SELF-ORGANIZING BLIND MIMO DECONVOLUTION USING LATERAL-INHIBITION

Inbar Fijalkow, Philippe Gaussier

Equipe de Traitement des Images et du Signal, URA CNRS 2235
ETIS / ENSEA - Univ. de Cergy-Pontoise, 6 av du Ponceau
95014 Cergy-Pontoise Cdx, France
fijalkow, gaussier@ensea.fr

ABSTRACT

In this paper[1], we address the architecture design of linear filters for the blind adaptive restoration of several sources from convolutive mixtures. First, we show the drawbacks of fixed architectures involving additive decorrelation constraints and a source extraction criterion. We introduce a self-organizing architecture based on the dynamical stability properties of neural network lateral inhibition rules. This exploratory search of an optimal design helped our understanding of the difficulties raised by using the correlation information. It opens directions for further improvements.
Keywords: Blind MIMO deconvolution, Self-organization, Neural competition, Adaptive deconvolution.

1. INTRODUCTION

The problem of blind source deconvolution is to recover the independent source signals from linear and convolutive mixtures (i.e., in time and space) without the knowledge of the mixing matricial transfer function. This problem is motivated by many potential applications. For instance, in the cocktail party task, several voices are received by several microphones after convolution by the room impulse response from each speaker location to each microphone. It is then desirable to restore each speech signal adaptively since the speakers can be slowly moving and without prior knowledge of the transfer function. Among other applications is the restoration of multiple signals transmitted in a multipath environement [15].

The problem of convolutive mixtures can not be reduced to the simpler instantaneous problem for which a lot of recent research has been performed. However, some criteria can be considered for both cases when each source to be restored is an independent and identically distributed sequence (i.i.d.). It is the case of some criteria based on a non-linearity used for blind equalization as well as for blind source extraction. In particular, the Constant Modulus (CM) criterion first proposed for blind equalization [5] has been widely used for sources extraction from spatial mixtures and more recently introduced for the deconvolution of MIMO spatio-temporal mixtures [15]. There are indeed other criteria based the fourth order moments that can be used instead of CM, this is not the purpose of this paper. Our concern is about the architecture that controls the interaction between the filters trying to extract different signals. This problem is raised by the fact that a blind criterion such as CM allows perfect restoration of **one arbitrary source, with an arbitrary delay** from convolutive mixtures. It requires the mixtures to have sufficient spatio-temporal effective diversity (see for instance [12]). The problem is therefore to control that each filter selects a different source. To do so one can control the filters via parametrization, using a deflation approach, see [2] in the case of sources separation. A subtraction of already extracted sources was also considered, in [16], for non adaptive processing. However, since the CM criterion allows perfect source extraction, it is sufficient to make sure that the filters outputs are not correlated.

Different decorrelation rules were proposed to control how different filters can restore different sources. A symmetrical decorrelation constraint was added to the CM criterion in [12], [10] and others. It raised some problems of undesired local minima and possible slow convergence due to two filters preventing each other because of their symmetrical behaviour, to escape from a given basin of attraction in which they both have been initialized. To prevent from this kind of problem, a hierarchical approach was proposed in [14] and proved to be asymptotically satisfactory. It imposes a hierarchical order so that each filter filter imposes on the later

[1] This study was performed while the authors were visiting the Research School of Information Science and Engineering at the Australian National University, Canberra, Australia.

ones to be in different basins of attractions, therefore solving the problem of preventing escape from a given basin of attraction. However, the hierarchical order being given a priori is sometimes not the best in terms of convergence speed, i.e., the second filter initialisation may be closer to a minimum from which it will be discarded. In this paper, we challenge this question by a self-organizing filters structure using dynamical stability properties exhibited by lateral inhibition mechanisms in neural networks (see the dynamic field theory for the control of motor behavior for instance [13]).

