

Some Control Theoretic Issues in Neural Networks*

James W. Howse

Chaouki T. Abdallah

Gregory L. Heileman

Department of Electrical and Computer Engineering
University of New Mexico
Albuquerque, NM 87131

ABSTRACT

We have observed that many neural network models can be written as a bilinear system with a specific form of nonlinear state-to-input feedback. This framework includes the ART architecture among others. There are two significant results which follow from this observation. First, the parameters of the model determine the controllability of the system. A system is controllable if there exists some input which transfers any initial state to any desired final state in a *finite* time. If for a given set of these parameters the system is *not* controllable, then there are regions of the state space which the system can *never* enter in a finite time for *any* input. Because of this restriction the learning ability of the system may be severely limited. Second, the multiplicative equation is *linear* in all of the parameters, and all of the adjustable weights. This means that a provably convergent learning algorithm can be devised for *all* of these quantities. This does not however circumvent the learning limitation since the learning algorithm is *not* guaranteed to converge in a finite time. In the paper, we will study these issues as they apply to the ART architecture.

1. Introduction

It was observed in [5] that many neural network models can be written as either a bilinear or a linear system with a specific form of nonlinear state-to-input feedback. Among other networks, this framework includes the ART architecture introduced in [2]. There are two significant results which follow from the this observation. First, the parameters of the model determine the controllability of the system. A system is controllable if there exists some input which transfers any initial state to any desired final state in a *finite* time. If for a given set of parameters the system is *not* controllable, then there are regions of the state space which the system can *never* enter in a finite time for *any* input. Because of this restriction the learning ability of the system may be severely limited. Intuitively, this is because the system can not follow a training signal into these portions of the state space in a finite time, no matter what training examples are given. Second, the dynamic equations for ART are *linear* in all of the parameters, and all of the adjustable weights. This means that a convergent learning algorithm similar to that presented in [6] can be constructed for finding *all* of these quantities, provided that a bounded input leads to a bounded state. This does not circumvent the learning limitation since this learning algorithm is *not* guaranteed to converge in a finite time. In Section 2 of this paper the dynamics of the ART model are discussed. Then in Section 3 a general ART network is cast into a bilinear form. Lastly, Section 4 discusses the controllability of the ART network.

2. Dynamics in the ART Network

In [2] a neural network architecture is defined which came to be called ART1. A general form for the node activation and weight update dynamics for an n_s node version of this system is

$$\frac{1}{\epsilon_i} \dot{x}_i = -\mathcal{A}_i x_i + (\mathcal{B}_i - \mathcal{D}_i x_i) \left[I_i^+(t) + \sum_{j=1}^{n_s} s_{ij}^+ w_{ij}^+ f_j^+(x_j) \right] - (\mathcal{C}_i + \mathcal{E}_i x_i) \left[I_i^-(t) + \sum_{j=1}^{n_s} s_{ij}^- w_{ij}^- f_j^-(x_j) \right], \quad (1a)$$

*This paper will be presented in an invited session on ART at the *International Conference on Neural Networks* in June, 1996.

$$\dot{w}_{ij}^+ = -\mathcal{G}_{ij} f_i^+(x_i) w_{ij}^+ + \mathcal{H}_{ij} f_i^+(x_i) f_j^+(x_j), \quad i, j \in \{1, \dots, n_s\}. \quad (1b)$$

Equation (1a) defines the node activation dynamics at *each* of the n_s nodes in the network, while Equation (1b) defines the weight update dynamics for the excitatory weights. Note that this network is *not* fully connected, nor are all of the weights adaptive.

