k|} y_k^\star \mathbf{x}\right] + 2\sigma_{kk}\mathbf{w}_k + \sum_{j=1}^{k-1} \sigma_{kj}^\star \mathbf{w}_j \ . \quad (9)$$

The optimal multipliers as functions of $\mathbf{w}_k$ can be found by solving equations $\mathbf{w}_h^H \frac{\partial J}{\partial \mathbf{w}_k} = 0$ for different values of the index h. In this case we have:

$$E\left[\frac{g'(|y_k|)}{|y_k|} y_k^\star y_h\right] + 2\sigma_{kk}\mathbf{w}_h^H \mathbf{w}_k + \sum_{j=1}^{k-1} \sigma_{kj}^\star \mathbf{w}_h^H \mathbf{w}_j = 0 \ ,$$

thus letting $h = k$ gives $\sigma_{kk}^{\text{opt}} = -\frac{1}{2}E[g'(|y_k|)|y_k|]$, while assuming $h \neq k$ leads to $\sigma_{kh}^{\text{opt}} = -E\left[\frac{g'(|y_k|)}{|y_k|} y_k y_h^\star\right]$. By plugging these expressions into equation (9) the formula for the optimal gradient of J is easily found to be:

$$\left(\frac{\partial J}{\partial \mathbf{w}_k}\right)^{\text{opt}} = E\left\{\frac{g'(|y_k|)}{|y_k|} y_k^\star \left[\mathbf{x} - \sum_{j=1}^{k} y_j \mathbf{w}_j\right]\right\} \ . \quad (10)$$

Finally, by defining the projection operator $\mathbf{P}_k \stackrel{\text{def}}{=} \mathbf{I} - \sum_{j=1}^{k} \mathbf{w}_j \mathbf{w}_j^H$ and $G(\zeta) \stackrel{\text{def}}{=} \frac{dg(|\zeta|)}{d|\zeta|}\frac{1}{|\zeta|}$, with $\zeta \in \mathcal{C}$, the new complex non-classic counterpart of GHA learning rule writes:

$$\frac{d\mathbf{w}_k}{dt} = \mathbf{P}_k E[G(y_k) y_k^\star \mathbf{x}] \ , \ k = 1, 2, \ldots, m \ . \quad (11)$$

The factor $E[G(y_k)y_k^\star \mathbf{x}]$ may be interpreted as a complex non-classic Hebbian term common to each neuron, while projector $\mathbf{P}_k$ is a deflating factor which pushes each weight-vector $\mathbf{w}_k$ into a different subspace. About function $g(\cdot)$, it can be chosen on the basis of the specific task for which the network is used. It deserves to note that assuming $g(u) = \frac{1}{2}u^2$ yields $G(\cdot) = 1$, thus in this case and in presence of real-valued data algorithm (11) coincides to well-known GHA rule by T.D. Sanger [9].

3. THE SUDJIANTO-HASSOUN INTERPRETATION

In [10], Sudjianto and Hassoun considered the problem of maximizing a criterion $J(\mathbf{w}) \stackrel{\text{def}}{=} E[S^2(\mathbf{w}^T\mathbf{x})]$ 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 $S(\cdot) \in [-1, +1]$. The authors 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 [10]. The GSA learning rule for the neuron is:

$$\frac{d\mathbf{w}}{dt} = \frac{\partial J}{\partial \mathbf{w}} = (\mathbf{I} - \mathbf{w}\mathbf{w}^T)E[\ell(y)\mathbf{x}] \ , \quad (12)$$

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 [10] that z will be uniformly distributed within $[-1,+1]$. The central idea developed by Sudjianto and Hassoun is that the learning rule (12) 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. In other words, the rule (12) behaves as a *probabilistic filter*.

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

$$U(\mathbf{w}) \stackrel{\text{def}}{=} E[S^2(|y|)] \ , \qquad (13)$$

for a complex weighted neuron with output $y = \mathbf{w}^H\mathbf{x}$. Its GSA maximization under the constraint $\mathbf{w}^H\mathbf{w} = 1$ yields the learning rule:

$$\frac{d\mathbf{w}}{dt} = (\mathbf{I} - \mathbf{w}\mathbf{w}^H)E\left[\ell(|y|)\frac{y^\star}{|y|}\mathbf{x}\right] \ , \qquad (14)$$

that closely recalls equation (11) for $k = 1$.In our case we can assume for $S(|y|)$ a function like:

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

where $q(\cdot)$ represents a generic probability density functions. By equating (14) to (11) it is possible to find the relationship between $q(\cdot)$ and $g'(\cdot)$, that is:

$$g'(u) = 2q(u)\int_0^u q(\eta)d\eta \ . \qquad (15)$$

Ultimately it is clear that training each neuron of a linear complex-weighted neural network by means of the learning rule (11) with the non-linearity (15) causes the network to learn connection strengths that filter the outputs so that the probability density function of the output moduli $|y_k|$ deviates away from $q(\cdot)$. In the next section it will be clarified how this principle could be employed for separating out independent complex signals from their linear mixtures.