2. PROBLEM SETTING

Let us consider P source symbols $(s_k(n))_{k=1,..,P}$ emitted from different locations and observed at the output of L spatially distributed sensors. At each instant, these observations are collected in a L-variate vector $\mathbf{y}(n)$ viewed as the sum of the P sources contributions as in:

$$\mathbf{y}(n) = \sum_{k=1}^{P} \sum_{m=0}^{Q} \mathbf{h}_k(m) s_k(n-m) \qquad (1)$$

The transfer function $\mathbf{h}_k(z) = \sum_{m=0}^{Q} \mathbf{h}_k(m) z^{-m}$ modelizes the transmission media between the location of $s_k(n)$ and the sensors, for $k = 1,..,P$. Let us denote $\mathbf{H}(z) = [\mathbf{h}_1(z), ..., \mathbf{h}_P(z)]$, the MIMO transfer function from the P sources to the L sensors.

In the absence of noise, the FIR linear deconvolution problem consists in finding P transfer functions $(\mathbf{g}_p(z))_{p=1,...,P}$ such that $\mathbf{G}(z) = (\mathbf{g}_1(z), \cdots, \mathbf{g}_P(z))$ satisfies,

$$\mathbf{G}(z)^\top \mathbf{H}(z) = \begin{bmatrix} z^{-\nu_1} & & \\ & \ddots & \\ & & z^{-\nu_P} \end{bmatrix}$$

up to some source permutation and scaling. $\nu_1, .., \nu_P$ are integer delays that are in the range of achievable delays for a given channel and equalizer length. In the sequel, we denote $N-1$ the different equalizers degree, so that $\mathbf{g}_p(z) = \sum_{k=0}^{N-1} \mathbf{g}_p(k) z^{-k}$. $\mathbf{g}_p$ denotes the corresponding impulse response $\mathbf{g}_p = (\mathbf{g}_p(0)^\top, ..., \mathbf{g}_p(N-1)^\top)^\top$.

2.1. System invertibility

This is known to be possible ([9]) under the following **invertibility assumptions**:

- $P < L$
- $\mathbf{h}_k(z)$ degree Q_k ; Rank $\mathbf{H}(z) = P, \forall z$; Rank $\mathbf{H}_C(z) = P, \forall z$,
- $\mathbf{G}(z)$ degree N; $N \geq \sum_{k=1}^{P} Q_k$.

The question is how to do it blindly, i.e., without knowledge of neither the input sequences or of the system transfer function.

2.2. One source extraction

It is also known that a blind algorithm such as CM algorithm (CMA) is able to adapt blindly the impulse response $\mathbf{g}$ of a degree $N-1$ filter $\mathbf{g}(z) = \sum_{k=0}^{N-1} \mathbf{g}(k) z^{-k}$, so that it extracts one arbitrary source sequence (with an arbitrary delay in the range of achievable delays) from the mixtures. The CMA adaptation is given by the following stochastic gradient descent equation:

$$\mathbf{g}^{(n+1)} = \underbrace{\mathbf{g}^{(n)} - \mu x(n)(|x(n)|^2 - 1)\mathbf{Y}^*(n)}_{CMA(n)} \qquad (2)$$

where $x(n) = (\mathbf{g}^{(n)})^\top \mathbf{Y}(n)$ is the equalizer output and $\mathbf{Y}(n) = (\mathbf{y}(n), ..., \mathbf{y}(n-N+1)$ is the regression vector.

The extraction is shown to be perfect (i.e., suppression of the contributions of other sources) if the system invertibility conditions are satisfied, see for instance [12]. Other reasons to consider an approach based on CMA are its robustness to additive noise, to an overestimation of the transfer function order and to a lack of system invertibility, see [3] for the one source case.

The use of such an approach to extract several sources requires to make sure that the same source is not selected twice, even with different delays. The control of the different delays makes the task much more complex than in the instantaneous source separation case.

3. EXISTING APPROACHES

In this section, we present only existing approaches based on a decorrelation constraint. It is important to understand the lacks of such fixed architectures in order to understand why we are looking for self-organizing ones.

