Appl. Math. and Comp. Sci., 1994, vol.4, No.3, 331–336

AN EXTENSION OF ADAPTATION ALGORITHMS FOR 2-D FEEDFORWARD NEURAL NETWORKS

TADEUSZ KACZOREK*

The generalized weight adaptation algorithms presented in (Kuschewski *et al.*, 1993; Żak and Sira-Ramirez, 1990) are extended for 2-D madaline and 2-D two-layer FNN's.

1. Introduction

A feedforward neural network (FNN) is a special type of non-recurrent artificial neural network (ANN). It consists of layers of neurons with weighted links connecting the outputs of neurons in one layer to the inputs of neurons in the next layer (Chen, 1990; IJC, 1992; Kuschewski *et al.*, 1993; Psaltis *et al.*, 1988; Widrow and Lehr, 1990; Żak and Sira-Ramirez, 1990). Applications of FNN's to dynamical system indentification and control have been considered in (Kuschewski *et al.*, 1993).

In this paper the well-known (Kuschewski *et al.*, 1993; Żak and Sira-Ramirez, 1990) generalized weight adaptation algorithm for the madaline (many adaptive linear elements) will be extended to the two-dimensional (2-D) case. The generalized weight adaptation algorithm will also be extended for the 2-D two-layer FNN. A need for such extension has arisen while solving some practical problems.

2. 2-D Madaline Weight Adaptation Algorithm

A block diagram of the 2-D madaline weight adaptation algorithm is shown in Fig. 1. It will be shown that the generalized delta rule proposed in (Żak and Sira-Ramirez, 1990; Kuschewski *et al.*, 1993) can be extended for the 2-D madaline.

Let $\theta : \mathbb{R}^n \to \mathbb{R}^n$ be an operator, for example

$$\theta(x) \doteq \begin{bmatrix} \mathrm{sgn}(x_1) \\ \vdots \\ \mathrm{sgn}(x_n) \end{bmatrix}, \qquad \mathrm{sgn}(x_k) = \begin{cases} +1 & \text{if} \quad x_k \geq 0 \\ -1 & \text{if} \quad x_k < 0 \end{cases}, \qquad k = 1, \dots, n$$

W_{ij} is a weight matrix with real entries depending on i and j and $i, j \in \mathbb{Z}_+, \mathbb{Z}_+$ is the set of non-negative integers. Let

$$e_{ij} \doteq y_d - y_{ij} \tag{1}$$

be the error vector at the point (i, j), where y_d is the desired output vector, y_{ij} is the output vector at the point (i, j).

* Institute of Control and Industrial Electronics, Warsaw University of Technology, ul. Koszykowa 75, 00–662 Warsaw, Poland

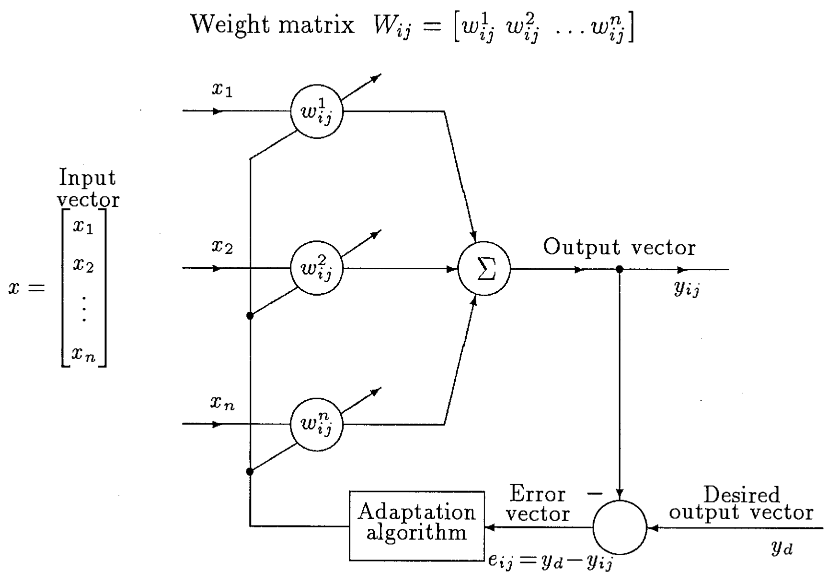

Fig. 1.

