Title
Convergent Activation Dynamics in ContinuousTime Networks

Author
Hirsch, Morris W

Publication Date
1989-01-01

Peer reviewed

Neural Networks, Vol. 2, pp. 331–349, 1989
Printed in the USA.

0893-6080/89 $3.00 + .00

REVIEW ARTICLE

Convergent Activation Dynamics in Continuous Time Networks

MORRIS W. HIRSCH

University of California

(Received and accepted 14 March 1989)

Abstract—*The activation dynamics of nets are considered from a rigorous mathematical point of view. A net is identified with the dynamical system defined by a continuously differentiable vector field on the space of activation vectors, with fixed weights, biases and inputs. Chaotic and oscillating nets are briefly discussed, but the main goal is to find conditions guaranteeing that the trajectory of every (or almost every) initial activation state converges to an equilibrium. Several new results of this type are proved. These are illustrated with applications to additive nets. Cascades of nets are considered and a cascade decomposition theorem is proved. An extension of the Cohen–Grossberg convergence theorem is proved for certain nets with nonsymmetric weight matrices.*

Keywords—Convergence, Cascade, Liapunov function, Excitatory net, Inhibitory net, Global asymptotic stability, Chaotic dynamics.

1. INTRODUCTION

A neural net with fixed weights is a dynamical system: given initial values of the activations of all the units, the future activations can be computed. This is the *activation dynamics*, with weights, biases, and inputs as parameters. On the other hand, there are many schemes for adaptively determining the weights of a network in order to achieve some particular kind of activation dynamics, for example, to classify input patterns in a particular way. Such a scheme determines a dynamical system in the space of weight matrices: this is the *weight dynamics*. A third, little explored possibility is to adapt the weights while running the activation dynamics. Such a procedure is a dynamical system in the Cartesian product of the weight space and the activation space.

In this article we consider activation dynamics from a rigorous mathematical point of view. We restrict attention to continuous time nets whose activation dynamics, with fixed weights, biases, and inputs, is governed by an autonomous system of ordinary differential equations defined by a continuously differentiable vector field. We identify a network with the dynamical system determined by such a system of differential equations.

The vast majority of all networks that have been simulated or theoretically analyzed have convergent activation dynamics: the trajectory of every initial state tends to some equilibrium. This is highly implausible behavior for biological networks whose units are nerve cells; but it may be descriptive of biological networks whose units are agglomerations of many nerve cells which tend to act coherently—such units have been variously termed cell assemblies, neuron pools, etc. On the other hand convergent networks have been designed to accomplish many interesting tasks, such as pattern recognition and classification, combinatorial optimization, conversion of printed documents to spoken words, and so forth. But no doubt the main reason for the common assumption of convergent dynamics is that it is exceedingly difficult to analyze or control the other kind.

In this article we present mathematical results which guarantee that the networks to which they apply are *convergent*: the orbit of every state converges to an equilibrium. Commonly this is done by finding a function which decreases on trajectories; but as we shall see, there are interesting nets where no such function is known, but which can be otherwise proved convergent. We also show that certain networks which may not be convergent are nevertheless "almost convergent."

A net has n units. To the ith unit we associate its *activation state* at time t, a real number $x_i = x_i(t)$; *output function* σ_i; bias θ_i; and *output signal* $R_i =$

Requests for reprints should be sent to Morris W. Hirsch, Department of Mathematics, University of California, Berkeley, CA 94720.

$\sigma_i(x_i + \theta_i)$. Usually we suppress notation for θ_i by incorporating it into σ_i.

The *weight* or connection strength on the line from unit j to unit i is a real number W_{ij}. When $W_{ij} = 0$ then there is no transmission line from unit j to unit i.

The *incoming signal* from unit j to unit i is $S_{ij} = W_{ij}R_j$. In addition there can be a vector I of any number of *external inputs* feeding into some or all units.

In all our nets the weights and biases are fixed.