4. APPLICATION TO COMPLEX INDEPENDENT COMPONENT ANALYSIS

Suppose input $\mathbf{x}$ contains a complex linear mixture of statistically independent signals [8], 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 σ^2. Then it is known that the modulus $|v|$ follows the Rayleigh distribution:

$$q_{\mathrm{R}}(u) = \frac{u}{\sigma^2}\exp\left(-\frac{u^2}{2\sigma^2}\right)\Gamma(u) \ ,$$

where $\Gamma(u)$ is the unit step. Then by formula (15) we find:

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

Figure 1 depicts the Rayleigh non-linearity $g'_{\mathrm{R}}(u)/u$ for a unit noise power. In this case it is possible to express

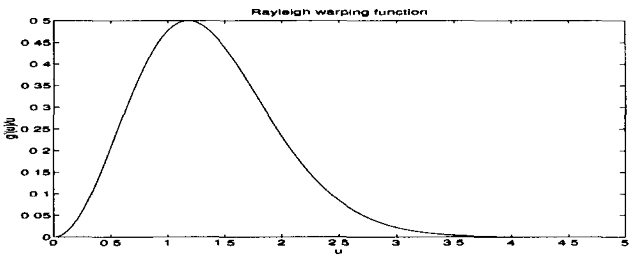

Figure 1: Rayleigh warping function for $\sigma = 1$.

the cumulative distribution function in closed form simply as:

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

By assuming in (11) the function $G(y_k)$ as the quantity $\frac{g'_{\mathrm{R}}(|y_k|)}{|y_k|}$, 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, 8].

5. COMPUTER SIMULATIONS

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 as in [8]. The mixture is computed as $\mathbf{x} = \mathbf{M}\mathbf{s}$, where $\mathbf{M}$ is a randomly generated 4×4 complex matrix. The first row of Figure 2 depicts the independent signals while second row shows the obtained four mixtures.

By means of the Sudjianto-Hassoun principle, a linear neural network with four inputs and four outputs, trained by the learning rule (11) with the Rayleigh nonlinearity should be able to separate out the independent signals up to a phase shift and a random permutation [2] after mixture prewhitening. Simulation results are shown in Figure 3: The first row depicts the result of prewhitening performed by means of the well-known Laheld-Cardoso's standardizing algorithm [8]; the second row shows the last 100 outputs of the network trained by (11) on the prewhitened data. The Fig-

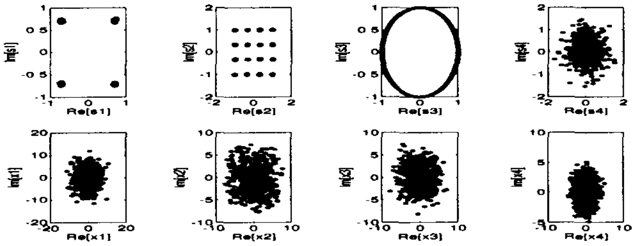

Figure 2: The four independent signals and the four mixtures of them.

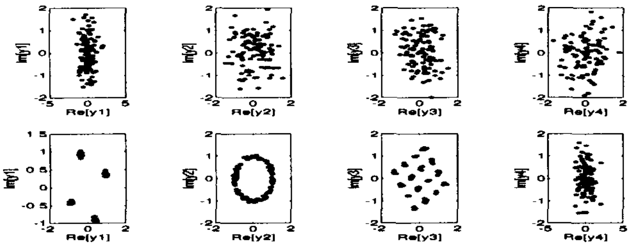

Figure 3: Network's output after learning by rule (11).

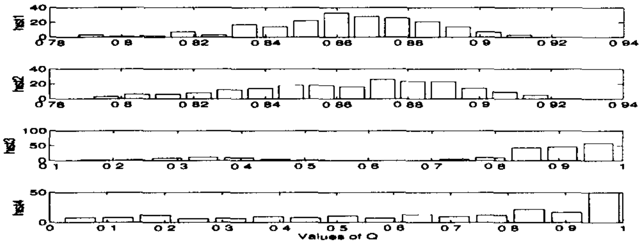

Figure 4: Functions $Q_R(\cdot)$ histograms.

ure 4 depicts instead the histograms of the last 200 samples of $Q_R(|y_1|), \ldots, Q_R(|y_4|)$.

Simulation results show that the network is able to recover the independent signals. The histograms $Q_R(|y_1|)$, $Q_R(|y_2|)$ and $Q_R(|y_3|)$ are in good accordance with the signals in Figure 3, while the presence of a peak in $+1$ on the histogram of $Q_R(|y_4|)$ 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

In this paper a new adapting rule for linear neural networks as generalization of Hebbian learning has been presented. It provides a generalization in that it applies to complex-weighted neural networks and embeds non-linearity in the classic Hebbian learning. A particular choice of the non-linearity is discussed by recalling the Sudjianto-Hassoun interpretation of non-classic Hebbian learning extended to the complex-case. Numerical results confirm that non-classic ('non-linear') complex generalized Hebbian learning is closely related to Independent Component Analysis and Blind Source Separation.

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

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

[5] 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

[6] S. Fiori and A. Uncini, *A unified approach to laterally-connected neural networks*, to appear on Proc. EUSIPCO, Rhodes, Greece, Sept. 8-11, 1998

[7] 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

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

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

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