Theorem 1. *If the weight matrix W_{ij} of the 2-D madaline is adapted according to the rule*

$$W_{i+1,j+1} = W_{i+1,j} + W_{i,j+1} - W_{ij} + \Delta W_{ij} \tag{2}$$

where

$$\Delta W_{ij} \doteq \begin{cases} (Ae_{i+1,j} + Be_{i,j+1} + Ce_{ij})\dfrac{\theta^T(x)}{\theta^T(x)x} & \text{if } \theta^T(x)x \neq 0 \\ 0 & \text{if } \theta^T(x)x = 0 \end{cases}$$

and the matrices A, B, C are chosen so that all the roots (zero-manifolds) of the equation

$$\det\Big[Iz_1z_2 + (A-I)z_1 + (B-I)z_2 + I + C\Big] = 0 \tag{3}$$

are located in the unit bidisk $U \doteq \{(z_1, z_2) \in \mathbb{C}\times\mathbb{C} : |z_1| < 1, \; |z_2| < 1\}$ ($\mathbb{C}$ is the field of complex numbers), then the error vector e_{ij} converges asymptotically to the zero vector for i and j tending to infinity.

Proof. Using (1), (2) and $y_{ij} = W_{ij}x$ we have

$$\begin{aligned} e_{i+1,j+1} - e_{i+1,j} - e_{i,j+1} + e_{ij} &= -y_{i+1,j+1} + y_{i+1,j} + y_{i,j+1} - y_{ij} \\ &= -(W_{i+1,j+1} - W_{i+1,j} - W_{i,j+1} + W_{ij})x \\ &= -(Ae_{i+1,j} + Be_{i,j+1} + Ce_{ij}) \qquad \text{if } \theta^T(x)x \neq 0 \end{aligned}$$

and

$$e_{i+1,j+1} = (I-A)e_{i+1,j} + (I-B)e_{i,j+1} - (I+C)e_{ij} \tag{4}$$

The characteristic equation of (4) has the form (3). If all the roots of eqn. (3) are located in the unit bidisk, then the vector e_{ij} converges asymptotically to 0 for i and j tending to infinity (Ahmed, 1980; Fernando and Nicholson, 1985; Fornasini and Marchesini, 1980; Kaczorek, 1985). ■

If $A = B = -C = I$, then from (3) we have $z_1 = z_2 = 0$. A 2-D madaline adaptation algorithm with $z_1 = z_2 = 0$ will be called the *deadbeat* one. Let $A = aI$, $B = bI$ and $C = cI$. Then

$$\begin{aligned}\det\Big[Iz_1z_2 + (A-I)z_1 + (B-I)z_2 + I + C\Big] \\ = \Big[z_1z_2 + (a-1)z_1 + (b-1)z_2 + 1 + c\Big]^n = 0\end{aligned} \tag{5}$$

It is easy to see that all the roots of eqn. (5) are located in the unit bidisk if for example $a = b = 1$ and $|c+1| < 1$.

3. 2-D Two-layer FNN Weight Adaptation Algorithm

A block diagram of the 2-D two-layer FNN weight adaptation algorithm is shown in Fig. 2. The input vector $x \in \mathbb{R}^n$ is constant (independent of i and j), $W_{ij}^1 \in \mathbb{R}^{n_1 \times n}$ and $W_{ij}^2 \in \mathbb{R}^{n_2 \times n_1}$ are weight matrices with real entries depending on i and j ($i, j \in \mathbb{Z}_+$), $z_{ij} \in \mathbb{R}^{n_1}$, $v_{ij} \in \mathbb{R}^{n_1}$ are real vectors, $\Gamma : \mathbb{R}^{n_1} \to \mathbb{R}^{n_1}$ is an activation operator satisfying the condition $\Gamma(-z_{ij}) = -\Gamma(z_{ij})$ for all $i, j \in \mathbb{Z}_+$ and $y_d \in \mathbb{R}^{n_2}$, $y_{ij} \in \mathbb{R}^{n_2}$ are the desired output vector and the current output vector at the point (i, j), respectively.

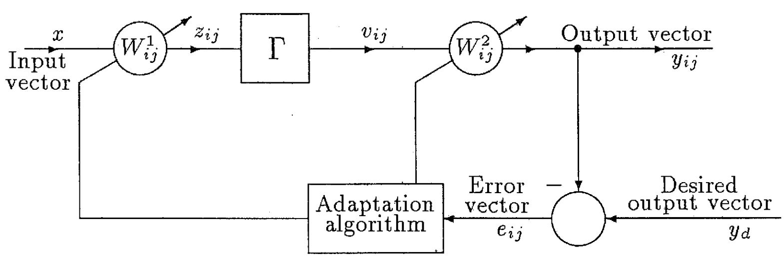

Fig. 2.