The future activation states are assumed to be determined by a system of n differential equations of the form

$$\dot{x}_i = \frac{dx_i}{dt} = G_i(x_i, S_{i1}, \ldots, S_{in}, I);$$

$$i = 1, \ldots, n \quad (1)$$

where the independent variable t represents time. Written out in full this is

$$\dot{x}_i = G_i(x_i, W_{i1}\sigma_1(x_1 + \theta_1), \ldots, W_{in}\sigma_n(x_n + \theta_n),$$
$$I_1, \ldots, I_m), \quad i = 1, \ldots, n. \quad (2)$$

With the W_{ij}, θ_j and I_k assumed known we write this as

$$\dot{x}_i = F_i(x_1, \ldots, x_n), \quad i = 1, \ldots, n. \quad (3)$$

The output functions σ_j are taken to be continuously differentiable and nondecreasing: $\sigma_j' \geq 0$; occasionally we require the stronger condition $\sigma_j' > 0$. We also assume that the *state transition functions* G_i in (1) satisfy $\partial G_i/\partial S_{ij} > 0$; in other words, an increase in the weighted signal $W_{ij}\sigma_j(x_j)$ from unit j to unit i tends to increase the activation of unit i.

We shall often assume nonnegative outputs: $\sigma_j \geq 0$. In this case we interpret the condition $W_{ij} > 0$ as meaning that "unit j excites unit i", since an increase in the output σ_j will cause the activation x_i to rise if other outputs are held constant; similarly, $W_{ij} < 0$ means "unit j inhibits unit i".

Equations (3) represent the network in a particular coordinate system, called *network coordinates*. These coordinates $(x_1, \ldots, x_n)$ are convenient because x_i is the activation level of unit i of the network, and we shall generally work in these coordinates. However it is important to emphasize that we identify the network not with this particular system of equations, but with the *underlying dynamical system*.

By this system we mean the collection of mappings $\{\varphi_t\}_{t \in \mathbb{R}}$ defined as follows. For each $y \in \mathbb{R}^n$ there is a unique solution x to (3) with $x(0) = y$; we set $\varphi_t(y) = x(t)$. If we introduce new coordinates, this same dynamical system will probably be represented by different differential equations. Of course the variables in a different coordinate system will not be the activation levels; for example they are sometimes chosen to be the outputs, but in principle they can be any invertible function of the activation levels, chosen for mathematical convenience. We shall see examples of this in Section 10, where we replace some of the variables by their negatives in order to obtain differential equations of the special type called "cooperative."

All the dynamical features of solutions to (3)—convergence, attractors, limit cycles, and so forth—are invariant under coordinate changes: they are properties of the underlying dynamical system. A fundamental mathematical task is to deduce important dynamical properties from the form of the equations. While a great deal is known about the dynamics of certain classes of equations, no methods are known that apply to all equations.

Equations (2) do not include all systems that have been used to model neural networks—see for example (5) below—but they are reasonably general, and can be used to illustrate mathematical results that apply to most network equations. Many of the methods and results we describe also apply to more general nets with little change.

Often the external inputs are "clamped"—held constant—during a particular run of the activation dynamics. In this case *the inputs are parameters that determine the activation dynamics*. It is important to realize that changing the clamped inputs will change the dynamics. Thus for nets of this type we cannot properly speak of equilibria, attractors, and so forth without first specifying a particular input pattern.

In vector notation we write (3) as $\dot{x} = F(x)$; here F is the vector field on Euclidean space $\mathbb{R}^n$ whose ith component is F_i. We always assume that F is continuously differentiable. We shall tacitly assume that all vector fields dealt with are at least continuous and satisfy the usual theorems on existence, continuity and uniqueness of solutions. These theorems hold for continuously differentiable vector fields.[3]

[2] Although this seems an innocuous assumption, in the neural net literature one frequently comes across discontinuous vector fields, to which these theorems do not apply in general. Even continuity of a vector field, without further assumptions, does not imply uniqueness of solutions. Vector fields built out of step functions are often used to define activation dynamics. But they are generally not continuous, and the standard theorems on differential equations cannot be assumed to apply to them. Greater attention ought to be paid to this point.

[3] They also hold for vector fields which satisfy a local Lipschitz condition. These include any field whose component functions are constructed by starting with continuously differentiable functions and applying the following operations a finite number of times: taking the maximum or minimum of two functions; composing functions; and performing the usual arithmetic operations on functions. An example of such a field is obtained from system (1) by taking the output functions σ_i to be piecewise linear (and continuous), i.e., a function whose graph is connected and composed of a finite set of nonvertical straight segments or rays. A typical example of a piecewise linear function is a *ramp* function, whose connected graph is made up of two horizontal rays and one segment of positive slope.

