Location and Stability of the High-Gain Equilibria of Nonlinear Neural Networks

Mathukumalli Vidyasagar, *Fellow, IEEE*

***Abstract*—This paper analyzes the number, location, and stability behavior of the equilibria of arbitrary nonlinear neural networks without resorting to energy arguments based on assumptions of symmetric interactions or no self-interactions. The class of networks studied consists of very general continuous-time continuous-state (CTCS) networks that contain the standard Hopfield network as a special case. The emphasis is on the case where the slopes of the sigmoidal nonlinearities become larger and larger, i.e., the *high-gain limit.* The following results are proved: Let $H = (0,1)^n$ and $\bar{H} = [0,1]^n$ denote the open and closed n-dimensional hypercubes, respectively, on which the neural network evolves, and let I denote the (constant) vector of external inputs. Then, as the neural sigmoid characteristics become steeper and steeper, it is shown that the following statements are true for all I except for those belonging to a set of measure zero. 1) There are only finitely many equilibria in any compact subset of H. If there are no self-interactions, then these equilibria cannot be exponentially stable, and under mild conditions they are in fact unstable. If the network has symmetric (nonlinear) interactions, whether or not it has self-interactions, then the stable manifolds of all these equilibria have the same dimension, which can be computed explicitly. If the network also has no self-interactions, then all of these equilibria are unstable. 2) There are only finitely many equilibria in any face of H. If there are no self-interactions, then there are no equilibria in an edge of H. If the network has symmetric interactions, then the stable manifolds of equilibria in parallel faces of H have the same dimension, which can be computed explicitly. If the network also has no self-interactions, then all equilibria in the faces of H are unstable. 3) A systematic procedure is given for determining which corners of H contain equilibria, and it is shown that all equilibria in the corners of H are asymptotically stable. One corollary of the above results is that the standard Hopfield network can have asymptotically stable equilibria only in the corners of H, and trajectories starting at almost all initial conditions approach the corners of H. It is important to note that the proofs here are *not* based on energy arguments. As a result, these results are "hardy" in the sense that they continue to hold even if the network dynamics are slightly perturbed.**

I. Introduction

Recently there has been a great deal of interest in artificial neural networks, especially those of the Hopfield type. Two types of Hopfield networks are widely studied. *Discrete-time discrete-state* (DTDS) networks are described by

$$V_i^{t+1} = \text{sat}\left[\sum_{j=1}^{n} t_{ij}V_j^t + I_i\right], \qquad i = 1, \cdots, n \qquad (1.1)$$

where n is the number of neurons; V_i^t is the state of neuron i at time t, and equals either 0 or 1; I_i is the (constant) external input to neuron i; and t_{ij} is the interconnection weight. Here the "sat" function is defined by

$$\text{sat}(x) = \begin{cases} 0, & \text{if } x \le 0, \\ 1, & \text{if } x > 0. \end{cases} \qquad (1.2)$$

Such networks are studied by Hopfield [1]. He defines the *energy function* as

$$E_d(V) = -\sum_{i=1}^{n}\left[I_iV_i + \sum_{j=1}^{n}\frac{1}{2}t_{ij}V_iV_j\right] \qquad (1.3)$$

where $V = [V_1 \cdots V_n]^t \in \{0,1\}^n$ is the state vector of the neural network. He then shows that if $t_{ii} = 0 \ \forall i$ (no self-interactions) and $t_{ij} = t_{ji} \ \forall i, j$ (symmetric interactions), and *if the neural states are updated asynchronously,* then

$$E_d(V^{t+1}) \le E_d(V^t) \qquad (1.4)$$

In other words, the energy is nonincreasing as a function of time. Hence, in a finite number of time steps, the neural state vector V^t will reach a "one-flip minimum," i.e., a vector $V_0 \in \{0,1\}^n$ with the property that

$$E(V_0) \le E(V) \quad \text{whenever } H(V, V_0) = 1 \qquad (1.5)$$

where $H(V, V_0)$ denotes the *Hamming distance* between V and V_0, i.e., the number of components where V and V_0 differ.

The second class of neural networks that has been studied consists of continuous-time, continuous-state (CTCS) networks[1] described by

$$C_i\dot{u}_i = -\frac{1}{R_i}u_i + \sum_{j=1}^{n} t_{ij}v_j + I_i,$$
$$v_i = g_i(\lambda u_i), \quad i = 1 \cdots n \qquad (1.6)$$

where n is the number of neurons; v_i is the neural current and u_i is the neural voltage; I_i is the external current input to the ith neuron, C_i is the membrane capacitance, and R_i is the neural resistance; g_i is the characteristic of the ith neuron, λ is a scaling parameter, and t_{ij} is the interconnection weight. The function $g_i : \Re \to (0,1)$ is a so-called sigmoidal nonlinearity. In other words, g_i is continuously differentiable, strictly increasing, $g_i(x) \to 0$ as $x \to -\infty$, and $g_i(x) \to 1$ as $x \to \infty$. The role of the constant λ is to scale the input to the

Manuscript received April 3, 1991; revised August 8, 1992.
The author is with the Centre for Artificial Intelligence and Robotics, Raj Bhavan Circle, High Grounds, Bangladore, India.
IEEE Log Number 9203981.

[1] The other two possible classes of networks—continuous-time, discrete-state networks, and discrete-time continuous-state networks—have not received much attention.

1045-9227/93$03.00 © 1993 IEEE

sigmoidal nonlinearites. Note that, as $\lambda \to \infty$, the function $i \mapsto g_i(\lambda u)$ "approaches" the "sat" function of (1.2) in some loose sense.

Hopfield [2] studies such networks, under the assumptions that 1) $t_{ii} = 0$ for all i (no self-interactions), and 2) $t_{ij} = t_{ji}$ for all i, j (symmetric interactions). He defines the *energy function*

$$E_c = \sum_{i=1}^{n} \left[\frac{1}{\lambda R_i} \int_0^{v_i} g_i^{-1}(v)\, dv - I_i v_i - \sum_{j=1}^{n} \frac{1}{2} t_{ij} v_i v_j \right]. \tag{1.7}$$

Note that, as the scaling parameter λ approaches infinity, the energy function E_c approaches the energy function E_d. Hopfield argues that, for this reason, networks of the form (1.6) can also be used to minimize quadratic functions of the form (1.3) over the Boolean set $\{0,1\}^n$, provided that the scale factor λ is sufficiently large. The dynamics of networks of the form (1.6) are further analyzed in [3] and [4]. In [3], it is shown that $\dot{E}_c(V) \leq 0$ for all V, and that $\dot{E}_c(V) < 0$ if V is not equilibrium. On this basis, it is concluded in [3] that the network is *totally stable,* i.e., that every solution trajectory approaches an equilibrium. Strictly speaking, the argument in [3] in incomplete. In order to make it complete, it is necessary to show in addition that no solution trajectory escapes to infinity in the u-space. This is established in [4]. Thus the results of [3] and [4] mean that, in the case where the interactions are symmetric, the neural network does not exhibit any nontrivial periodic solutions.

Neural networks of the form (1.1) or (1.6) are claimed to be extremely versatile and powerful. In [5], and [6], it is claimed that several important problems, such as the Traveling Salesman Problem, analog-to-digital conversion, and threshold decision making, can be solved using such networks.

In a practical implementation of a neural network of the form (1.6), two difficulties can arise. The first difficulty is that the scaling constant λ need not be the same for all neurons. Thus instead of (1.6), one can have

$$\cdots, v_i = g_i(\lambda_i u_i) \qquad i = 1, \cdots, n. \tag{1.8}$$

The consequences of this are not serious. In fact, it is only necessary to modify the energy function of (1.7) by replacing λ by λ_i; that is,

$$E_c = \sum_{i=1}^{n} \left[\frac{1}{\lambda_i R_i} \int_0^{v_i} g_i^{-1}(v)\, dv - I_i v_i - \sum_{j=1}^{n} \frac{1}{2} t_{ij} v_i v_j \right]. \tag{1.9}$$

With this modification, the arguments of [3], and [4] continue to apply, and the neural network is totally stable. The second difficulty is that it is unrealistic to assume that the interactions are symmetric, since this often requires guaranteeing that two physical quantities (such as resistances or the gains of operational amplifiers) are *exactly* equal. The consequences of even *slight* asymmetries in the interactions are disastrous to the theory of [3], and [4]. If $t_{ij} = t_{ji}$ for all i, j, then the network description (1.6) can be rewritten as

$$C_i \dot{u}_i = -\partial E_c / \partial v_i, \qquad i = 1, \cdots, n. \tag{1.10}$$

However, if $t_{ij} \neq t_{ji}$ for even a single pair (i, j), then (1.10) is no longer true, and it does not matter how small the asymmetry $|t_{ij} - t_{ji}|$ is. In essence, the theory of [3], and [4] is based on the relationship (1.10), and hence cannot be modified to account for asymmetric interactions. As of now, there is very little theory to analyze the behavior of networks of the form (1.6) in the case of asymmetric interactions.