Let

$$e_{ij} \doteq y_d - y_{ij} \tag{6}$$

be the error at the point (i, j) and θ_1 and θ_2 be some non-linear operators.

Theorem 2. *If the weight matrices W_{ij}^1, W_{ij}^2 of a 2-D two-layer FNN are adapted according to the rule*

$$W_{i+1,j+1}^k = W_{i+1,j}^k + W_{i,j+1}^k + W_{ij}^k + \Delta W_{i+1,j}^k + \Delta W_{i,j+1}^k + \Delta W_{ij}^k, \quad k = 1, 2 \tag{7}$$

where

$$W_{i+1,j}^1 = W_{i,j+1}^1 = -W_{ij}^1$$

$$\Delta W^1_{i+1,j} = \Delta W^1_{i,j+1} = -\Delta W^1_{ij} = \begin{cases} \dfrac{2z_{ij}\theta_1^T(x)}{\theta_1^T(x)x} & if \quad \theta_1^T(x)x \neq 0 \\ 0 & if \quad \theta_1^T(x)x = 0 \end{cases} \tag{8}$$

$$\Delta W^2_{i+1,j} = \begin{cases} -2W^2_{i+1,j} - \dfrac{Ae_{i+1,j}\theta_2^T(\nu_{ij})}{\theta_2^T(\nu_{ij})\nu_{ij}} & if \quad \theta_2^T(\nu_{ij})\nu_{ij} \neq 0 \\ 0 & if \quad \theta_2^T(\nu_{ij})\nu_{ij} = 0 \end{cases}$$

$$\Delta W^2_{i,j+1} = \begin{cases} -2W^2_{i,j+1} - \dfrac{Be_{i,j+1}\theta_2^T(\nu_{ij})}{\theta_2^T(\nu_{ij})\nu_{ij}} & if \quad \theta_2^T(\nu_{ij})\nu_{ij} \neq 0 \\ 0 & if \quad \theta_2^T(\nu_{ij})\nu_{ij} = 0 \end{cases} \tag{9}$$

$$\Delta W^2_{ij} = \begin{cases} -2W^2_{i,j} - \dfrac{Ce_{i,j}\theta_2^T(\nu_{ij})}{\theta_2^T(\nu_{ij})\nu_{ij}} & if \quad \theta_2^T(\nu_{ij})\nu_{ij} \neq 0 \\ 0 & if \quad \theta_2^T(\nu_{ij})\nu_{ij} = 0 \end{cases}$$

and the matrices A, B, C are chosen so that all the roots (zero manifolds) of the equation

$$\det\Big[Iz_1z_2 - (I+A)z_1 - (I+B)z_2 + I - C\Big] = 0 \tag{10}$$

are located in the unit bidisk, then the error vector e_{ij} converges asymptotically to the zero vector for i and j tending to infinity.

Proof. Using (6), (7) and $z_{ij} = W^1_{ij}x$, $\nu_{ij} = \Gamma(z_{ij})$, $y_{ij} = W^2_{ij}\nu_{ij}$ we obtain

$$\begin{aligned} e_{i+1,j+1} - e_{i+1,j} - e_{i,j+1} + e_{ij} &= -y_{i+1,j+1} + y_{i+1,j} + y_{i,j+1} - y_{ij} \\ &= -W^2_{i+1,j+1}\nu_{i+1,j+1} + W^2_{i+1,j}\nu_{i+1,j} + W^2_{i,j+1}\nu_{i,j+1} - W^2_{ij}\nu_{ij} \\ &= -W^2_{i+1,j}\Big[\Gamma(z_{i+1,j+1}) - \Gamma(z_{i+1,j})\Big] - W^2_{i,j+1}\Big[\Gamma(z_{i+1,j+1}) - \Gamma(z_{i,j+1})\Big] \\ &\quad -W^2_{ij}\Big[\Gamma(z_{i+1,j+1}) + \Gamma(z_{ij})\Big] - \Delta W^2_{i+1,j}\Gamma(z_{i+1,j+1}) \\ &\quad -\Delta W^2_{i,j+1}\Gamma(z_{i+1,j+1}) - \Delta W^2_{ij}\Gamma(z_{i+1,j+1}) \end{aligned}$$