A much studied class of network dynamics are the *additive nets*:

$$x_i' = -c_i x_i + \sum_j W_{ij}\sigma_j(x_j - \theta_j) + I_i, \quad i = 1, \ldots, n \tag{4}$$

with constant *decay rates* c_i and external inputs I_i (Amari, 1972, 1982; Aplevich, 1968; Cowan, 1967; Grossberg, 1969; Hopfield, 1984; Malsburg, 1973; Sejnowski, 1977). Our results will be illustrated by additive nets, although many of them apply more generally.

A closely related type of net is composed of units which are differentiable analogs of linear threshold elements. The dynamics are given by

$$y_i' = -c_i y_i + \rho_i\left(\sum_j W_{ij} y_j - \theta_i\right), \quad i = 1, \ldots, n \tag{5}$$

where each ρ_i is a sigmoid function. This system is not in the form (4). As has been noted by several authors, in case all the c_i are equal we can substitute $x_i = \sum_j W_{ij} y_j$ in (5) and obtain a system of type (4) with $\sigma_i = \rho_i$. When the weight matrix is invertible then the inverse transformation is also possible.[3] While many of the results given below for systems of type (4) can be recast in forms valid for (5), in this paper we do not consider (5) further.

2. INPUT AND OUTPUT

Consider a net represented by equation (1). In running the net we must specify the external input vector I and the initial activation vector $x(0)$. Both I and $x(0)$ are ways of feeding data into the net, but they play different dynamic roles. When I is specified the dynamics is determined, and $x(0)$ is the initial value of a trajectory. A different I determines a different dynamical system, whereas if I is held fixed, a different $x(0)$ is the initial value of another trajectory of the same dynamical system.

In the activation dynamics of feed-forward nets operating in discrete time, only the initial values of the input units is specified. This is because the initial values of the other units is irrelevant; their future values are functions of the inputs alone. But in a net governed by differential equations, even if it is feed-forward, *all* activations must be given initial values, because a solution of a differential equation is not determined until initial values of all variables have been specified. The initial values of the non-input units are generally reset to the same conventional value (usually zero) each time the net is run.

It is biologically more realistic *not* to reset the activations of the non-input units when inputs are changed, but rather to simply take as the new initial value whatever the activation level happens to be when the input is changed. This, however, greatly complicates the analysis of the net's behavior under a sequence of inputs.

To see how such a net might work, we suppose that for each input vector the dynamics is such that almost every initial value lies in the basin of some point attractor. After the first input vector $I_{(1)}$ is chosen, the activation is in some initial state z. Suppose that this state is in the basin of an attracting equilibrium $p = (p_1, \ldots, p_n)$. Under the dynamics determined by $I_{(1)}$, the trajectory of the state approaches p. Now we change to a second input $I_{(2)} \neq I_{(1)}$. The dynamics are now different, and p is probably not an equilibrium for the new dynamics. We assume p lies in the basin of some attractor $q \neq p$ for the dynamics corresponding to the new input $I_{(2)}$. The activation vector then tends to q. Suppose the third input $I_{(3)}$ coincides with the first: $I_{(3)} = I_{(1)}$. We are back in the same dynamical system as we started with, but we are computing the trajectory of the state q, rather than the state z which initialized the system. There is no guarantee that q and z are in the basin of the same attractor for the $I_{(1)}$ dynamics. If they are not, then the activation will evolve to some new attractor $r \neq p$. The upshot is that for a net of this type, run without resetting initial values, we cannot use the dynamics to define a mapping from inputs to attractors.

Evidently such a net cannot function as a classifier for the input patterns I, or as an associative content addressable memory. Instead it tends to behave like a rather unreliable finite state automaton, the states of the automaton being the various attracting equilibria. An interesting generalization of the supervised learning problem is the question of how to teach a network of this type to emulate a given automaton.