The objective of the present paper is to analyze the number, location, and stability behavior of neural networks described by (1.6), *without* the assumptions of no self-interactions and symmetric interactions. It turns out however that the method of analysis used here is *not* limited to neural networks with *linear* interconnections. To exploit this feature, the object of study in this paper is the neural network described by

$$\begin{aligned} C_i \dot{u}_i &= -\frac{1}{R_i} u_i + \psi_i(V) + I_i \\ v_i &= g_i(\lambda u_i) \qquad i = 1, \cdots, n \end{aligned} \tag{1.11}$$

where C_i, R_i, u_i, v_i, I_i are the same as in (1.6), $V = [v_1 \cdots v_n]^t$, and $\psi_i : (0,1)^n \to \Re$ is some function representing the effects of the interconnections amongst the neurons; $g_i : \Re \to (0,1)$ is a sigmoidal function, and λ is a scaling parameter, as described earlier. The assumptions on the functions ψ_i are stated in the next section, but they are very simple and natural, and include the Hopfield networks of (1.6) as a special case. Hence all the results derived here are applicable to Hopfield networks but apply as well to a far larger class. In particular, since the results derived here are *not* based on energy-type arguments, they apply to systems of the form (1.6) even when the interconnection matrix T is not symmetric, but is only "nearly" symmetric.

Note that, in the system description (1.11), it is assumed that the *same* scaling factor λ appears in all the sigmoidal characteristics. Strictly speaking, this is not realistic; as mentioned earlier, it would be more realistic to assume a relationship of the form (1.8). However, it turns out that this assumption is not crucial to the contents of the paper. The only reason for making it is to simplify notation. Section VIII describes how the arguments in the paper can be modified to cover the more general description (1.8).

The following results are proved in the paper: Consider (1.11) as evolving on the open n-dimensional hypercube $H = (0,1)^n$ in the "V-space," and let $I = [I_1 \cdots I_n]^t$ denote the external input vector. Then, as $\lambda \to \infty$ so that the sigmoid characteristic become steeper and steeper, the following statements are true for all I except for those belonging to a set of measure zero. 1) There are only finitely many equilibria in any compact subset of H. If there are no self-interactions, then these equilibria cannot be exponentially stable, and under mild conditions they are in fact unstable. If the network has symmetric (nonlinear) interactions, whether or not it has self-interactions, then the stable manifolds of all these equilibria have the same dimension, which can be computed explicitly. If the network also has no self-interactions, then all these equilibria are unstable. 2) There are only finitely many equilibria in any face of H. If there are no self-interactions, then there are no equilibria in any edge of H. If the network has symmetric interactions, then the stable manifold of equilibria in parallel faces of H have the same dimension, which can be computed explicitly. If the network also has no self-interactions, then all equilibria in the faces of H are unstable. A corollary of these results is that, in the standard Hopfield-type network, there can be asymptotically stable equilibria only at the corners of H, and trajectories starting at almost all initial conditions approach the corners of $\bar{H}$. 3) A systematic procedure is given for determining which corners of H contain equilibria, and it is shown that all equilibria in the corners of H are asymptotically stable. It is important to note that the proofs here are *not* based on energy arguments. As a result, these results are "hardy" in the sense that they continue to hold even if the network dynamics are slightly perturbed.

II. Preliminaries

In this section, the various assumptions made throughout the paper are briefly summarized.

The input–output relationship of the ith neuron is given by the sigmoid function

$$v_i = g_i(\lambda u_i) \tag{2.1}$$

where g_i is given sigmoid function and λ is a scaling constant. The only assumptions made on the sigmoid function are the following.

A. Assumptions on the Sigmoid Nonlinearities

The $g_i(x)$ is continuously differentiable, strictly increasing, and $g_i(x) \to 1$ as $x \to \infty$, $g_i(x) \to 0$ as $x \to -\infty$. Furthermore, $xg_i'(x) \to 0$ as $|x| \to \infty$.

The assumptions about g_i are quite standard. The assumption about g_i' are almost a consequence of the fact that $g_i(x)$ has a definite limit as $|x| \to \infty$. Since the function $1/x$ is not integrable over any infinite interval, it follows that

$$g_i(x) \to 1 \quad \text{as } x \to \infty \Rightarrow \liminf xg_i'(x) = 0 \quad \text{as } x \to \infty \tag{2.2}$$

and similarly as $x \to -\infty$. So all we have done is to replace "lim inf" by "lim." Note that the commonly used sigmoid function $1/(1 + e^{-x})$ satisfies these assumptions. As $\lambda \to \infty$, the sigmoid becomes steeper and steeper and eventually "approaches" the "sat" function of (1.2). Note that each neuron can have a different switching function, but for simplicity it is assumed that all neurons have the same scaling constant. This assumption is not essential—see Section VIII for a discussion of how this assumption can be relaxed.

B. Assumptions on the Interconnection Nonlinearities

At various stages, we impose a variety of conditions on the functions ψ_i in (1.11). Naturally, the more structure we impose on ψ_i, the more conclusions we are able to draw. But it is interesting to note that some conclusions can be drawn with virtually no assumptions.

(N0): There exists a finite constant μ such that

$$\left|\frac{\partial \psi_i}{\partial v_i}\right| \le \mu \qquad \forall V \in H. \tag{2.3}$$

Since H is a precompact subset of $\mathfrak{R}^n$ (i.e., its closure is a compact set), Condition (N0) is quite mild. In fact, (2.3) is satisfied if each ψ_i has a C^1 extension to $\bar{H} = [0,1]^n$.

(N1) (No Self-Interactions): Condition (N0) is true, and in addition,

$$\frac{\partial \psi_i}{\partial v_i} = 0 \qquad \forall V \in H. \tag{2.4}$$

This says that ψ_i is independent of v_i, but it does not in any way limit the nature of the dependence of ψ_i on v_j, $j \neq i$.

(N2) (Symmetrical Interactions): The function ψ_i has the form

$$\psi_i(V) = \phi_i\left[\sum_{j=1}^{n} t_{ij}\theta_j(v_j)\right] \tag{2.5}$$

where $\phi_i : \mathfrak{R} \to \mathfrak{R}$, $\theta_i : (0,1) \to \mathfrak{R}$ are continuously differentiable and strictly increasing, and t_{ij} are real numbers with

$$t_{ij} = t_{ji}, \qquad \forall i,j. \tag{2.6}$$

In addition, there exists a finite constant μ such that

$$0 < \frac{\partial \phi_i}{\partial v_i} \le \mu, \qquad 0 < \frac{\partial \theta_i}{\partial v_i} \le \mu. \tag{2.7}$$

Finally, the matrix $T = [t_{ij}]$ is hyperbolic; i.e., T has no eigenvalues with zero real part.

Note that (N2) implies (N0) but is independent of (N1). One can think of networks satisfying (N2) as generalized Hopfield-type networks, whereby each neuronal current v_j is first passed through a nonlinearity θ_j, the resulting signals are weighed by t_{ij} and then summed, and finally the weighted sum is fed into another nonlinearity ϕ_i. It is clear that by taking both θ_i and ϕ_i to be identity maps, one recovers the standard Hopfield model (1.6).

(N3) (Symmetric Interactions Plus): Condition (N2) is true. In addition, all principal submatrices[2] of T of size 2×2 or larger are hyperbolic, i.e. none of their eigenvalues has a zero real part.

All of the matrices proposed by Tank and Hopfield satisfy these assumptions. Note that, if the interconnection matrix T has zero diagonal elements, then the assumption of hyperbolicity implies that each principal submatrix of T of dimension 2×2 or larger has at least one eigenvalue with positive real part. This is because the trace of a matrix is equal to the sum of the eigenvalues. Thus if the trace of T is zero, and it has no eigenvalues on the imaginary axis, then it must have some

[2] Recall that a *submatrix* of an $n \times n$ matrix T is obtained by choosing two nonempty subsets $J, K \subseteq \{1, \cdots, n\}$, and forming the matrix consisting of all elements (t_{jk}), $j \in J$, $k \in K$. A *principal* submatrix of T is obtained when $J = K$. Note that a principal submatrix is necessarily square. Moreover, if T is symmetric, so are all principal submatrices of T.

eigenvalues with positive real part and others with negative real part. This is true whether or not T is symmetric.

Define H to be the open hypercube $(0,1)^n$, and $\bar{H}$ to be the closed hypercube $[0,1]^n$. The symbol $\boldsymbol{b}$ denotes the binary set $\{0,1\}$, and $\boldsymbol{b}^n$ denotes the set of n-dimensional binary vectors. Note that the set $\boldsymbol{b}^n$ consists precisely of the 2^n corners of the hypercube $\bar{H}$. The *faces* of the hypercube $\bar{H}$ consist precisely of those vectors $x \in \bar{H}$ with the property that $x_i \in \boldsymbol{b}$ for some but not all values of i. In other words, a face of $\bar{H}$ is a set of the form

$$\left\{x \in \bar{H} : x_i \in \boldsymbol{b}\ \forall i \in I, x_i \in (0,1)\ \forall i \in J\right\} \tag{2.8}$$

where I, J is a nontrivial partition of the set $\{1, \cdots, n\}$.

Among other things, we are interested in the location of the equilibria of (1.11) as the sigmoid gain λ approaches ∞. Three types of equilibria are identified.

1) If $V \in \bar{H}$ is an equilibrium and $v_i \in (0,1)\ \forall i$, then the equilibrium is said to be in the *interior* of $\bar{H}$.
2) If all components of V approach either 0 or 1 as $\lambda \to \infty$, then the equilibrium is said to be in a corner of $\bar{H}$.
3) If some components of V approach 0 or 1 as $\lambda \to \infty$ while others approach some value in $(0,1)$, then the equilibrium is said to be in a *face* of $\bar{H}$.