3.1. Symmetrical approach

Since source extraction is perfectly achievable, a natural idea is to add a decorrelation constraint between all equalizers outputs at all possible delays to prevent from selecting twice the same source. The intercorrelation between the outputs $x_i(n)$ and $x_j(n)$ of two filters, $\mathbf{g}_i$ and $\mathbf{g}_j$ (with $x_k(n) = \mathbf{g}_k^{(n)\top} \mathbf{Y}(n)$), is denoted by $E[x_i(n) x_j(n-m)] = \mathbf{g}_i^\top \mathcal{R}(m) \mathbf{g}_j$ with $\mathcal{R}(m) = E[\mathbf{Y}(n)\mathbf{Y}(n-m)^\top]$. The decorrelation constraint can be expressed as $\sum_{i \neq j} \sum_{m=-M}^{M} (\mathbf{g}_i^\top \mathcal{R}(m) \mathbf{g}_j)^2$ where M overestimates $L(N+Q)$, for instance one can take $2LN$.

A stochastic algorithm can be deduced with the following updating equation,

$$\mathbf{g}_k^{(n+1)} = CMA_k(n) - \mu_2 \underbrace{\sum_{l \neq k} \sum_{m=-M}^{M} \mathbf{g}_k^{(n)\top} \hat{\mathcal{R}}(m) \mathbf{g}_l^{(n)} \hat{\mathcal{R}}(m) \mathbf{g}_l^{(n)}}_{\mathbf{C}_{k,l}(n)} \quad (3)$$

where $CMA_k(n)$ is the CMA updating term defined in (2) for $\mathbf{g}_k$. $\hat{\mathcal{R}}(m)$ is an estimation of $\mathcal{R}(m)$ updated at each iteration with a forgetting factor.

Note that since the decorrelation term is not the mean expected value of some expression, (3) is not the stochastic gradient descent algorithm of the constraint cost function. The analytical study of the performances of such an algorithm is difficult. However, from simulations, we have noticed that in the case of an initialization in one basin of attraction for two filters (for instance, in the worst possible case with the same initial setting), the two filters may push each other in and out the basin of attraction for a very large number of iteration because of their similar updating rule. The hierarchical approach was proposed in order to prevent from such a behaviour.

3.2. Hierarchical approach

In the hierarchical organized algorithm the decorrelation term is taken into account only to constraint the following (in the given order) filters. The resulting algorithm was proposed in [14],

$$\mathbf{g}_k^{(n+1)} = CMA_k(n) - \mu_2 \sum_{l > k} \mathbf{C}_{k,l}(n) \quad (4)$$

The structure was shown to have the desired asymptotical mean behaviour when $\mu_2/\mu > 2\rho_p$ where $\rho_p = E[s_p^4]/E[s_p^2]^2$ is the normalized fourth order moment of the input source to be extracted.

If the filters initializations correspond to the order of the given hierarchy, the algorithm is going to converge very quickly. However, if the filter at the head of the hierarchy is far from convergence and if its convergence pushes over filters out of the basins of attraction in which they were initialized, very slow convergence may be induced.

4. SELF-ORGANIZING HIERARCHY

4.1. Aim

We want hierarchical order to adapt itself to the system, so that if one filter is closer to convergence than the others it will push the others outside of the basins of attractions corresponding to extraction of the same source. The resulting algorithm should of the form

$$\mathbf{g}_k^{(n+1)} = CMA_k(n) - \mu_2 \sum_{l \neq k} \theta_{k,l}(n) \mathbf{C}_{k,l}(n) \quad (5)$$

where the scalar $\theta_{k,l}(n)$ should be adapted in order to match the optimal hierarchy and to suppress the decorrelation terms when no longer required (i.e., each filter is in the basin of attraction of a different source). This is not necessary to ensure asymptotic convergence but it would decrease the residual jitter when steady-state is close to being reached. In practical (non-asymptotical) situations, reducing the jitter is all the more crucial than the number of input sources is large.