If the activation dynamics are globally asymptotically stable for every input vector, then the initial state doesn't matter, since for any fixed input, all trajectories tend to the same limit. Such nets realize a mapping sending each input to the corresponding equilibrium state. They are discussed in Section 8.

So far we have assumed the external input I is clamped. Alternatively, I may be a *single pulse*: $I(t)$ is specified during the time interval $0 \le t \le t_1$ and is clamped at zero (or some other conventional value) after time t_1. Thus the system has different dynamics for $t \le t_1$ and $t > t_1$. One way of using single pulse inputs is the following. Each input vector is such that it quickly drives $x(t)$—regardless of $x(0)$—to some desired region of activation space, for example, the basin of an attractor associated to the input. Then when the pulse is shut off, the activation vector tends to that attractor. For a net run in this mode the initial activation values are irrelevant, pro-

[3] This change of variables is interesting for the following reason: if the y_i in (5) denote the activations of a physical network, the "network" with activations $x_i = \sum_j W_{ij} y_j$ represented by (4) is only conceptual. This shows that it can be worthwhile studying network-type equations that do not correspond to a presumed real class of physical networks.

vided the input pulses are strong enough. Such a net can thus be run without resetting activations.

From now on we assume inputs are clamped.

3. CONVERGENT, OSCILLATORY, AND CHAOTIC DYNAMICS

In order to say anything interesting about the activation dynamics, the response functions G_i in equations (1) must be further restricted. At this point the relevant question is: What kind of dynamics do we expect of our net? In order to discuss this we consider three large categories of dynamical systems (assuming weights and external inputs are fixed):

- *Convergent*: Every trajectory $x(t) = (x_1(t), \ldots, x_n(t))$ converges to some *equilibrium* (stationary state) (as $t \to +\infty$);
- *Oscillatory*: Every trajectory is asymptotic to a periodic (perhaps stationary) orbit;
- *Chaotic*: Most trajectories do not tend to periodic orbits.[4] ("Most" must be made precise: for example, excluding a set of initial values having measure zero.)

Almost all, but not all, nets that have actually been simulated or analyzed are convergent (or assumed to be so). In particular, feed-forward nets are convergent (practically by definition).[5] The classical additive nets (4) are known to be convergent in certain cases: when the weight matrix $W = [W_{ij}]$ is symmetric (section 5), when the state transition functions G_i have a special algebraic form (section 6), and when the derivatives σ_i' and the weights obey certain inequalities (section 7). In certain cases cascades of convergent nets can be proved to be convergent (section 9). We explain in section 10 that excitatory nets are "almost convergent," as are certain inhibitory and other types of nets.

There is good reason for wanting nets to be convergent. If we think of the activation dynamics as eventually retrieving (or perhaps creating) information, it is natural to want this information to be in the form of a single unchanging n-tuple of numbers, that is, an equilibrium (stationary state) of the dynamical system.

We are so accustomed to storing data as numbers or symbols—discrete, constant entities—that it is hard to imagine any other way. Suppose, for example, that our net is oscillatory. We input an initial state, and the dynamics eventually homes in on a cycle (nonconstant periodic orbit). How can we retrieve the "information" embodied in this cycle? We might for example calculate the period of the cycle, or its amplitude, or some function of the Fourier coefficients of its components; or the average of some function over the cycle.

In fact there are fundamental differences between cycles and equilibria. In a finite dimensional state space any state is an equilibrium for some dynamics, thus the set of possible equilibria is identical with the state space, and thus is finite dimensional. But the set of possible cycles is a much richer, infinite dimensional space.

An equilibrium is a finite, static kind of object—mathematically, merely a point. But a cycle is not only an infinite set of points, it is an object whose dynamic interpretation necessarily involves *time*. Moreover the interpretation of an equilibrium is not highly dependent on the coordinate system; ordinarily it is a standard, usually routine mathematical task to change from one coordinate system to another, should this be desirable (e.g., in order to exhibit the local dynamics more perspicuously). But our interpretation of a cycle may be intimately tied to particular coordinates. In one coordinate system a periodic orbit can appear extremely simple, perhaps with only one nonzero Fourier component—simple harmonic motion—while this same orbit may appear very complex in other coordinates. These coordinates may be natural ones adapted to the network, and the complexity of the cycle may reflect accurately the behavior of the "real" system modeled by the mathematics. On the other hand it may be a merely mathematical artifact due to our choice of state variables, with no intrinsic significance.