In Equation (1a) the term $-\mathcal{A}_i x_i$ is a passive decay which causes x_i to go to zero if the other two terms are zero, and where the constant \mathcal{A}_i determines the rate of decay. The term $(\mathcal{B}_i - \mathcal{D}_i x_i) [I_i^+(t) + \sum_{j=1}^{n_s} s_{ij}^+ w_{ij}^+ f_j^+(x_j)]$ is the positive (e.g., excitatory) feedback which tries to increase x_i . Finally $(\mathcal{C}_i + \mathcal{E}_i x_i) [I_i^-(t) + \sum_{j=1}^{n_s} s_{ij}^- w_{ij}^- f_j^-(x_j)]$ is the negative (e.g., inhibitory) feedback which tries to decrease x_i . These descriptions depend upon x_i , $f_i^+(x_i)$ and $f_i^-(x_i)$ and all of the parameters and weights being non-negative. If that is not the case, then the behavior of the terms will be the reverse of their previous descriptions. The excitatory and inhibitory connection weights *to* the i th node *from* the j th node are given by w_{ij}^+ and w_{ij}^- respectively. All of the connection weights incident to a specific node do not have to be given equal consideration. For instance, the connections from nodes that are physically closer to the given node may be considered more important than those from nodes which are farther away. If all of the connection weights into a node are viewed as a field of values in weight space, then this field is sampled by some function. The sample values applied to the excitatory and inhibitory connections to the i th node from the j th node are given by s_{ij}^+ and s_{ij}^- respectively. The excitatory and inhibitory external inputs are given by $I_i^+(t)$ and $I_i^-(t)$ respectively. The terms $-\mathcal{D}_i x_i$ and $-\mathcal{E}_i x_i$ have the effect of forcing the activation levels \mathbf{x} of all of the nodes to belong to the closed and bounded set $\mathcal{X} = \{x_i \in \mathbb{R} : -\frac{\mathcal{C}_i}{\mathcal{E}_i} < x_i < \frac{\mathcal{B}_i}{\mathcal{D}_i} \ \forall \ i = 1, \dots, n_s\}$. Note that this bound is only valid if the initial values of all the activations $\mathbf{x}(t_0)$ are contained in the set \mathcal{X} . Also the time constants, which determine the speed at which x_i approaches these upper and lower bounds, are given by $\frac{1}{\mathcal{D}_i}$ and $\frac{1}{\mathcal{E}_i}$ respectively.

Equation (1b) defines the dynamics of the excitatory connection weights. The term $-\mathcal{G}_{ij} w_{ij}^+$ is a passive decay term where \mathcal{G}_{ij} is a constant which determines the decay rate. The constant \mathcal{H}_{ij} determines the growth rate of the connection weight w_{ij}^+ if the nodes at both ends of the connection are active. Notice that under this learning rule a weight can not decay unless the node which the connection is incident *to* has a non-zero output. Also notice that the equilibrium value of a weight under this rule is the output value of the node that the weight is incident from. The activation dynamics in Equation (1a) are called *multiplicative* dynamics. An extensive study of these dynamics, when the connection weights are constant, was begun in [3]. The weight dynamics in Equation (1b) are called *gated* dynamics, and some reasons for selecting such dynamics are discussed in [4].

3. ART as a Bilinear System

It is instructive to look at the system defined in Equation (1) from the viewpoint of control theory. Consider an ART network with n_1 nodes in the first layer (i.e., the *F1* layer) and n_2 nodes in the first layer (i.e., the *F2* layer). Define the sum and product of these two numbers as $n_s = n_1 + n_2$ and $n_p = n_1 n_2$ respectively. Note that such an ART network contains $2 n_p$ adjustable connection weights w_{ij}^+ , and has a total of $n_t = n_s + 2 n_p$ elements in the state vector $\mathbf{z} \equiv [\mathbf{x}_{n_s \times 1} \parallel \mathbf{w}_{2n_p \times 1}^+]^\dagger$. A control diagram of a general ART network appears in Figure 1. This figure makes it clear that an ART network is a bilinear system with a specific type of nonlinear state-to-input feedback. The input $\mathbf{v}(t)$ is a $(3 n_s + n_p)$ element vector consisting of $\mathbf{v}(t) \equiv [\mathbf{I}_{n_s \times 1}^+ \mid \mathbf{I}_{n_s \times 1}^- \parallel \mathbf{0}_{n_s \times 1} \mid \mathbf{0}_{n_p \times 1}]^\dagger$. Note that in the standard ART model, the inhibitory external input \mathbf{I}^- is zero. The nonlinear feedback function $\mathbf{f}(\mathbf{x})$ is a $(2 n_s + n_p)$ element vector consisting of $\mathbf{f}(\cdot) \equiv [\mathbf{f}_{n_s \times 1}^+ \mid \mathbf{f}_{n_s \times 1}^- \parallel \mathbf{p}_{n_p \times 1}^+]^\dagger$. In this vector, the last term contains the appropriate elements of the upper triangular part of the outer product of \mathbf{f}^+ with itself, given by $p_i^+ = f_j^+ f_k^+$, where $i = 1, \dots, n_p$, $j = 1, \dots, n_1$, and $k = 1, \dots, n_2$. The symmetry of this outer product is used to halve the number of terms needed in \mathbf{p}^+ . Note that in the standard ART model, the excitatory node output function \mathbf{f}^+ and the inhibitory node output function \mathbf{f}^- are the identical.