The self-organized competition is aimed to choose a winner (principle of Winner Take All (WTA) mechanism [11]) the success of which can be measured by a small CM error and decorrelation from the other filters outputs. The instantaneous CM errors of all filter outputs, $(|x_k(n)|^2 - 1)^2$ with $k \in [1, P]$, could be considered to set up a hierarchy. However, tests show that making instantaneous decisions induces lots of instability since the errors are jittering due to the stochastic character of the algorithm. An integration in time of the competition variables is therefore required to regulate the competition between the filters. This integration needs to be performed inside the competition, see [1]. Many competition rules can be experimented. In the sequel, we introduce this approach for our problem and exhibit examples to prove the interest of such an approach.

4.2. Two sources case

For sake of simplicity, let us first consider the case of two sources. In the case of Winner Take All (WTA) with fixed inhibition weights, the competition is performed using a measure of success of source restoration denoted $\phi_k(n)$ (the simplest expression being a function of the CMA error obtained from $x_k(n)$), as in Figure 1.

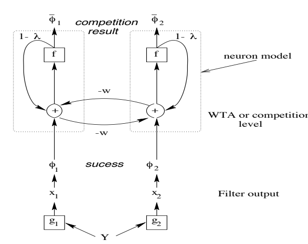


Figure 1: Competition scheme

The neurons output $\bar{\phi}_k(n)$ are competing using the following equations:

$$\bar{\phi}_{k_1}(n+1) = f\left((1-\lambda)\bar{\phi}_{k_1}(n) + \alpha_1 \phi_{k_1}(n) - w\bar{\phi}_{k_2}(n)\right) \quad (6)$$

for $k_1 \neq k_2$. λ and α_1 are respectively the forgetting and acquiring factors of the integration of the success measure. w is the anti-Hebbian term setting the fixed inhibition. The neuron output $\bar{\phi}_k(n)$ resulting from (6) is forced into the interval $[0,1]$ using f. $\bar{\phi}_k(n)$ tends to 1 as it is winning. The weighting factors for decorrelation can be easily described by

$$\theta_{k_1,k_2}(n) = (1 - \bar{\phi}_{k_1}(n)), \quad k_1 \neq k_2 \quad (7)$$

Note that θ_{k_1,k_2} is no longer explicitly depending on k_2. This mechanism enhances the filter that has more success over a time duration depending on the fixed factors in (6). The study of such competition mechanism [6, 7] shows that it stabilizes asymptotically on one winner in static conditions when the instantaneous successes are different enough from each other. This requires indeed some conditions on the parameters λ, α_1 and w depending on the input distributions.

The main difficulty in applying such a competition mechanism to our deconvolution problem consists in choosing the measure of success $\phi_k(n)$ for the deconvolution task. It seems natural to use the CMA error the minimization of which ensure restoration of one source. One could think of choosing the inverse of the CMA error $1/(|x_k(n)|^2 - 1)^2$ for $\phi_k(n)$, but it is not bounded. We considered $\phi_k(n) = f(1 - (|x_k(n)|^2 - 1)^2)/e_{max})$ where e_{max} is maximal error allowed. When one CMA error is larger than e_{max}, $\phi_k(n) = 0$ so that the associate neuron output will decrease and will be less likely to win. This is not a problem since we assume that such error level is too large for a successful source deconvolution. This is certainly not the optimal choice, but only an exploratory one and the simplest linear function of the CMA error.

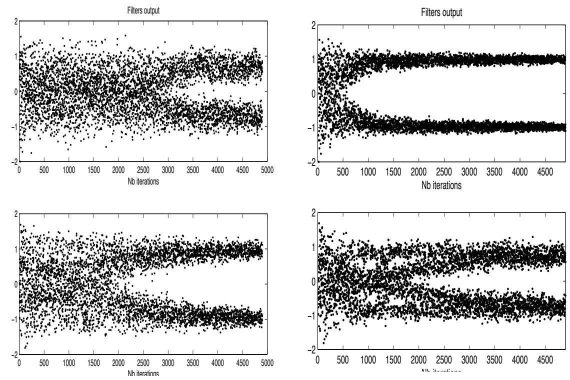
Simulation: The following simulation setting considers 2 sources and 3 sensors with the transfer function satisfying the invertibility conditions given in Table I. The sources are binary (± 1) i.i.d. sequences, independent from each other. For the WTA parameters, we have selected $e_{max} = 2$, $\lambda = 0.025$, $\alpha_1 = 0.7$, $w = 0.6$. For the CMA $\mu = 0.01$.