An interesting problem arises if the output of our net serves as input to another net, as is often the case in biological systems. There is no difficulty about this if the first net is convergent. But what if it is oscillatory? The second net then has *oscillatory inputs*. The theory—or practice—of nets with oscillatory input does not exist, apart from the general subject of dynamical systems with oscillatory forcing.

[4] Chaotic dynamics is often defined more loosely to mean that long-term behavior of trajectories is extremely sensitive to initial values, but for present purposes the definition given here is convenient. On the other hand, R. Thom says, "'Chaos' and 'chaotic' should be reserved for systems that cannot be explicitly described either quantitatively or qualitatively (there are plenty of them). Hence such chaotic systems have no equations . . . It is to be expected that after the present initial period of word play, people will realize that the term 'chaos' has in itself very little explanatory power . . ." (Skarda & Freeman, 1987, p. 182). I agree wholeheartedly with the last sentence, but the concept of "systems that cannot be explicitly described"—needs to be more explicitly described!

[5] This is true for discrete-time net, but not for all continuous-time nets run by equation (1), unless $\partial G_i/\partial x_i = 0$. Consider a three-layer, three-unit net; the clamped input is a real number I, the activation x of the hidden layer is determined by $x' = f(x, I)$, and the output $y(t)$ is governed by $y' = g(x(t), y(t))$. Suppose $x(0)$ and $y(0)$ are specified. Since the system is assumed bounded, $x(t) \to$ constant. Without further assumptions, however, this does not imply $y(t) \to$ constant.

While there are many unanswered questions about convergent nets (and no doubt many interesting unasked questions!), there are several widely accepted methodologies for convergent nets. There is general agreement on what it means to input information to a net (in several ways); on how to read the output of a net; on how a net can function as an associative memory, a pattern analyzer, an optimizer, and so forth; on what supervised and unsupervised learning mean. There is no such general understanding as to how oscillatory nets might function.

One of the great successes of the study of neural nets has been the development of nets which can be analyzed through the use of well-understood mathematical methods: gradient descent, Liapunov functions, probability theory, linear algebra, group theory, dynamical systems theory, differential equations, combinatorics. For oscillatory nets much of the relevant mathematics is either not sufficiently developed or is too complex to be useful. For example, the existence of a strict Liapunov function is a simple and usable criterion for every trajectory to converge to equilibrium. But there is no known analogous criterion for every trajectory to converge to a cycle. The detection of equilibria for a given dynamical system amounts to solving a system of algebraic equations, and determining their stability properties is the problem of estimating eigenvalues of matrices. While the computations may be arduous, there are standard methods of carrying them out, and well-understood theories behind them. The analogous problems for cycles are extremely difficult, and there are no general methods—practical or theoretical—for locating cycles and determining their stability.

Oscillatory dynamics in network models of short-term memory were studied by Elias and Grossberg (Elias & Grossberg, 1975). Nets that store data as stable oscillations and methods for training their weights have been examined as models of the olfactory bulb by Baird (1986, 1989). Li and Hopfield (in press) have also studied storage of oscillations in such models.

More perplexing questions arise with chaotic nets. The limit set of a chaotic orbit is generally some sort of fractal; in what sense can it represent useful information? How do we retrieve information from a fractal? How can we use it as input to another net? How can we train the weights? In what sense can a chaotic network be stable? How can we determine if the net is actually chaotic?

On the other hand, there are good reasons for trying to understand and use oscillatory and chaotic nets. If we take seriously the basic creed of the neural net enterprise—that we have much to learn from the networks in the brain—then it is a striking experimental fact that brain dynamics have never been observed to be convergent, and are generally oscillatory or chaotic (see e.g., Freeman & Viana Di Prisco, 1986a, 1986b).