In the following discussion a bold calligraphic letter is used to denote a *diagonal* matrix with the corresponding constants along the diagonal (e.g., the matrix \mathcal{H}^u is a diagonal matrix containing the constants \mathcal{H}_{ij} ,

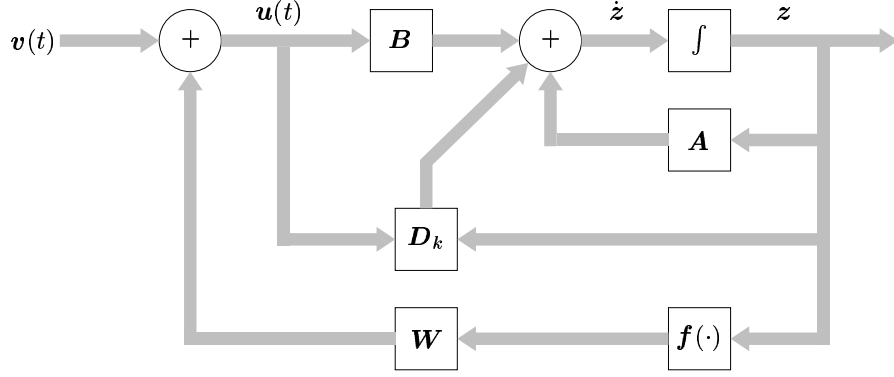


Fig. 1: A control theory diagram of the general ART system defined in Equation (1). This system is a $(3n_s + n_p)$ -input n_t -output bilinear system with a specific form of nonlinear state-to-input feedback.

$j > i$ along the diagonal). The matrix \mathbf{A} is an $(n_t \times n_t)$ matrix of the form $\mathbf{A} \equiv \left(\begin{array}{c|c} -\mathbf{A}_{n_s \times n_s} & \mathbf{O}_{n_s \times 2n_p} \\ \hline \mathbf{O}_{2n_p \times n_s} & \mathbf{O}_{2n_p \times 2n_p} \end{array} \right)$. The matrix \mathbf{B} is $(n_t \times 3n_s + n_p)$, and the matrix \mathbf{W} is $(3n_s + n_p \times 2n_s + n_p)$. They are given by

$$\mathbf{B} \equiv \left(\begin{array}{c|c|c} \mathbf{B}_{n_s \times n_s} & -\mathbf{C}_{n_s \times n_s} & \mathbf{O}_{n_s \times n_s + n_p} \\ \hline \mathbf{O}_{2n_p \times 2n_s} & \mathbf{O}_{n_p \times n_s} & \mathcal{H}_{n_p \times n_p}^u \\ \hline & \mathbf{O}_{n_p \times n_s} & \mathcal{H}_{n_p \times n_p}^l \end{array} \right) \quad \mathbf{W} \equiv \left(\begin{array}{c|c|c} (\mathbf{S}^+ \circ \mathbf{W}^+)_{n_s \times n_s} & \mathbf{O}_{n_s \times n_s} & \mathbf{O}_{2n_s \times n_p} \\ \hline \mathbf{O}_{n_s \times n_s} & -(\mathbf{S}^- \circ \mathbf{W}^-)_{n_s \times n_s} & \mathbf{O}_{2n_s \times n_p} \\ \hline \mathcal{I}_{n_s \times n_s} & \mathbf{O}_{n_s \times n_s} & \mathbf{O}_{n_s \times n_p} \\ \hline & \mathbf{O}_{n_p \times 2n_s} & \mathcal{I}_{n_p \times n_p} \end{array} \right) \quad (2)$$

where the operation \circ denotes the *Schur product* which is defined as $[\mathbf{A} \circ \mathbf{B}]_{ij} = a_{ij}b_{ij}$, and where \mathcal{I} denotes the identity matrix. Lastly consider the $(n_t \times n_t)$ matrices \mathbf{D}_k where $k = 1, \dots, 3n_s + n_p$. There are four cases that must be considered.