This section is concluded by recalling a few definitions [7]. Consider a differential equation

$$\dot{x} = f(x) \tag{2.9}$$

where $x \in H$ and $f : H \to \Re^n$ is continuously differentiable. Then a vector $x_e \in H$ is called an *equilibrium* of (2.9) if $f(x_e) = 0$. Now define

$$A = \left[\frac{\partial f}{\partial x}\right]_{x=x_e} \tag{2.10}$$

Then the equilibrium x_e is said to be *hyperbolic* if the Jacobian matrix A has no eigenvalues with zero real part, i.e., if the matrix A is also hyperbolic. Let m denote the number of eigenvalues of a hyperbolic matrix A with positive real part; then A has $n - m$ eigenvalues with negative real part. The ordered pair $(m, n - m)$ is called the *signature* of the (hyperbolic) equilibrium x_e.

III. Motivation: Single-Neuron Case

Much of what happens in a neural network as the neuron characteristics become steeper and steeper can be understood by studying the behavior of (1.6) when $n = 1$. In this case, the network dynamics are described by

$$\dot{u} = -u/\alpha + bg(\lambda u) + y \tag{3.1}$$

where

$$\alpha = R_1C_1, \qquad b = t_{11}/C_1, \qquad y = I_1/C_1. \tag{3.2}$$

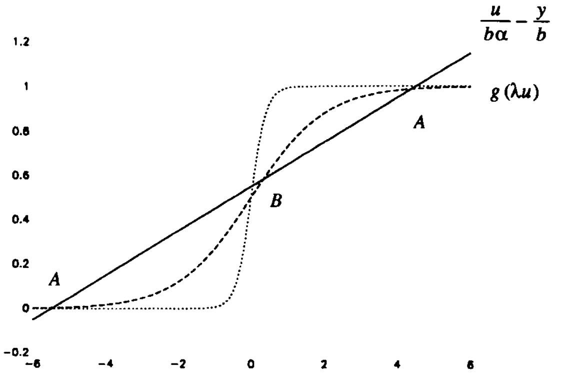


Fig. 1.

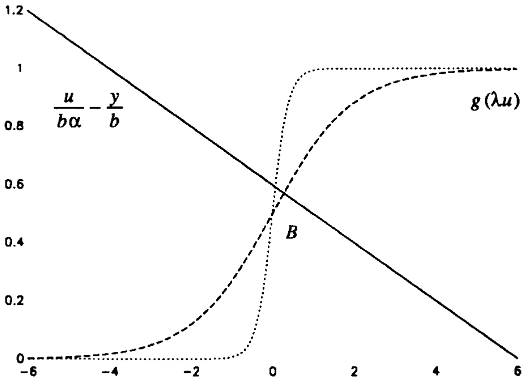


Fig. 2.

So the equilibria of this network are at the solutions of

$$\frac{u}{\alpha} = bg(\lambda u) + y. \tag{3.3}$$

Fig. 1 shows where the solutions of this equation can lie as $\lambda \to \infty$ when $b > 0$, while Fig. 2 does the same when $b < 0$. These figures show that as $\lambda \to \infty$, there can be two types of equilibria. First, those where u_{eq} approaches a finite number, and v_{eq} approaches 0 if $u_{\text{eq}} < 0$ and 1 if $u_{\text{eq}} > 0$; these types of equilibria are labelled as type A in Fig. 1 and 2. Second, those where $u_{\text{eq}} \to 0$ but v_{eq} approaches a number strictly between 0 and 1; this type of equilibrium is labeled as type B in Fig. 1. In Section VI we will see that, in the case of networks containing multiple neurons, it is possible for *some components* of u_{eq} to approach a nonzero value while the remaining components approach zero; in such a case, some components of v_{eq} approach 0 or 1 while the remaining components approach a value strictly between 0 and 1.

IV. Equilibria in the Interior of $\bar{H}$

To analyze the equilibria of (1.11), define

$$\alpha_i = R_iC_i \qquad A = \text{Diag}\{\alpha_1, \cdots, \alpha_n\}, \tag{4.1}$$

$$C = \mathrm{Diag}\{C_1, \cdots, C_n\} \qquad R = \mathrm{Diag}\{R_1, \cdots, R_n\}. \tag{4.2}$$

Define maps $G : \Re^n \to H$ and $\Psi : H \to \Re^n$ by

$$[G(u)]_i = g_i(u_i) \qquad [\Psi(V)]_i = \psi_i(V), \quad i = 1, \cdots, n. \tag{4.3}$$

Then the network equations (1.11) can be rewritten compactly as

$$\dot{u} = -A^{-1}u + C^{-1}[\Psi(V) + I] \qquad V = G(\lambda u) \tag{4.4}$$

where it should be obvious that

$$u = [u_1 \cdots u_n]^t \qquad V = [v_1 \cdots v_n]^t \qquad I = [I_1 \cdots I_n]^t. \tag{4.5}$$

Now the equilibria of (4.4) are the solutions of

$$A^{-1}u = C^{-1}\Psi[G(\lambda u)] + C^{-1}I. \tag{4.6}$$

In this section we are interested in the equilibria of (4.4) in the interior of $\bar{H}$. If all components of V are to stay away from the limits 0 and 1 as $\lambda \to \infty$, then u must approach 0, while λu approaches some well-defined limit. Substituting $u = 0$ in (4.6) and noting that $V = G(\lambda u)$ gives

$$\Psi(V) = -I. \tag{4.7}$$

The question is: How many solutions does (4.7) have and what is their nature?

Proposition 4.1: Let S be any compact subset of H. Then, for all $I \in \Re^n$ except those belonging to a set of measure zero, the network (4.4) has only finitely many equilibria in S as $\lambda \to \infty$.

Proof: The result is virtually a direct consequence of Sard's Theorem [8]. For the convenience of the reader, the relevant definitions and the theorem itself are summarized. A vector $p \in H$ is called a *critical point* of the differentiable map Ψ if the Jacobian matrix $J_\Psi(p) = [\partial\Psi/\partial V](p)$ is singular; otherwise it is called a *regular point.* A vector $q \in \Re^n$ is called a *regular value* of the map Ψ if every point in the preimage $\Psi^{-1}(q)$ is a regular point; otherwise, q is called a *critical value.* Note that if q is a regular value of Ψ, then every point in the set $\Psi^{-1}(q)$ is *isolated;* i.e., every point $p \in \Psi^{-1}(q)$ has a neighborhood that does not contain any other point of $\Psi^{-1}(q)$. This is a ready consequence of the fact that the Jacobian matrix of Ψ evaluated at p is nonsingular. Now a standard compactness argument shows that if q is a regular value of Ψ, then any compact subset $S \subset H$ can contain at most a finite number of points $\Psi^{-1}(p)$, i.e., at most a finite number of solutions of the equation $\Psi(p) = q$. Now Sard's theorem [8] says, quite simply, that the set of critical values of a differentiable map has measure zero.

Proposition 4.2: Suppose the function Ψ satisfies the no self-interactions assumption (N1), and that all functions in (2.5) are twice continuously differentiable. Let S be a compact subset of H, and let $I \in \Re^n$. Then, as $\lambda \to \infty$, the equilibria of (4.4) inside S, if any, are not exponentially stable.

Proof: Suppose $p \in S$ satisfies $\Psi(p) = -I$, and let J_m denote the Jacobian matrix of a map M. Then, as $\lambda \to \infty$, the network has an equilibrium approaching p. Let us linearize the network around this equilibrium. For this purpose, define $\bar{u} = \lambda u$. Then, as $\lambda \to \infty$, we have that $\bar{u} \to G^{-1}(p)$. Now

$$\frac{d}{d\bar{u}}\left[-A^{-1}u + C^{-1}\Psi(V) + C^{-1}I\right] = -\lambda^{-1}A^{-1} + C^{-1}J_\Psi \cdot [G(\bar{u})]J_G(\bar{u}). \tag{4.8}$$

As $\lambda \to \infty$, the first term approaches zero, and we are left with

$$C^{-1}J_\Psi(p)J_G\left[G^{-1}(p)\right] = B, \text{say}. \tag{4.9}$$

Now, by the no self-interactions assumption, it follows that the diagonal elements of J_Ψ are all zero, from which it follows that $b_{ii} = 0 \ \forall i$. Therefore the sum of the eigenvalues of B, equal to the trace of B, is also zero. Thus there are only two possibilities.

1) B has at least one eigenvalue with positive real part, in which case the equilibrium is unstable.
2) All of eigenvalues of B have zero real parts. In this case, the equilibrium is not exponentially stable.

Although it cannot be stated as a theorem, one can see that case 2) is quite unlikely. It is much more likely that B has at least one eigenvalue with positive real part, in which case any equilibria that stay in the interior of $\bar{H}$ as $\lambda \to \infty$ are unstable. This observation perhaps sheds light on why it is useful to prohibit neural networks from having self-interactions.

Proposition 4.3: Suppose the function Ψ satisfies assumption (N2) and let S be any compact subset of H. Let $I \in \Re^n$. Then, as $\lambda \to \infty$, the equilibria of (4.4) that remain inside S are all hyperbolic. Moreover, the dimensions of the stable manifolds of all these equilibria are all the same, and equal the number of negative eigenvalues of the interconnection matrix T. If Ψ satisfies Assumption (N1) as well as (N2), then all equilibria inside S are unstable.