Table I : Impulse response of 2-sources / 3-sensors system.

$\mathbf{h}_1(0)$	$\mathbf{h}_1(1)$	$\mathbf{h}_1(2)$	$\mathbf{h}_2(0)$	$\mathbf{h}_2(1)$	$\mathbf{h}_2(2)$
0.0400	0.9359	0.3686	0.7575	0.4756	0.6541
0.7591	0.0907	0.2895	0.8320	0.9996	0.9651
0.5619	0.2927	0.7802	0.7853	0.1786	0.1635

The initialization of the two filters are center-spike vectors corresponding to basins of attraction of the same source (with the non-zero component at the respective positions 3 and 10 for initilization (a) and swapped for initialization (b)).

We compare the filters output for the hierarchical algorithm (Figure 2) and the WTA algorithm (Figure 3).



(a) (b)

Figure 2: Hierarchical order: filters output with swapped initialisations

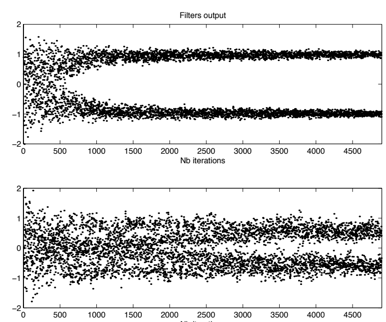
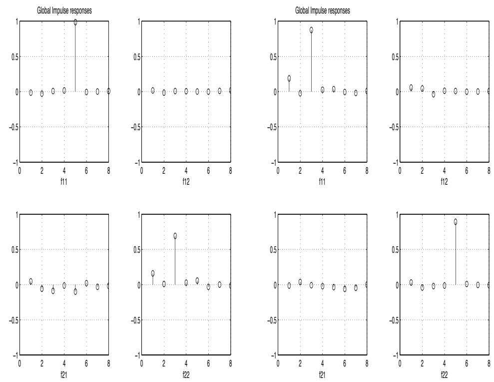


Figure 3: WTA: filters output

On this simulation, alike in many others we have performed, the WTA seems to behave as an average between the hierarchical algorithm best performances (i.e., when the the given order is satisfactory) and worst performances (i.e., when the the given order is not satisfactory).

Figure 4 displays the global impulse responses of the system averaged over the last 500 iterations. The quadrant i, j represents the impulse response between the source i and the output j. We can that the sources are well separated but that the second one is not very well equalized (some small non-zero taps).

If we extend the WTA rule to the case of more than



(a) (b)

Figure 4: Hierarchical order: global impulse responses with swapped initialisations

two sources, the inhibition term becomes $-w \sum_{k \neq k_1} \bar{\phi}_k(n)$. Still, we get one winner for all filters. Such a processing would require to wait for the decorrelation term to be very small before another filter is allowed to select another source. In the next paragraph, we let the WTA inhibition factor w to adapt itself, by taking the decorrelation into account, to allow several filters to win the competition when they restore different sources.