1. For $k = 1, \dots, n_s$, \mathbf{D}_k is the zero matrix with the single non-zero element $D_{kk} = -\mathcal{D}_k$.
2. For $k = n_s + 1, \dots, 2n_s$, \mathbf{D}_k is the zero matrix with the single non-zero element $D_{(k-n_s)(k-n_s)} = -\mathcal{E}_k$.
3. For $k = 2n_s + 1, \dots, 3n_s$, there are two sub-cases.
 - (a) For $k = 2n_s + 1, \dots, (2n_s + 1) + (n_1 - 1)$, \mathbf{D}_k is the zero matrix with the n_2 non-zero elements $D_{(2k-(3n_s+1)+j)(2k-(3n_s+1)+j)} = -\mathcal{G}_{(k-2n_s)((j+1)+n_1)}$, $j = 0, 1, \dots, n_2 - 1$.
 - (b) For $k = (2n_s + 1) + n_1, \dots, 3n_s$, \mathbf{D}_k is the zero matrix with the n_1 non-zero elements $D_{(k-(2n_s-n_2-n_p)+n_2j)(k-(2n_s-n_2-n_p)+n_2j)} = -\mathcal{G}_{(k-2n_s)(j+1)}$, $j = 0, 1, \dots, n_1 - 1$.
4. For $k = 3n_s + 1, \dots, 3n_s + n_p$, \mathbf{D}_k is the zero matrix.

Note that all non-zero elements in \mathbf{D}_k occur on the diagonal regardless of the value of k . So this is a $(3n_s + n_p)$ -input n_t -output system, where the inputs are the external excitatory and inhibitory signals. The dynamics of the system illustrated in Figure 1 are

$$\dot{\mathbf{z}} = \mathbf{A}\mathbf{z} + \mathbf{B}(\mathbf{v}(t) + \mathbf{W}\mathbf{f}(\mathbf{z})) + \sum_{k=1}^{3n_s+n_p} \mathbf{D}_k(\mathbf{v}(t) + \mathbf{W}\mathbf{f}(\mathbf{z}))_k \mathbf{z}, \quad (3)$$

where $(\mathbf{v}(t) + \mathbf{W}\mathbf{f}(\mathbf{x}))_k$ denotes the k th element of this $(3n_s + n_p)$ -dimensional vector.

4. Controllability of ART

Although there are several definitions of controllability, the one used in this paper is the following. A system $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u})$ is *controllable* if there exists some input $\tilde{\mathbf{u}}$ which transfers any initial state $\tilde{\mathbf{x}}_i$ to any final state $\tilde{\mathbf{x}}_f$ in a finite time. A well known but often unstated fact is that for *any* system, the controllability is *not*

changed by *any* form of state-to-input feedback. This can be seen by considering the input in the previous system to be $\mathbf{u} = \mathbf{v} + \mathbf{g}(\mathbf{x})$, where \mathbf{v} is a new external input. Suppose that the system without feedback (i.e., $\mathbf{u} = \mathbf{v}$) is controllable. Then for every pair of points $(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_f)$ there exists an input $\tilde{\mathbf{u}}$ which transfers the system between these two points. Now consider the system with feedback. Clearly the input $\mathbf{v} = \tilde{\mathbf{u}} - \mathbf{g}(\mathbf{x})$ will transfer the feedback system between every pair of points $(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_f)$. Therefore a controllable system can not be made uncontrollable using state-to-input feedback. Now suppose that the system without feedback is uncontrollable. Then there exists at least one pair of points $(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_f)$ which no input can transfer the system between in a finite time. Suppose that for the feedback system some input $\tilde{\mathbf{v}}$ exists which will transfer the system between $\tilde{\mathbf{x}}_i$ and $\tilde{\mathbf{x}}_f$. In that case there must exist an input $\mathbf{u} = \tilde{\mathbf{v}} + \mathbf{g}(\mathbf{x})$ which transfers the system without feedback between these two points, which is a contradiction. Therefore an uncontrollable system can not be made controllable using state-to-input feedback.