Proof: Suppose $p \in S$ satisfies $\Psi(p) = I$, and linearize (4.4) around the equilibrium $G^{-1}(p)$. Define maps $\Phi : \Re^n \to \Re^n$ and $\Theta : H \to \Re^n$ in the obvious way [cf. (2.5)]. Now, as in the proof of Proposition 4.2, it follows that as $\lambda \to \infty$, the quantity $\bar{u}$ approaches $G^{-1}(p)$. Let us compute the matrix B of (4.9), noting that in the present case

$$\Psi(V) = \Phi[T\Theta(V)]. \tag{4.10}$$

Hence

$$J_\Psi(V) = J_\Phi[T\Theta(V)]TJ_\Theta(V). \tag{4.11}$$

Substituting from (4.11) into (4.9) shows that the matrix B has the form

$$B = PTQ \tag{4.12}$$

where P and Q are diagonal matrices with positive entries. Now it is a well-known result (see e.g., [9, p. 297]) that if M is any nonsingular matrix, then T and M^tTM have the same *signature,* i.e., the same number of positive, zero, and negative eigenvalues.[3] Next, note that P and Q commute, since they are both diagonal. Define $D = P^{1/2}$, $S = Q^{1/2}$, and note that both D and S are also diagonal with positive entries, and that D and S commute. Therefore,

$$B = PTQ = S^{-1}D[DSTSD]D^{-1}S \tag{4.13}$$

is similar to $W = DSTSD$, and as a consequence both B and W have the same eigenvalues. In turn W has the same signature as T. Hence each equilibrium of (4.4) inside S is hyperbolic, and the desired conclusion follows. Finally, if Ψ also satisfies Assumption (N1), then $t_{ii} = 0 \; \forall i$. This, plus the fact that T is hyperbolic, implies that T has at least one positive eigenvalue. Hence all equilibria inside S are unstable.

The various neural networks proposed by Tank and Hopfield have the feature that the interconnection matrix T is hyperbolic and has zero diagonal elements. Hence T has at least one positive eigenvalue. Thus Proposition 4.3 shows that in such neural networks almost all trajectories move away from the interior of the hypercube $\bar{H}$.

It is important to note that Propositions 4.1–4.3 remain valid even if the dynamics of the neural network are slightly perturbed. In particular, if Assumption (N2) is violated in the sense that the interconnection matrix T is not symmetric but is "close" to a symmetric matrix, then the matrix $B = PTQ$ will not in general have only real eigenvalues, but B will continue to be hyperbolic and to have the same "signature" as T, in the sense that both B and T have the same number of eigenvalues with negative real part.

Finally, suppose the network is of the type (1.6), i.e., it is a standard Hopfield-type network. This is a special cas of Assumption (N2) with all ϕ_i and θ_i set equal to the identity map. In this case the only thing we gain is that (4.7) has a unique solution, namely $V = T^{-1}I$. Hence, in Proposition 4.1, one can replace "a finite number" by "at most one."

Example 4.4: As an illustration of Proposition 4.1, consider the A/D converter circuit of [6]. If we study the four-bit converter, then $n = 4$, and

$$T = \begin{bmatrix} 0 & -2 & -4 & -8 \\ -2 & 0 & -8 & -16 \\ -4 & -8 & 0 & -32 \\ -8 & -16 & -32 & 0 \end{bmatrix} \qquad I = \begin{bmatrix} 1 \\ 2 \\ 4 \\ 8 \end{bmatrix} x - \begin{bmatrix} 0.5 \\ 2 \\ 8 \\ 32 \end{bmatrix} \tag{4.14}$$

where x is the real number which is to be quantized. This neural network evolves on the four-dimensional open hypercube $H = (0,1)^4$. The objective of the example is to determine the range of values of x for which the network has an equilibrium in the interior or H, and to determine the dimensions of its stable and unstable manifolds.

Taking the second question first, it is easy to verify that T has one negative and three positive eigenvalues. Thus if the network has an equilibrium in the interior of $\bar{H}$, it is hyperbolic, and its stable and unstable manifolds have dimensions one and three, respectively.

Next, we compute

$$V_{\text{eq}} = -T^{-1}I = \begin{bmatrix} -2 \\ -0.75 \\ -0.125 \\ -0.1875 \end{bmatrix} + \begin{bmatrix} 1/3 \\ 1/6 \\ 1/12 \\ 1/24 \end{bmatrix} x. \tag{4.15}$$

It is routine to verify that the above vector belongs to the open hypercube H if and only if $6 < x < 9$. Thus the neural network corresponding to the four-bit A/D converter has an equilibrium in the interior of $\bar{H}$ if and only if x belongs to the open interval $(6, 9)$.

V. Equilibria in the Corners

In this section, we study whether any equilibria of the system (4.4) approach the corners of the hypercube $\bar{H}$ as the sigmoid gain $\lambda \to \infty$. Recall that $\boldsymbol{b}$ denotes the Boolean set $\{0, 1\}$, so that $\boldsymbol{b}^n$ is the set corners of the closed hypercube $\bar{H}$. Now, since the differential equation (4.4) evolves on the *open* hypercube H, no vector in $\boldsymbol{b}^n$ can actually be an equilibrium of this system. However, it is possible that, as $\lambda \to \infty$, some equilibria of (4.4) *approach* a vector in $\boldsymbol{b}^n$

Proposition 5.1: Let e be an arbitrary vector in $\boldsymbol{b}^n$. Then an equilibrium of (4.4) approaches e as $\lambda \to \infty$ if and only if e satisfies the *parity condition,* defined as follows: Let $z = \Psi(e) + I$. Then

$$z_i > 0 \quad \text{if } e_i = 1, \qquad z_i < 0 \quad \text{if } e_i = 0, \qquad i = 1, \cdots, n. \tag{5.1}$$

Remark: Note that the parity condition can also be expressed as

$$e_i = \text{sat}[\Psi(e) + I]_i \qquad i = 1, \cdots, n \tag{5.2}$$

where the "sat" function is defined in (1.2).

Proof: Put $\dot{u} = 0$ in (4.4). This gives

$$0 = -A^{-1}u + C^{-1}[\Psi(V) + I] \tag{5.3}$$

or

$$u = AC^{-1}[\Psi(V) + I] = R[\Psi(V) + I]. \tag{5.4}$$

Now, if we substitute $G(\lambda u) = V = e \in \boldsymbol{b}^n$, then we get

$$u_{\text{eq}} = R[\Psi(e) + I] = Rz. \tag{5.5}$$

Thus as $\lambda \to \infty$, $V = G(\lambda u_{\text{eq}}) \to e$, *provided*

$$(u_{\text{eq}})_i > 0 \quad \text{if } e_i = 1 \qquad (u_{\text{eq}})_i < 0 \quad \text{if } e_i = 0 \qquad i = 1, \cdots, n \tag{5.6}$$

But since $u_{\text{eq}} = Rz$, $(u_{\text{eq}})_i = R_i z_i$ for all i, and it follows that each component of u_{eq} has the same sign as the corresponding component of z. Hence (5.6) is equivalent to (5.1).

Proposition 5.1 is not very surprising, since it is very similar to a related result for (DTDS) networks of the form (1.1). A

[3] Recall that the eigenvalues of a symmetric matrix are real.

binary vector $e \in \boldsymbol{b}^n$ is called a *fixed point* of the DTDS network (1.1) if

$$V^0 = e \Rightarrow V^t = e \quad \forall\, t \geq 0. \tag{5.7}$$

In other words, e is a fixed point if, whenever the network starts in the state e, it remains there. It is easy to see [10, p. 7] that e is a fixed point if and only if

$$e_i = \text{sat}[Te + I]_i \qquad i = 1, \cdots, n. \tag{5.8}$$

Now consider the associated continuous-time, continous-state (CTCS) network (1.6), where t_{ij} and I_i are the same as in (1.1). In this case, the interconnection function $\Psi(e)$ of (4.3) is given by

$$\Psi(e) = Te + I. \tag{5.9}$$

Hence the parity condition (5.1) (or (5.2)) is precisely (5.8). The conclusion can be stated as follows: The CTCS network (1.6) has an equilibrium that approaches $e \in \boldsymbol{b}^n$ as $\lambda \rightarrow \infty$ if and only if e is a fixed point of the associated DTDS network (1.11).

Proposition 5.2: Suppose that an equilibrium of (4.4) approaches an element of $\boldsymbol{b}^n$ as $\lambda \rightarrow \infty$. Then this equilibrium is exponentially stable for all sufficiently large λ.

Proof: Linearize (4.4) around the equilibrium u_{eq} of (5.5). The Jacobian matrix of the right side of (4.4) at u_{eq} is

$$-A^{-1} + J_{\Psi}(e) J_G(\lambda u_{\text{eq}})\lambda. \tag{5.10}$$

By assumption, $\lambda J_G(\lambda u_{\text{eq}}) \rightarrow 0$ as $\lambda \rightarrow \infty$. Hence the Jacobian approaches $-A^{-1}$, whose eigenvalues are $-1/\alpha_1, \cdots, -1/\alpha_n$. Since all of these eigenvalues are negative, it follows from the linearization theorem [11, p. 188] that the equilibrium u_{eq} is exponentially stable.

Remark: An informal, but informative, way to state the above proposition is: "All equilibria approaching the corners of $\bar{H}$ are asymptotically stable."

Proposition 5.2 brings out an important difference between DTDS networks of the form (1.1) and CTCS networks of the form (1.6). In the case of DTDS networks, not all fixed points need be attractive. In fact, there are very few results concerning the attractivity of fixed points (see e.g., [12], [13], and [10, pp. 38 *et seq.*]). In contrast, in the case of CTCS networks of the form (1.6) (or the more general (4.3)), *every* equilibrium near a corner of $\bar{H}$ is exponentially stable. The difference arises because of the difference between the two models. Suppose $e \in \boldsymbol{b}^n$ is a fixed point of the DTDS network (1.1). By previous remarks, it follows that an equilibrium of (1.6) approaches e as $\lambda \rightarrow \infty$. Let u_{eq} denote this equilibrium. Proposition 5.2 states that u_{eq} is exponentially stable. This means that, if the initial state of the network (1.6) is *sufficiently close* to u_{eq}, then the resulting solution trajectory will converge to u_{eq}. But in the case of the network (1.1), the state vectors are discretized. Hence, in this network, there is no concept of "sufficiently small" perturbations of the initial state. The only possible perturbations of e are to change some of the 1's to 0's or vice versa. With such a perturbation, it is quite possible that the resulting trajectory will not converge to e.