From eqn. (8) it follows that

$$z_{i+1,j} = z_{i,j+1} = -z_{ij}$$

and using (7) we have

$$\begin{aligned} z_{i+1,j+1} = W^1_{i+1,j+1}x &= z_{i+1,j} + z_{i,j+1} + z_{ij} + \Delta W^1_{i+1,j}x + \Delta W^1_{i,j+1}x + \Delta W^1_{ij}x \\ &= z_{i+1,j} + \Delta W^1_{i+1,j}x = z_{i+1,j} - \frac{2z_{i+1,j}\theta_1^T(x)x}{\theta_1^T(x)x} = -z_{i+1,j} \end{aligned} \tag{11}$$

Taking into account expression (11), the assumption $\Gamma(-z_{ij}) = -\Gamma(z_{ij})$, and eqn. (9) we obtain

$$\begin{aligned} &e_{i+1,j+1} - e_{i+1,j} - e_{i,j+1} + e_{ij} \\ &= 2\Big(W^2_{i+1,j}\Gamma(z_{i+1,j}) + W^2_{i,j+1}\Gamma(z_{i,j+1}) - W^2_{ij}\Gamma(z_{ij})\Big) \\ &-\Big(\Delta W^2_{i+1,j} + \Delta W^2_{i,j+1} + \Delta W^2_{ij}\Big)\nu_{ij} = Ae_{i+1,j} + Be_{i,j+1} + Ce_{ij} \end{aligned}$$

and

$$e_{i+1,j+1} = (I+A)e_{i+1,j} + (I+B)e_{i,j+1} + (C-I)e_{ij} \tag{12}$$

If all the roots of eqn. (10) are located in the unit bidisk, then the vector e_{ij} converges asymptotically to 0 for i and j tending to infinity (Ahmed 1980; Fernando and Nicholson, 1985; Fornasini and Marchesini, 1980; Kaczorek, 1985). ■

4. Concluding Remarks

The generalized weight adaptation algorithms have been extended for 2-D madaline and 2-D two-layer FNN's. Theorem 2 can be extended for 2-D three-layer FNN and generally for 2-D n-layer FNN's $(n > 2)$. Note that for the adaptation algorithm (7) another choice of $\Delta W^k_{i+1,j}$, $\Delta W^k_{i,j+1}$, and ΔW^k_{ij} for $k = 1, 2$ can also guarantee the convergence of e_{ij} for $i, j \to \infty$.

With slight modifications the above considerations can be extended for n-D, madaline $(n > 2)$ and n-D many-layers FNN's.

References

Ahmed A.R.E. (1980): *On the stability of two-dimensional discrete systems.* — IEEE Trans. Automat. Control, v.AC-25, No.3, pp.551–552.

Chen F. (1990): *Back-propagation neural networks for non-linear self-tuning adaptive control.* — IEEE Contr. Syst. Mag., v.10, No.3, pp.44–48.

Fernando K.V. and Nicholson H. (1985): *Stability assessment of two-dimensional state-space systems.* — Trans. Automat. Control, v.CAS-32, No.5, pp.484–487.

Fornasini E. and Marchesini G. (1980): *Stability of 2-D systems.* — IEEE Trans. Circuits System, v.CAS-127, No.12, pp.1210–1217.

Kaczorek T. (1985): *Two-Dimensional Linear Systems.* — Berlin: Springer-Verlag.

Kuschewski J.G., Hui S. and Żak S.H. (1993): *Application of feedforward neural networks to dynamical system identification and control.* — IEEE Trans. Control Systems Technology, v.1, No.1, pp.37–49.

Pandolfi L. (1984): *Exponential stability of 2-D systems.* — Systems & Control Letters, v.4, pp.381–385.

Psaltis D., Sideris A. and Yamamura A.A. (1988): *A multilayered neural network controller.* — IEEE Contr. Syst. Mag., v.8, No.2, pp.17–21.

IJC (1992): *Special issue on intelligent control.* — Int. J. Contr., v.56, No.2.

Widrow B. and Lehr M.A. (1990): *30 years of adaptive neural networks: perceptron, madaline, and back-propagation.* — Proc. IEEE, v.78, No.9, pp.1415–1442.

Żak S.H. and Sira-Ramirez H.J. (1990): *On the adaptation algorithms for feedforward neural networks,* In: Theoretical Aspects of Industrial Design (Field D.A. and Komkov V., Eds.). — New York: SIAM, pp.116–133.