4.3. Lateral-Inhibition rule

We follow the work of [4] to adapt the inhibition factor in (6). This factor is adapted so to increase inhibition when the two filters outputs are correlated and not successful. On the opposite, the inhibition from a given source should stop if the filters outputs are uncorrelated (or sufficiently decorrelated). The updating equation is given by: $w_{k_1,k_2}(n+1) =$

$$(1-\lambda_2) w_{k_1,k_2}(n) + \alpha_2 \phi_{k_1}(n) \phi_{k_2}(n) f_s(c\bar{o}rr_{k_1,k_2}(n)) \tag{8}$$

for $k_1 \neq k_2$. λ_2 is a forgetting factor. The usual Hebbian term to learn the neurons correlation is given by $\alpha_2 \phi_{k_1}(n) \phi_{k_2}(n)$. In (8), the Hebbian term is modulated by the way the filters outputs are correlated denoted by $f_s(c\bar{o}rr_{k_1,k_2}(n))$. At this stage, the difficulty concerns the way to take into account the measure of the integrated correlation $c\bar{o}rr_{k_1,k_2}(n)$. between $x_{k_1}(n)$ and $x_{k_2}(n)$. $c\bar{o}rr_{k_1,k_2}(n)$ is adapted by (9) bellow. We choose f_s to be a binary function defined by $f_s(c) = 0$ if c is smaller than a given threshold t_s, 1 otherwise. Indeed, this choice is far from being optimal, and one may think of adapting the threshold during the process.

The integrated correlation is simply adapted as,

$$c\bar{o}rr_{k_1,k_2}(n) = (1-\lambda_3) c\bar{o}rr_{k_1,k_2}(n) + \lambda_3 \frac{corr_{k_1,k_2}(n)}{||\mathbf{g}_{k_1}(n)||||\mathbf{g}_{k_2}(n)||} \tag{9}$$

with $corr_{k_1,k_2}(n) = \mathbf{g}_{k_1}^{\top}(n) \mathbf{C}_{k_1,k_2}(n)$. λ_3 is the forgetting factor.

4.4. Simulations

The simulation setting is exactly the same as previously. $w_{k_1,k_2}(0) = 1$, $\lambda_2 = 0.001$, $\alpha_2 = 1$, $t_s = 0.2$ $\lambda_3 = 0.01$.

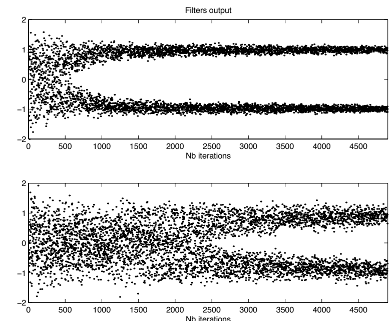


Figure 5: Adaptive WTA: filters output

First note that we obtain the same simulation results independently of a permutation of the initial filters settings. When the decorrelation term is no longer effective after 2000 iterations, both filters are updated by CMA only. Figure 6 displays the global impulse responses of the system averaged over the last 500 iterations. Obvisously the two different sources have been restored almost perfectly, the improvement is consequent compared with the hierarchical algorithm in Figure 4. Nevertheless, the second filter output (lower part of Figure 5) shows a quite large variance. We believe that this jitter around ± 1 is due to the slow convergence of the filter adaptation with spatio-temporal diversity (ill conditioned covariance matrix of $\mathbf{Y}$) in tracking mode. Some filter taps with small coefficients are lacking excitation to be tuned correctly. To overcome this phenomenon, one may think of introducing a forgetting factor (or leakage) in (2). An alternative solution would be to switch to a decision directed mode. It means that $CMA_K(n)$ is replaced by an LMS type of error with the desired value being replaced by a decision taken from the filter output.

Finally in Figure 7, the neuron outputs are displayed. $\bar{\phi}_1$ wins almost instantaneously, and remains at 1. $\bar{\phi}_2$ is first set to 0, being therefore subject the decorrelation constraint from the first neuron. After only 1000 iterations, the two filters outputs are decorrelated enough meaning that the second filter succeeded

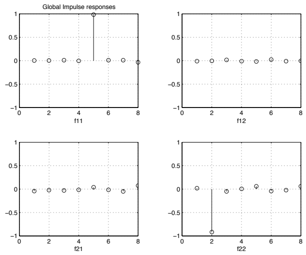


Figure 6: Adaptive WTA: global system impulse responses

in being in a new basin of attraction. This induces a decrease of the inhibition weight, so that the second neuron is also allowed to win. The filters adaptations are then completly decoupled (CMA adaptation only). This shows the relevance of adapting the inhibition weight.