Example 5.3: Consider again the four-bit A/D converter of Example 4.4. In [6] it is claimed that, if x is any real number and if the neural network is started from the zero initial state (i.e., $u_i = 0$ for all i), then eventually the vector V will converge to the correct binary quantization of the real number x. However, it is observed in [6] that sometimes the vector V converges to a binary number which is either one less or one more than the correct quantization of x. This problem is referred to in [6] as "hysteresis." (See [6, Fig. 3].) Hence, for a given x there could be more than one stable equilibrium of the neural network, and depending on the initial condition the solution trajectory of the neural network could converge to an incorrect binary vector. If x is not kept fixed but is changed periodically, then it is necessary to "re-initialize" the network each time x is changed. Otherwise the solution trajectory will converge to an incorrect value.

Since the neural network has four neurons, there are $2^4 = 16$ possible binary vectors, or 16 corners to the hypercube $\bar{H}$. By taking each corner in turn, it is possible to determine the values of x for which an equilibrium exists at that corner. This can be done using Proposition 5.1. By Proposition 5.2, each such equilibrium is asymptotically stable. Hence, for some initial values of u at least, the solution trajectory will converge to that corner.

To illustrate that application of Proposition 5.1, consider the corner $e = [1\,0\,1\,1]^t$. Note that the first component represents the lowest or least significant bit whereas the last component represents the highest bit. Hence this vector corresponds to the binary representation of the integer 13. To determine for what values of x an equilibrium exists near this corner, we compute the vector $Te + I$, as per Proposition 5.1. This gives

$$Te + I = -\begin{bmatrix} 12.5 \\ 28 \\ 44 \\ 72 \end{bmatrix} + \begin{bmatrix} 1 \\ 2 \\ 4 \\ 8 \end{bmatrix} x. \tag{5.11}$$

Now, in order for an equilibrium to exist near this corner, a necessary and sufficient condition is that

$$-12.5 + x > 0 \qquad -28 + 2x < 0 \qquad -44 + 4x > 0$$
$$-72 + 8x > 0. \tag{5.12}$$

Solving these inequalities shows that an equilibrium exists near this corner if and only if

$$12.5 < x < 14. \tag{5.13}$$

The same process can be repeated at all 16 binary vectors, and corresponding intervals of x can be computed. This is displayed in Table I. (It is easy to show, using Proposition 5.1, that the set of values of x corresponding to a given binary vector is always an interval.) For ease of presentation, the 16 binary vectors have been shown in terms of the corresponding decimal integer.

From Table I one can see that, corresponding a given real number x for which it is desired to find a binary quantization, there can be as many as *three* distinct asymptotically stable equilibria. Moreover, some of these equilibria need not be anywhere close to the correct binary quantization. For example, if

TABLE I

e	x	e	x	e	x
0	$x < 0.5$	1	$0.5 < x < 2$	2	$1 < x < 2.5$
3	$2.5 < x < 5$	4	$2 < x < 4.5$	5	$4.5 < x < 6$
6	$5 < x < 6.5$	7	$6.5 < x < 11$	8	$4 < x < 8.5$
9	$8.5 < x < 10$	10	$9 < x < 10.5$	11	$10.5 < x < 13$
12	$10 < x < 12.5$	13	$12.5 < x < 14$	14	$13 < x < 14.5$
15	$14.5 < x$				

$x = 4.3$, then there are three asymptotically stable equilibria, at (in decimal representation) $e = 3, 4, 8$. As per the convention of Tank and Hopfield, if $3.5 < x < 4.5$, then the correct binary quantization is 4. Hence one would hope that the neural network would converge toward the corner $e = 4 = [0\,0\,1\,0]^t$. But since there are two other asymptotically stable equilibria, for suitable initial conditions the neural network will in fact converge toward the corners $e = 3$ or 8. If the network which should converge to 4, in fact converges to 3, then one can consider it as "hysteresis," as mentioned in [6], since the difference between 3 and 4 is only one. But as Table I shows, it is possible for the network to converge to a corner that is at a (Euclidean) distance *more than one* from the correct value This phenomenon is not mentioned in [6]. Indeed, they do not offer any systematic procedure for identifying *all* attractive equilibria, as is done here. Similarly, if $x = 10.3$, then there are asymptotically stable equilibria at $e = 7, 10$, and 11. Once again, for suitable initial conditions the network will converge to the corner 7 when it should converge to 10, and of course the distance between 7 and 10 is more than one. This brings up the question of whether there is an improved version of an A/D converter which does not exhibit such multiple asymptotically stable equilibria. The answer is "yes," as shown in [14], and [15].

One final comment: Although the parity test of Proposition 5.1 gives a systematic procedure for identifying all the equilibria of a given neural network near the corners of $\bar{H}$, the number of operations needed to apply the parity test of order 2^n where n is the number of neurons. Hence, as an *analysis* tool, the parity test is not very useful. However, it is very useful as a *synthesis* tool, i.e., as a method for constructing an neural network with equilibria near prescribed corners of $\bar{H}$. For example, in [14] and [15], the parity test is used to construct an analog to digital converter neural network that has only a single, globally attractive equilibrium for almost all values of the input. Moreover, as stated in the remark following Proposition 5.1, the problem of constructing a CTCS network of the form (1.6) with equilibria near prescribed corners of $\bar{H}$ is mathematically equivalent to the problem of constructing a DTDS network of the form (1.1) with fixed points at the same corners. Hence, the known methods for achieving this in DTDS networks, e.g., [1] and [13], can also be used to construct CTCS networks.

VI. Equilibria in the Faces of $\bar{H}$

Thus far we have studied the existence of equilibria in the interior of $\bar{H}$, and near the corners of $\bar{H}$. In this section, we complete the analysis by studying conditions under which there exist equilibria in the *faces* of $\bar{H}$, i.e., equilibria where some components approach 0 or 1 while the remaining components remain bounded away from 0 and 1 as $\lambda \to \infty$.

We are searching for solutions to

$$R^{-1}u = \Psi(e) + I = z \tag{6.1}$$

where *some components* of e belong to $\{0, 1\}$ while the remaining components belong to the open interval $(0, 1)$. Note that if some component of e belongs to $(0, 1)$, then the corresponding component of u_{eq} (and of z) must be zero; otherwise $g(\lambda u) \to 0$ or 1 as $\lambda \to \infty$.

Let us first define a few terms. Suppose $1 \leq k \leq n-1$, and let $\pi = \{\pi_1, \cdots, \pi_k\}$ be a subset of $\{1, \cdots, n\}$. Then the set

$$\{e \in \bar{H} : e_i \in (0,1) \text{for } i \in \pi, e_i \in \{0,1\} \text{for } i \notin \pi\} = F \tag{6.2}$$

defines a *face* of the hypercube $\bar{H}$ of dimension k. Once the set π is defined, there are 2^{n-k} ways of choosing the components of e_i, $i \notin \pi$. Each choice defines a distinct face of $\bar{H}$, and these faces are said to be *parallel.*

Now back to the problem of studying equilibria in the faces of $\bar{H}$. Fix an integer k such that $1 \leq k \leq n-1$, as well as a subset π as above, and consider the problem of studying the equilibria in a corresponding k-dimensional face F of $\bar{H}$. By renumbering the indices if necessary, it can be assumed without loss of generality that the set π equals $\{1, \cdots, k\}$. So suppose a binary vector $e_b = [e_{k+1} \cdots e_n]^t \in \boldsymbol{b}^{n-k}$ is specified, and let F defined in (6.2) denote the corresponding face of $\bar{H}$. Define

$$e_a = [e_1 \cdots e_k]^t \qquad e_b = [e_{k+1} \cdots e_n]^t \tag{6.3}$$

and partition I, z and Ψ commensurately. Then, in partitioned form, (6.1) becomes

$$0_k = \Psi_a(e_a, e_b) + I_a \tag{6.4}$$

$$z_b = \Psi_b(e_a, e_b) + I_b. \tag{6.5}$$

Proposition 6.1: Given an input $I \in \mathfrak{R}^n$ and a binary vector $e_b \in \boldsymbol{b}^{n-k}$, there exists an equilibrium of (4.4) approaching the face F as $\lambda \to \infty$ if and only if the following statement is true: Equation (6.4) has a solution $e_a \in (0,1)^k$, and the corresponding z_b given by (6.5) satisfies the parity condition, namely

$$(z_b)_i > 0 \quad \text{if } (e_b)_i = 1, \qquad (z_b)_i < 0 \quad \text{if } (e_b)_i = 0. \tag{6.6}$$

The proof is virtually the same as that of Proposition 5.1 and is therefore omitted.

Proposition 6.2: Fix $e_b \in \boldsymbol{b}^{n-k}$ and let F be the corresponding face of $\bar{H}$ defined in (6.2). Suppose S is a compact subset of F. Then, for all $I \in \mathfrak{R}^n$ except for those belonging to a set of measure zero, there are only finitely many equilibria of (4.4) that approach S as $\lambda \to \infty$.

The proof is virtually the same as that of Proposition 4.1 and is therefore omitted.