The importance of this result in determining the controllability of the ART network is that rather than having to determine the controllability of the entire nonlinear system, which can be quite difficult, only the controllability of the bilinear part of the system must be determined. Referring to Figure 1, the branch containing $\tilde{\mathbf{W}}$ and $\tilde{\mathbf{f}}(\cdot)$ can be neglected when analyzing the controllability. So the interconnections between the nodes have *no effect* on the controllability of this network, it is determined solely by the activation dynamics without any of the terms containing the connection weights w_{ij} . This bilinear portion of the ART network has two properties that make determining the controllability quite easy. First, each of the n_t state components only receives feedback from itself, all other feedback is provided by the loop containing $\tilde{\mathbf{W}}$ and $\tilde{\mathbf{f}}(\cdot)$. Second, only the n_1 state components represented by the nodes in the *F1* layer receive any input from outside the network. Collectively these observations imply that each state component of the bilinear part of the network acts as an independent system, and that overall this part *can not* be controllable because some (in fact most) of the state elements have neither an external input, nor any interaction with other state elements. This implies that the entire ART network is uncontrollable.

A formal analysis of the controllability of the ART network can be conducted using either the method discussed in [8] or the one in [1]. We will use the method in [8] because it minimizes the need for Lie algebra. In [7] one of the necessary conditions given for a bilinear system to be controllable is that the matrix \mathbf{A} must have distinct non-zero eigenvalues. Clearly this criteria is not satisfied by the ART network since the \mathbf{A} matrix has $2n_p$ zero eigenvalues. This leaves open the question of whether the subspace consisting of the only the node activations \mathbf{x} is controllable. In order to answer this question, the following theorem, paraphrased from [8], will be used. Note that the conditions of this theorem are sufficient but not necessary.

Theorem 4.1 ([8]). *Consider the class of all piecewise-continuous inputs $\mathbf{u}(t)$, whose domain is $[0, \infty)$ and whose range \mathcal{U} is a compact connected set containing the origin. The bilinear system*

$$\dot{\mathbf{x}} = \tilde{\mathbf{A}}\mathbf{x} + \tilde{\mathbf{B}}(\tilde{\mathbf{v}}(t) + \tilde{\mathbf{W}}\tilde{\mathbf{f}}(\mathbf{x})) + \sum_{k=1}^{2n_s} \tilde{\mathbf{D}}_k(\tilde{\mathbf{v}}(t) + \tilde{\mathbf{W}}\tilde{\mathbf{f}}(\mathbf{x}))_k \mathbf{x}, \quad (4)$$

where $\mathbf{u}(t) = \tilde{\mathbf{v}}(t) + \tilde{\mathbf{W}}\tilde{\mathbf{f}}(\mathbf{x})$ is completely controllable with respect to this class of inputs if

- 1) there exist constant control values $\tilde{\mathbf{u}}^+$ and $\tilde{\mathbf{u}}^-$ in \mathcal{U} such that the real parts of the eigenvalues of $\mathbf{R} \equiv \tilde{\mathbf{A}} + \sum_{k=1}^{2n_s} \tilde{\mathbf{D}}_k \tilde{\mathbf{u}}_k$ are positive and negative for $\tilde{\mathbf{u}} = \tilde{\mathbf{u}}^+$ and $\tilde{\mathbf{u}} = \tilde{\mathbf{u}}^-$ respectively;
- 2) the equilibrium points $\mathbf{x}^e(\tilde{\mathbf{u}}) = -\mathbf{R}^{-1}\tilde{\mathbf{B}}\tilde{\mathbf{u}}$ corresponding to $\tilde{\mathbf{u}} = \tilde{\mathbf{u}}^+$ and $\tilde{\mathbf{u}} = \tilde{\mathbf{u}}^-$ are contained in a connected subset of the equilibrium set;
- 3) for each control $\tilde{\mathbf{u}}$ and the matrices \mathbf{R} and $\mathbf{T} \equiv \tilde{\mathbf{B}} + [\tilde{\mathbf{D}}_1 \mathbf{x}^e \mid \tilde{\mathbf{D}}_2 \mathbf{x}^e \mid \cdots \mid \tilde{\mathbf{D}}_{2n_s} \mathbf{x}^e]$, there exists a $\mathbf{y} \in \mathbb{R}^{2n_s}$ such that the n_s vectors $\mathbf{T}\mathbf{y}$, $\mathbf{R}\mathbf{T}\mathbf{y}$, \dots , $\mathbf{R}^{n_s-1}\mathbf{T}\mathbf{y}$ are linearly independent.