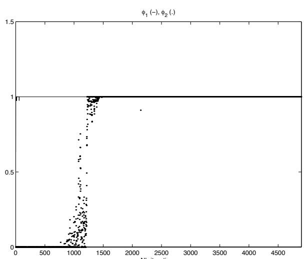


Figure 7: Adaptive WTA: measures of succes $\bar{\phi}_1$, $\bar{\phi}_2$

5. CONCLUSION

In this paper, we explore ways to improve the convergence properties of MIMO convolutive mixtures deconvolution approaches based on decorrelation constraints. In the case of two sources a Winner Take All architecture is applied and seems to be quite satisfactory.

The dynamical properties of the proposed WTA mechanism are usefull to allow the adaptive filters to specialize on different sources. These properties are crucial when sources changes (or appear or deseppear) during the experiment. The competition mechanism allows to change winner very quickly if the winner is not considered as being safe. The mechanism is also robust to perturbations when a safe winner has been found. When our system has succeed to restore corretly 2 sources and that we change them it reactivates in few iterations (10) the competition mechanism and finds the new solution in the same time than the presented simulation (no differences in the figures).

The new approches should be even more outperforming in the case of more than two sources, especially if the CMA adaptation is switched to a decision directed mode when possible. These feature and the optimisation of the proposed design are yet to be explored.

6. REFERENCES

[1] S. Amari, *Dynamics of pattern formation in lateral-inhibition type neural fields*, in Biological Cybernetics, (27):77-87, 1977.

[2] N. Delfosse, P. Loubaton, *Adaptive separation of independent sources: a deflation approach*, in Proc. ICASSP-94, 4:41-44, 1994.

[3] I. Fijalkow, A. Touzni, J.R. Treichler, *Fractionally-spaced equalization using CMA: Robustness to channel noise and lack of disparity*, IEEE trans. on SP, 45(1):56-66, January 1997.

[4] P. Földiák, *Adaptive network for optimal linear feature extraction*, in Intern. Joint Conf. on Neural Networks, pp.401-405, 1989.

[5] D. Godard, *Self-recovering equalization and carrier tracking in two dimensional data communication systems*, IEEE Trans. on Com., 28:1867-1875, 1980.

[6] S. Grossberg, *Nonlinear neural networks: principles, mechanisms, and architectures*, in Neural Networks, 1:17-61, 1988.

[7] S. Haykin, *Neural Networks - A comprehensive foundation*, Mc Millan College Pupl. Co., 1994.

[8] D.O. Hebb, *The Organization of Behavior*, Wiley, New York, 1949.

[9] T. Kailath, *Linear Systems*,

[10] C.B. Papadias, A. Paulraj, *A constant Modulus algorithm for multi-user signal separation in presence of delay spread using antenna arrays*, SP Letters.

[11] D.E. Rumelhart et al., *Parallel Distributed Processing*, MIT Press, Cambridge, 1986.

[12] A. Touzni, I. Fijalkow, *Blind adaptive equalization and simultaneous separation of convolutional mixtures*, in Proc. International DSP conf., July 1997.

[13] G. Schöner, M. Dose, C. Engels, *Dynamics of behavior: theory and applications for autonomous robot architectures*, in Robotics and Autonomous System, 16:(2-4)213–245, December, 1995.

[14] A. Touzni, I. Fijalkow, M.G. Larimore and J.R. Treichler, *A globally convergent approach for blind MIMO adaptive deconvolution*, in Proc. ICASSP98, May 1998. A journal version has been submitted to the IEEE Trans. on SP in April 1998.

[15] J.R. Treichler and M.G. Larimore, *New processing techniques based on the constant modulus adaptive algorithm*, IEEE Trans. ASSP, 33:420-431, April 1985.

[16] J.K. Tugnait, *Blind spatio-temporal equalization and impulse response estimation for MIMO channels using a Godard cost function*, IEEE Tr. on SP, 45(1):268-271, January 1997.