Proposition 6.3: Suppose the function Ψ satisfies the no self-interactions assumption (N1). Then, for all $I \in \mathfrak{R}^n$ except those belonging to a set of measure zero, no equilibrium of (4.4) approaches an edge of $\bar{H}$.

Proof: An edge of $\bar{H}$ is a face with $k = 1$. In this case, Assumption (N1) implies that $\Psi(e_a, e_b)$ is independent of e_a. Hence (6.4) reduces to

$$0 = \Psi_a(e_b) + I_a \tag{6.7}$$

which can only be satisfied if $I_a = \Psi_a(e_b)$. This defines a linear variety in $\Re^n$, which has measure zero. Now $\bar{H}$ has only finitely many edges, and a finite union of sets of measure zero one again has measure zero.

Example 6.4: (Three-Bit A/D Converter) As an illustration of Proposition 6.2, consider the Tank and Hopfield A/D converter circuit of [6], but this time with only three neurons, so that it does a three-bit quantization of a given real number. In this case,

$$T = \begin{bmatrix} 0 & -2 & -4 \\ -2 & 0 & -8 \\ -4 & -8 & 0 \end{bmatrix} \qquad I = -\begin{bmatrix} 0.5 \\ 2 \\ 4 \end{bmatrix} + \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix} x. \tag{6.8}$$

Let $x = 3.2$; we show that it is possible to obtain a complete characterization of all equilibria of the neural network.

First, compute

$$V_{\text{eq}} = -T^{-1}I = \begin{bmatrix} 0.35 \\ 0.425 \\ 0.4625 \end{bmatrix}. \tag{6.9}$$

Since $V_{\text{eq}} \in (0,1)^3$, there is indeed an equilibrium at this point as $\lambda \to \infty$ i.e., as the neural characteristics approach those of an ideal switch. Next, let us check for equilibria in the corners of $\bar{H} = [0,1]^3$. Using the procedure of Proposition 5.1 as illustrated in Example 5.2, one finds that there are (asymptotically stable) equilibria only at $e = [0\,0\,1]^t = 4$ and at $e = [1\,1\,0]^t = 3$. Finally, let us check for solutions of (6.4) in the faces of $\bar{H}$. First, since all diagonal elements of T are zero, it follows from Proposition 6.3 that there are no equilibria along the edges of the cube $\bar{H}$. Next we try setting one component of e equal to zero and solving for the other two. If we set $e_1 = 0$, then solving (6.4) gives

$$\begin{bmatrix} e_2 \\ e_3 \end{bmatrix} = \begin{bmatrix} 0.6 \\ 0.55 \end{bmatrix} \in (0,1)^2. \tag{6.10}$$

Thus it can be concluded that, as $\lambda \to \infty$, there will be an equilibrium near $V = [0\ 0.6\ 0.55]^t$. Similarly it can be verified that there will be another equilibrium near $[1\ 0.1\ 0.3]^t$, and that these are the only equilibria along the faces of $[0,1]^3$.

Next, let us study the stability of equilibria in the faces of $\bar{H}$.

Proposition 6.5: Suppose the function Ψ satisfies Assumption (N3). Suppose $2 \le k \le n-1$, and that $\pi = \{\pi_1, \cdots, \pi_k\}$ is a given subset of $\{1, \cdots, n\}$. Define T_π to be the $(n-k) \times (n-k)$ principal submatrix of T given by

$$T_\pi = [t_{ij}, i, j \notin \pi]. \tag{6.11}$$

Let $e_b \in \boldsymbol{b}^{n-k}$ be chosen arbitrarily, and define F to be the corresponding face of $\bar{H}$ given by (6.2). Let S be a compact subset of F. Then, as $\lambda \to \infty$, any equilibria of (4.4) that approach S are hyperbolic. Moreover, the unstable manifold of all such equilibria have the same dimension, and it equals the number of positive eigenvalues of T_π. If Ψ satisfies (N1) as well as (N3), then all equilibria in the faces of $\bar{H}$ are unstable.

Remark: Once the index set π is fixed, there are 2^{n-k} different possible choices for the matrix e_b. Proposition 6.5 makes it clear that the equilibria in *each* of these faces, if any, have the same signature. To put it another way, equilibria in parallel faces have the same signature.

Proof: For convenience, renumber the indices such that $\pi = \{1, \cdots, k\}$. Now define

$$\Lambda_k = \begin{bmatrix} \lambda I_k & 0 \\ 0 & I_{n-k} \end{bmatrix} \qquad u_k = \Lambda_k u. \tag{6.12}$$

Then from (4.4) it follows that

$$\dot{u}_k = \Lambda_k \dot{u} = -\Lambda_k A^{-1} u + C^{-1}\Lambda_k\{\Psi[G(\lambda u)] + I\} \tag{6.13}$$
$$= -A^{-1}u_k + C^{-1}\Lambda_k\{\Psi[G(\lambda\Lambda_k^{-1}u_k)] + 1\}. \tag{6.14}$$

Here we have used the obvious fact that

$$\Lambda_k A^{-1} \Lambda_k^{-1} = A^{-1} \tag{6.15}$$

since all matrices are diagonal. Now let $\lambda \to \infty$ and suppose an equilibrium e approaches S, i.e.,

$$G(\lambda\Lambda_k^{-1}u_k) \to \begin{bmatrix} e_a \\ e_b \end{bmatrix} = e. \tag{6.16}$$

Define

$$u* = G^{-1}(e). \tag{6.17}$$

Now linearize (4.4) around the equilibrium in u_k-space. The Jacobian matrix is

$$-A^{-1} + C^{-1}\Lambda_k J_\Phi[T\Theta(e)]TJ_\Theta(e)J_G(\lambda\Lambda_k^{-1}u*)\lambda\Lambda_k^{-1} \tag{6.18}$$

Now consider separately the matrix

$$M = J_\Theta(e)J_G(\lambda\Lambda_k^{-1}u*)\lambda\Lambda_k^{-1}. \tag{6.19}$$

This is a diagonal matrix; moreover

$$m_{ii} = \theta_i'(e_i)g_i'[u_i*] \qquad \text{for } 1 \le i \le k, \tag{6.20}$$

$$m_{ii} = \theta_i'(e_i)\lambda g_i'[\lambda u_i*] \to 0 \quad \text{as } \to \infty, \quad \text{for } k+1 \le i \le n. \tag{6.21}$$

Hence, as $\lambda \to \infty$

$$M \to \begin{bmatrix} M_a & 0 \\ 0 & 0 \end{bmatrix} \tag{6.22}$$

and the Jacobian matrix approaches

$$\begin{bmatrix} -A_a^{-1} + \lambda C_a^{-1}(J_\Phi)_a T_{aa} M_a & 0 \\ C_b^{-1}(J_\Phi)_b T_{ba} M_a & -A_b^{-1} \end{bmatrix} \tag{6.23}$$

where we use the obvious notation partitioning the matrices A, C, J_Φ, and T. Note that T_{aa} is just the matrix T_π defined in (6.11). Now as $\lambda \to \infty$, the term A_a^{-1} in top left corner becomes insignificant compared to the term

$$\lambda C_a^{-1}(J_\Phi)_a T_{aa} M_a = \lambda W, \text{say}. \qquad (6.24)$$

Thus the eigenvalues of the linearized system approach

$$\operatorname{spec}(\lambda W) \cup \operatorname{spec}(-A_b^{-1}) \qquad (6.25)$$

where "spec" denotes the set of eigenvalues of a matrix. Of course, the eigenvalues of $-A_b^{-1}$ are just $\{-1/\alpha_{k+1}, \cdots, -1/\alpha_n\}$. Now one can show, just as in the proof of Proposition 4.3, that W has the same signature as $T_{aa} = T_\pi$. The result follows. Finally, if $t_{ii} = 0\ \forall i$, then T_π has at least one positive eigenvalue, whence the equilibria in F are unstable.

Corollary 6.6: Consider the neural network (4.4), where the interaction function Ψ has the form (2.5). Suppose all functions in (2.5) are twice continuously differentiable, and that the interconnection matrix T satisfies $t_{ii} = 0 \forall i$. Under these conditions, the network can have exponentially stable equilibria only at the corners of $\bar{H}$.

Remark: Note that the class of networks covered by Corollary 6.6 includes the standard Hopfield model (1.6) as a special case. The corollary states that, merely by avoiding self-interactions, one can ensure that the network can have exponentially stable equilibria only at the corners of $\bar{H}$. This corollary sheds some light on the role played by the "no self-interaction" assumption on neural network dynamics. This result is important because the energy arguments of [3], and [4] require *only* the symmetric interactions assumption, and do not require the no self-interaction assumption. Thus a network with self-interactions is still totally stable, provided that the interactions are symmetric. But, in such a case, it is possible, for example, that all solution trajectories will converge to an exponential stable equilibrium in the interior of $\bar{H}$. However, if the network has no self-interactions, then there can be exponentially stable equilibria only at the corners of $\bar{H}$, provided that the interactions are symmetric, or "nearly" so. The next corollary gives an even stronger result, but at the expense of more assumptions.

Proof: The hypotheses ensure that the interaction function Ψ satisfies the no self-interaction assumption (N1). Hence, it follows from Proposition 4.2 that there cannot be any exponentially stable equilibria in the interior of $\bar{H}$. Next, it follows from Proposition 6.3 that no equilibrium approaches an edge of $\bar{H}$. Finally, suppose an equilibrium approaches a face of $\bar{H}$, and denote it by u_{eq}. Now u_{eq} is exponentially stable if and only if the eigenvalues of the linearization around u_{eq} all have negative real parts. From (6.25), if follows that these eigenvalues include those of the matrix λW. However, from (6.24) it follows that the diagonal elements of W are all zero, since $t_{ii} = 0\ \forall i$, and the remaining matrices in (6.24) are all diagonal. Hence the sum of the eigenvalues of λW, which equals the trace of the matrix, is also zero. In particular, it is not possible for all of them to have negative real parts.