Note that in Equation (4), $\tilde{\mathbf{v}}(t)$ and $\tilde{\mathbf{f}}(\cdot)$ are the vectors obtained by using only the elements to the *left* of the double bar $\|$ in $\mathbf{v}(t)$ and $\mathbf{f}(\cdot)$ respectively. Furthermore, the matrices $\tilde{\mathbf{A}}$, $\tilde{\mathbf{B}}$, and $\tilde{\mathbf{W}}$ are obtained by taking those elements that are above and left of the double bars in \mathbf{A} , \mathbf{B} , and \mathbf{W} . Lastly the matrices $\tilde{\mathbf{D}}_k$ are $(n_s \times n_s)$ and are constructed using Cases 1 and 2 of \mathbf{D}_k .

First we investigate the conditions on the network needed to satisfy Condition 1. It can be shown that for the ART network (or any system whose activation dynamics are given by Equation (1a)) the matrix $\mathbf{R} =$

$[-\text{Diag}((\mathcal{A}_1 + \mathcal{D}_1(I_1^+ + \mathbf{w}_1^+ \mathbf{f}^+(\mathbf{x}^e)) + \mathcal{E}_1(I_1^- + \mathbf{w}_1^- \mathbf{f}^-(\mathbf{x}^e))), \dots, (\mathcal{A}_{n_1} + \mathcal{D}_{n_1}(I_{n_1}^+ + \mathbf{w}_{n_1}^+ \mathbf{f}^+(\mathbf{x}^e)) + \mathcal{E}_{n_1}(I_{n_1}^- + \mathbf{w}_{n_1}^- \mathbf{f}^-(\mathbf{x}^e))), (\mathcal{A}_{n_1+1} + \mathcal{D}_{n_1+1}(\mathbf{w}_{n_1+1}^+ \mathbf{f}^+(\mathbf{x}^e)) + \mathcal{E}_{n_1+1}(\mathbf{w}_{n_1+1}^- \mathbf{f}^-(\mathbf{x}^e))), \dots, (\mathcal{A}_{n_2} + \mathcal{D}_{n_2}(\mathbf{w}_{n_2}^+ \mathbf{f}^+(\mathbf{x}^e)) + \mathcal{E}_{n_2}(\mathbf{w}_{n_2}^- \mathbf{f}^-(\mathbf{x}^e)))]$, which is an $(n_s \times n_s)$ diagonal matrix. Because this matrix is diagonal, its eigenvalues are merely the values of the diagonal elements. If it is assumed that \mathcal{A}_i , \mathcal{D}_i , and \mathcal{E}_i are non-negative for all values of i and that all of the elements of \mathbf{W}^+ and \mathbf{W}^- are non-negative, then the only way for all the eigenvalues of \mathbf{R} to change sign is for *all* elements of $\mathbf{f}^+(\cdot)$ and/or $\mathbf{f}^-(\cdot)$ to return both positive and negative values. For example the function $f_i^+(x_i) = \tanh(\mathcal{M}_i x_i)$ satisfies this criteria, whereas $f_i^+(x_i) = \frac{1}{1+e^{-\mathcal{M}_i x_i}}$ does not. Since all of the constants are assumed to be positive, $f_i^+(x_i)$ and $f_i^-(x_i)$ must become sufficiently negative in order for the diagonal terms to be positive. If $\mathbf{f}^+ = \mathbf{f}^- \equiv \mathbf{f}$, one sufficient condition is $\min(f_i(x_i)) \leq -\frac{\mathcal{A}_{max}}{n_s(\mathcal{D}_{min} w_{min}^+ + \mathcal{E}_{min} w_{min}^-)}$. In this expression $w_{min}^+ = \min_{i=1, \dots, n_s} (\sum_{j=1}^{n_s} w_{ij}^+)$, in other words take the sum of the elements in each row of \mathbf{W}^+ and choose the smallest sum.