Corollary 6.7: Consider the neural network (4.4), where the interaction function Ψ satisfies Assumptions (N1) and (N3). Then, as $\lambda \to \infty$, all equilibria except those approaching the corners of $\bar{H}$ are unstable. Hence, trajectories starting from almost all initial conditions approach the corners of $\bar{H}$. In particular, this is true of Hopfield-type networks of the form (1.6).

Example 6.8: Let us continue Example 6.4. The analysis previously carried out shows that there is an equilibrium at $V_{\text{eq}} = [0.34\ 0.425\ 0.4625]^t$. Now the matrix T of (6.8) has one negative and two positive eigenvalue. Accordingly, from Proposition 4.1, this equilibrium has a stable manifold of dimension one and an unstable manifold of dimension two. Next, there are asymptotically stable equilibria at $e_1 = [0\,0\,1]^t$ and $e_2 = [1\,1\,0]^t$. Now consider equilibria in the faces. Letting $\pi = \{2.3\}$ and assigning $e_1 = 0$ leads to the equilibrium at $V_1 = [0\ 0.6\ 0.55]^t$, whereas assigning $e_1 = 1$ leads to the equilibrium $V_2 = [1\ 0.1\ 0.3]^t$. These equilibria are in opposite faces of the three-dimensional cube $[0,1]^3$. Now

$$T_\pi = \begin{bmatrix} t_{22} & t_{23} \\ t_{32} & t_{33} \end{bmatrix} = \begin{bmatrix} 0 & -8 \\ -8 & 0 \end{bmatrix}. \qquad (6.26)$$

The matrix has one positive eigenvalue. This shows that both V_1 and V_2 have stable manifolds of dimension two and an unstable manifold of dimension one.

The most important point to note about this example is that *all of the above conclusions remain valid even if the interconnection matrix is perturbed slightly from its original symmetric value.* Of course, the actual values of the various equilibria will change slightly in a continuous fashion, but the dimensions of the various stable and unstable manifolds will not change.

VII. Specialized Results

Thus far the emphasis has been on general nonlinear neural networks. In the present section, some specialized results are presented for Hopfield-type neural networks described by (1.6), where all the sigmoidal characteristics are identical (representing identical neurons).

A. Rate of Convergence of Trajectories

In this subsection some preliminary results are given about the rate at which the equilibria of the system (1.6) approach the corners of $\bar{H}$, and the rate at which the solution trajectories approach the equilibria.

Suppose e is a vector in $\boldsymbol{b}^n$, i.e., suppose e is a corner point of the hypercube $\bar{H}$. Then Proposition 5.1 states that the system (1.6) has an equilibrium approaching e if and only if the vector $z = Te + I$ has the same "parity" as e. Thus in Proposition 5.1, only the signs of the various components of z are pertinent, and their magnitudes do not play any role in determining whether or not *there exists* an equilibrium near a particular corner. Now it is shown that the magnitudes of the components of z do determine the *speed of convergence* of the equilibrium to e as the sigmoid gain $\lambda \to \infty$.

To be specific, suppose all the neural characteristics are identical, and are given by

$$g_i(x) = \frac{1}{1 + e^{-x}} \qquad \text{for } i = 1, \cdots, n. \qquad (7.1)$$

Suppose $e \in \boldsymbol{b}^n$ and that $z = Te + I$ has the same parity as e. In accordance with (5.5), define

$$u_{\mathrm{eq}i} = \frac{\alpha_i}{c_i} z_i = R_i z_i \qquad \text{for } i = 1, \cdots, n. \tag{7.2}$$

Now define

$$V_{\mathrm{eq}} = G(\lambda u_{\mathrm{eq}}) \tag{7.3}$$

and let $\lambda \rightarrow \infty$, i.e., let the sigmoid characteristics become steeper and steeper.

Proposition 7.1: Let all symbols be as defined above. Then

$$\lim_{\lambda \rightarrow \infty} \frac{\ln|e_i - v_{\mathrm{eq}i}|}{\ln|e_j - v_{\mathrm{eq}j}|} = \frac{|u_{\mathrm{eq}i}|}{|u_{\mathrm{eq}j}|}. \tag{7.4}$$

Proof: Suppose first that $u_i > 0$, $e_i = 1$ (by the parity condition). Then, as $\lambda \rightarrow \infty$, we have

$$u_{\mathrm{eq}i} = \frac{1}{1 + \exp(-\lambda u_{\mathrm{eq}i})} \approx 1 - \exp\,(-\lambda u_{\mathrm{eq}i}), \tag{7.5}$$

$$\ln\,(1 - v_{\mathrm{eq}i}) \approx -\lambda u_{\mathrm{eq}i}. \tag{7.6}$$

Now suppose $u_i < 0$, $e_i = 0$. Then, as $\lambda \rightarrow \infty$, we have

$$\begin{aligned} \ln v_{\mathrm{eq}i} = -\ln\,[1 + \exp\,(-\lambda u_{\mathrm{eq}i})] &\approx -\ln\,[\exp\,(-\lambda u_{\mathrm{eq}i})] \\ &= \lambda u_{\mathrm{eq}i}. \end{aligned} \tag{7.7}$$

The relationship (7.4) now follows readily from (7.6) and (7.7).

Proposition 7.1 address the issue of the rapidity with which V_{eq} approaches the corner e as $\lambda \rightarrow \infty$. Basically, the larger the value of $|u_{\mathrm{eq}i}|$, the more rapidly $v_{\mathrm{eq}i}$ approaches e_i. One can also explore the time behavior of the solution trajectories of (1.6) for a fixed "large" value of λ.

Proposition 7.2: Let all symbols be as defined earlier. Then

$$\lim_{\lambda \rightarrow \infty} \lim_{t \rightarrow \infty} \frac{\ln|v_i(t) - v_{\mathrm{eq}i}|}{\ln|v_j(t) - v_{\mathrm{eq}j}|} = \frac{\alpha_j}{\alpha_i}. \tag{7.8}$$

Proof: Suppose λ is "large" and that the initial condition $u_i(0)$ is "near" $u_{\mathrm{eq}i}$. Then it follows from (5.10) that

$$u_i(t) \approx u_{\mathrm{eq}i} + [u_i(0) - u_{\mathrm{eq}i}] \exp(-t/\alpha_i). \tag{7.9}$$

Suppose $u_{\mathrm{eq}i} > 0$. Then, in analogy with (7.5), we have

$$\begin{aligned} v_i(t) &= \frac{1}{1 + \exp\,[-\lambda u_i(t)]} \approx 1 - \exp\,[-\lambda u_i(t)] \\ &\approx 1 - \exp\left\{\lambda\left[u_{\mathrm{eq}i} + (u_i(0) - u_{\mathrm{eq}i})e^{-t/\alpha_i}\right]\right\} \\ &= 1 - \exp\,(\lambda u_{\mathrm{eq}i}) \cdot \exp\left[\lambda(u_i(0) - u_{\mathrm{eq}i})e^{-t/\alpha_i}\right] \\ &\approx 1 - \exp\,(\lambda u_{\mathrm{eq}i}) \cdot \left[1 + \lambda(u_i(0) - u_{\mathrm{eq}i})e^{-t/\alpha_i}\right] \\ &\approx v_{\mathrm{eq}i} - \lambda \exp\,(\lambda u_{\mathrm{eq}i})[u_i(0) - u_{\mathrm{eq}i}] \exp\,(-t/\alpha_i), \end{aligned} \tag{7.10}$$

$$\ln\,[|v_i(t) - v_{\mathrm{eq}i}|] \approx -\frac{t}{\alpha_i} + \lambda u_{\mathrm{eq}i} + \ln \lambda |u_i(0) - u_{\mathrm{eq}i}|. \tag{7.11}$$

As $t \rightarrow \infty$ for a fixed λ, the first term on the right side dominates the rest. A similar approximation applies when $u_{\mathrm{eq}i} < 0$. The desired result (7.8) now follows readily.

Proposition 7.2 shows that, in the case where all neurons are identical (i.e., $\alpha_i = \alpha$ for all i), the trajectory in the V-space converges to the equilibrium at essentially the same rate in all components.

TABLE II

e	z
$[1\ \ 0\ \ 1\ \ 0]^t$	$[0.9\ \ -1.2\ \ 9.6\ \ 27.2]^t$
$[0\ \ 1\ \ 1\ \ 0]^t$	$[-1.1\ \ 0.8\ \ 5.6\ \ -36.8]^t$
$[0\ \ 0\ \ 0\ \ 0]^t$	$[-3.1\ \ -7.2\ \ -18.4\ \ 11.2]^t$

Example 7.3: Consider again the four-bit A/D converter of Examples 4.2 and 5.3. Suppose the input x equals, say, 5.4. In this case, from Table I, one sees that there are three equilibria, namely at $e = [1\,0\,1\,0]^t = 5$, $[0\,1\,1\,0]^t = 6$, and $[0\,0\,0\,1]^t = 8$. Table II shows the corresponding values of $z = Te + I$.

From Table II one can see that, in two out of the three cases (in fact the two which represent the best digital approximations to the given input x), the components of z are smallest in magnitude corresponding to the least significant bits, and largest in magnitude corresponding to the most significant bits. Thus as the sigmoid nonlinearities become steeper and steeper ($\lambda \rightarrow \infty$), one would expect that the most significant bits to converge most rapidly to the "correct" values. The same phenomenon can be observed for almost all values of the input variable x. The details are routine and are left to the reader.

B. Existence of Equilibria in the Corners

Proposition 5.1 states that *if* the system (1.6) has any equilibria near the corners of $\bar{H}$, then these are asymptotically stable. But, under certain circumstances, there might be ***no*** equilibria near the corners of $\bar{H}$.

First a positive result.

Proposition 7.4: Suppose the interconnection matrix T satisfies the following conditions:

1) T is symmetric, and all of its diagonal elements are zero.
2) Every principal submatrix of T of size 2×2 or larger, including T itself, is hyperbolic and has at least one positive eigenvalue.

Under these conditions, for all inputs I except those belonging to a set of measure zero, there exists at least one binary vector $e \in \boldsymbol{b}^n$ such $Te + I$ has the same parity as e.

Proof: The assumptions ensure that the neural network exhibits total stability, i.e., every solution trajectory converges to an equilibrium [3]. Propositions 4.1, 6.1, and 6.2 show that there can be no asymptotically stable equilibria except near the corners of $\bar{H}$, while Proposition 6.2 guarantees that there can only be a finite number of equilibria in the faces of $\bar{H}$. All these facts plus total stability lead one to conclude that there must exist at least one asymptotically stable equilibrium near a corner of $\bar{H}$. By Proposition 5.1, this is equivalent to the parity condition being satisfied at some corner of $\bar{H}$. This is the desired conclusion.

Now an example to show that Proposition 7.4 is ***not*** valid if the interconnection matrix T is perturbed.

Example 7.5: Consider a two-neuron network with the interconnection matrix

$$T = \begin{bmatrix} -\epsilon & -1 \\ -1 & \epsilon \end{bmatrix}. \tag{7.12}$$

Then, applying Proposition 5.1, one can verify that if

$$0 < i_1 < \epsilon \qquad i_2 < -\epsilon \tag{7.13}$$

then *none of the four vectors* in $\boldsymbol{b}^2$ satisfies the parity condition. But by applying Proposition 6.2, one can see that there is an equilibrium near

$$e_1 = \frac{i_1}{\epsilon} \qquad e_2 = 0. \tag{7.14}$$

To determine the signature of this equilibrium, let

$$u_1 = g_1^{-1}(e_1) \qquad m_{11} = g_1'(u_1) > 0. \tag{7.15}$$

Then, by (6.25), the eigenvalues of the linearized system around the equilibrium are asymptotically equal to

$$\{-\lambda m_{11}, -\alpha_2\}. \tag{7.16}$$

Hence this equilibrium is asymptotically stable.

The point of Proposition 7.4 and Example 7.5 is as follows: Under ideal conditions, there is (almost) always an asymptotically stable equilibrium near a corner of $\bar{H}$. Since the parity condition of Proposition 5.1 is just an algebraic relationship, it is easy to see that, *for each fixed input vector I,* there is a small allowed perturbation such that there continues to exist an equilibrium near some corner of $\bar{H}$. But Example 7.5 shows that the order of the quantifiers cannot be interchanged: It is *not* true that there exists a small allowed perturbation for which there continues to exist an equilibrium near some corner of $\bar{H}$.

As a final comment, observe that the proof of Proposition 7.4 is quite round-about and unsatisfactory. The parity condition involves only linear algebra, and as such one would expect to be able to find a proof of the proposition based purely on linear algebra.

VIII. Conclusions

In this paper we have given a complete analysis of the location and stability of the equilibria, in the high-gain limit, of arbitrary nonlinear neural networks. The class of networks studied here is quite general and includes the standard Hopfield-type networks as a special case. The method of analysis does *not* depend on energy function arguments. As a result, the results presented here continue to hold even if the neural dynamics are slightly perturbed, in contrast with the results based on energy arguments.

In this paper, it has been assumed solely for notational convenience that the scaling factor λ is the same for all neurons; see (1.6). However, this assumption is not necessary in order to establish the results proved here. Consider the more general description (1.8), where each neuron has a separate scaling constant. Suppose these constants all approach infinity in such a way that they are all of the same order, i.e., suppose there exist positive constants σ and μ such that

$$\sigma \leq \lambda_i/\lambda_j \leq \mu \qquad \forall i,j \in \{1,\cdots,n\}. \tag{8.1}$$

Then all the results of the paper remain valid. The only modification needed in the proofs is to replace the *scalar* λ by the *diagonal matrix*

$$\Lambda = \mathrm{diag}\{\lambda_1,\cdots,\lambda_n\} \tag{8.2}$$

in appropriate places. The details are easy and are left to the reader. To repeat, the point is that all results remain valid provided all scale factors are of the same *order*—they need not all have the same *value.*

If a Hopfield-type neural network has symmetric interconnections, then [3] the network exhibits *total* stability, i.e., all solutions approach an equilibrium. This means, for example, that there are no nontrivial periodic solutions. This conclusion depends heavily on the ability to construct a total Lyapunov or energy function, and the energy function of [3] is only valid if the interconnection matrix is symmetric. Thus it is still an open question as to whether a network with "nearly" symmetric interconnections can exhibit limit cycles, and if so, under what conditions.

Another issue that is as yet unresolved, even in the symmetric interconnections case, is that of calculating (or at least estimating) the basin or domain of attraction of each asymptotically stable equilibrium, which we now know can only lie in the corners of the hypercube $\bar{H}$ if the interconnection matrix has zero diagonal elements. This is a topic for further research.

Acknowledgment

The author thanks Prof. B. Leung of the University of Waterloo for drawing his attention to this problem, and Prof. H. Schneider of the University of Wisconsin for his valuable help.

References

[1] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational capabilities," *Proc. Nat. Acad. Sci. U.S.,* vol. 79, pp. 2554–2558, 1982.

[2] ——, "Neurons with graded response have collective computational capabilities like those of two-state neurons," *Proc. Nat Acad. Sci. U.S.,* vol. 81, pp. 3088–3092, 1984.

[3] M.W. Hirsch, "Convergence in neural nets," in *Proc. Int. Joint Conf. Neural Networks,* vol. II, 1987, pp. 115–125.

[4] F.M.A. Salam, Y.-W Wang, and M.-R. Choi, "On the analysis of dynamic feedback neural networks," *IEEE Trans. Circuits Syst.,* vol. 38, pp. 196–201, 1991.

[5] J.J. Hopfield and D.W. Tank, "'Neural' computation of decision optimization problems," *Biological Cybern.,* vol. 52, pp. 141–152, 1985.

[6] D.W. Tank and J.J. Hopfield, "Simple 'neural' optimization networks: An A/D converter, signal decision circuit, and a linear programming circuit," *IEEE Trans. Circuits Syst.,* vol. CAS-33, pp. 533–541, 1986.

[7] M.W. Hirsch and S. Smale, *Differential Equations, Dynamical Systems, and Linear Algebra.* New York: Academic, 1974.

[8] A. Sard, "The measure of critical values of differentiable maps," *Bull. Amer. Math. Soc.,* vol. 48, pp. 883–890, 1942.

[9] F.R. Gantmacher, *Matrix Theory.* New York: Chelsea Publishers, 1959, vol. I.

[10] Y. Kamp and M. Hasler, *Recursive Neural Networks for Associative Memory.* Chichester, U.K.: Wiley, 1990.

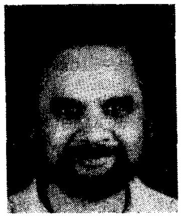
[11] M. Vidyasagar, *Nonlinear System Analysis.* Englewood Cliffs, NJ: Prentice-Hall, 1978.

[12] S. Amari, "Learning patterns and pattern sequences by self-organizing nets of threshold elements," *IEEE Trans. Comput.,* vol. C-21, pp. 1197–1206, 1972.

[13] L. Personnaz, I. Guyon, and G. Dreyfus, "Collective computational properties of neural networks: New learning mechanism," *Phys. Rev. A,* vol. 34, pp. 4217–4228, 1986.

[14] M. Vidyasagar, "Improved neural networks for analog to digital conversion," in *Proc. IJCNN,* Washington, DC, 1990, vol. III, pp. 517–522.

[15] M. Vidyasagar, "Improved neural networks for analog to digital conversion," *Circuits, Syst., Signal Processing,* vol. 11, no. 3, pp.387–398, 1992.



Mathukumalli Vidyasagar (S'69–M'69–SM'78–F'83) was born in Guntur, Andhra Pradesh, India, on September 29, 1947. He received the B.S., M.S., and Ph.D. degrees, all in electrical engineering, from the University of Wisconsin, in 1965, 1967, and 1969, respectively.

He taught at Marquette University, Milwaukee, WI, from 1969 to 1970, Concordia University, Canada, from 1970 to 1980, and the University of Waterloo, Canada, from 1980 to 1989. Since June 1989, he has been the Director of the Centre for Artificial Intelligence and Robotics (under the Defence Research and Development Organisation) in Bangalore. In addition he has held visiting positions at several universities including M.I.T., the University of California, Berkeley, and Los Angeles, C.N.R.S. Toulouse, France, and the Indian Institute of Science. He is the author or coauthor of six books and more than one hundred papers in archival journals. His current research interests are control theory, robotics, and neural networks.

Dr. Vidyasagar has received several honors in recognition of his research activities, and is a Fellow of the Indian Academy of Sciences.