Next, examine the restrictions needed to satisfy Condition 3. It can be shown that $\mathbf{T} = [\text{Diag}(\mathcal{B}_1 - \mathcal{D}_1 x_1^e, \dots, \mathcal{B}_{n_s} - \mathcal{D}_{n_s} x_{n_s}^e) \mid -\text{Diag}(\mathcal{C}_1 + \mathcal{E}_1 x_1^e, \dots, \mathcal{C}_{n_s} + \mathcal{E}_{n_s} x_{n_s}^e)]$, which is an $(n_s \times 2n_s)$ block diagonal matrix. Because \mathbf{R} is diagonal, there always exists a $\mathbf{y} \in \mathbb{R}^{2n_s}$ such that the vectors $\mathbf{T}\mathbf{y}$, $\mathbf{R}\mathbf{T}\mathbf{y}$, \dots , $\mathbf{R}^{n_s-1}\mathbf{T}\mathbf{y}$ are linearly independent, unless *all* the diagonal elements of \mathbf{R} are equal. In this case let $R_{ii} = \mathcal{K}$ for all $i = 1, \dots, n_s$ where $\mathcal{K} \in \mathbb{R}$. It can be shown that for a given value of \mathcal{K} there is a *unique* input vector $\hat{\mathbf{u}}$ which makes this condition true. Since $\hat{\mathbf{u}}$ depends on \mathcal{K} and since \mathcal{K} is arbitrary, some of these inputs will *always* fall within the range \mathcal{U} specified for the inputs \mathbf{u} . This means that at best there will always be a finite number of points at which the states are uncontrollable. Regarding Condition 2, we believe that an argument analogous to that used in [8] can be applied to show that the equilibrium set is connected for this system.

5. Conclusion

We have shown that the ART network presented in [2] is *uncontrollable* as a dynamical system. From a learning point of view this means that the system can not follow a training signal into certain regions of the state space in a finite time, no matter what sort of training examples are given. We show that choosing node output functions which return both positive and negative values allows the node activation subspace for the ART network to be controllable. In addition we provide a least upper bound for the negative values returned by the node output functions. One point should be noted here. Controllability generally assumes that the input can assume any value in its range at any given time. In neural networks the only inputs to nodes in the hidden or output layers generally depend only on the outputs of other nodes. So in this scenario it is not clear that these node inputs can take an arbitrary value at a given time. Another unresolved issue is the behavior of the uncontrollable region when the number of hidden layer nodes is increased. This would address the issue of whether increasing the number of nodes allows an uncontrollable network to accurately learn an actual system. We also observe that the dynamic equations for ART are linear in all of the parameters, and all of the adjustable weights. This means that a convergent learning algorithm similar to that presented in [6] can be constructed for finding all of these quantities.

Acknowledgments

This research was supported by a grant from Boeing Computer Services under Contract W-300445. The authors would like to thank Vangelis Coutsias, Tom Caudell, Don Hush, and Bill Horne for stimulating discussions and insightful suggestions.

References

- [1] R. Brockett, "System theory on group manifolds and coset spaces," *SIAM Journal on Control*, vol. 10, no. 2, pp. 265-284, 1972.
- [2] G. Carpenter and S. Grossberg, "A massively parallel architecture for a self-organizing neural pattern recognition machine," *Computer Vision, Graphics, and Image Processing*, vol. 37, no. 1, pp. 54-115, 1987.

- [3] S. Grossberg, "Contour enhancement, short term memory, and constancies in reverberating neural networks," *Studies in Applied Mathematics*, vol. 52, no. 3, pp. 213–257, 1973.
- [4] S. Grossberg, "Nonlinear neural networks: Principles, mechanisms, and architectures," *Neural Networks*, vol. 1, no. 1, pp. 17–61, 1988.
- [5] J. Howse, *Gradient and Hamiltonian dynamics: Some applications to neural network analysis and system identification*. PhD thesis, University of New Mexico, Department of Electrical Engineering, December 1995.
- [6] J. Howse, C. Abdallah, and G. Heileman, "Gradient and Hamiltonian dynamics applied to learning in neural networks," in *Advances in Neural Information Processing Systems: Proceedings of the 1995 Conference*, (D. Touretzky, M. Mozer, and M. Hasselmo, eds.), The MIT Press, 1996. To Appear.
- [7] R. Mohler, *Nonlinear systems: Applications to bilinear systems*. Vol. II, Englewood Cliffs, NJ: Prentice-Hall, Inc., 1991.
- [8] R. Rink and R. Mohler, "Completely controllable bilinear systems," *SIAM Journal on Control*, vol. 6, no. 3, pp. 477–486, 1968.