4. ARE CHAOTIC DYNAMICS BIOLOGICALLY USEFUL?

We digress to discuss an interesting paper by Skarda and Freeman (1987) who suggest ways in which chaotic dynamics might be useful, and even necessary, in the olfactory system of rabbits: "During late inhalation and early exhalation a surge of receptor input reaches the bulb, depolarizes the mitral cells, sensitizes the bulb, and induces an oscillatory burst. This is a bifurcation from a low-energy chaotic state to a high energy state with a narrow temporal spectral distribution of its energy, suggesting that it is governed by a limit cycle attractor." They suggest further that ". . . multiple limit cycle attractors exist, one for each odorant an animal has learned to discriminate behaviorally, and each one leading to regular oscillation in a burst." In this model the dynamics is chaotic in the absence of a recognized odor; upon intake of a previously learned odor the dynamics bifurcates, and the state vector finds itself in the basin of an attracting cycle of the new dynamics, corresponding to the particular odor.

How is it useful for the dynamics to be chaotic? Skarda and Freeman: "We conjecture that chaotic activity provides a way of exercising neurons that is guaranteed not to lead to cyclic entrainment or to spatially structured activity . . . It also allows rapid and unbiased access to every limit cycle attractor on every inhalation, so that the entire repertoire of learned discriminanda is available to the animal at all times for instantaneous access. There is no search through a memory store. Moreover the chaotic well during inhalation provides a catch-basin for failure of the mechanism to converge to a known attractor . . . the chaotic well provides an escape from all established attractors, so that an animal can classify an odorant as 'novel' with no greater delay than for the classification of any known sample, and it gains the freedom to maintain unstructured activity while building a new attractor . . . Most remarkably, 'signals' are not detected 'in' the chaos because the mechanism turns the chaos 'off' when it turns a signal 'on'." They go on to make the provocative suggestion that "without chaotic behavior the neural system cannot add a new odor to its repertoire of learned odors."

This is an interesting and original biological role for chaotic dynamics. But notice that the chaos is only in the background: When the rabbit sniffs a previously learned odor, the dynamics *bifurcates*—it radically changes its global behavior—and the state which was in a chaotic attractor for the old dynamics suddenly finds itself in a limit cycle attractor for the new dynamics.[6] The Skarda–Freeman scheme is thus based on the bifurcation of a chaotic attractor into several limit cycle attractors upon presentation of learned inputs. How the network learns this bifurcation is not explained.

It is worth emphasizing that the mathematical theory of chaotic attractors is in its infancy: there is very little rigorous mathematical treatment beyond bifurcation of equilibria and limit cycles. Concerning chaotic attractors there is much simulation but few theorems, or even conjectures; and consequently little in the way of predictive power. Even simulations are controversial: it is not clear that what is observed is truly a chaotic attractor; it might well be a complicated transient on its way to a limit cycle. In any case, omnipresent rounding and truncation errors, which are harmless for convergent dynamics, render dubious conclusions based solely on chaotic-looking simulations. It is extremely difficult to prove or disprove that a suspected attractor is in fact chaotic. There is not even a rudimentary classification. As Thom pointed out in his critique of Skarda and Freeman (note 4 above), "the invariants associated with the present theory—Liapunov exponents, Hausdorff dimension, Kolmogoroff-Sinai entropy . . . show little robustness in the presence of noise."

There are many cogent criticisms of Skarda and Freeman's innovative thesis in the same journal. Of these, Grossberg's is relevant to the dynamical issues. He points out that "a data phenomenon, despite its *correlation* with a particular functional property, may not be *necessary* to achieve that functional property. When this is true it is not possible to assert that the system has been *designed* to generate the property for that functional purpose. One can defeat the claim that the property in question is necessary by providing a mathematical counterexample of its necessity." Grossberg goes on to claim that the Adaptive Resonance Theory architecture ART I (Carpenter & Grossberg 1987) defeats Skarda and Freeman's claim of the necessity of chaos because (in our terminology) it is a convergent system, which exhibits with mathematical rigor the same behavior for which chaos is claimed to be necessary.

Skarda and Freeman suggest ways in which chaotic network dynamics may be useful and even necessary for efficient learning; this needs a great deal of investigation by computer simulation, mathematical analysis, and neurological experiment. The mathematical model they refer to (Freeman, 1987) is a complicated system of coupled nonlinear second-order differential equations, which has not been subjected to mathematical analysis. There is no theory of how the model works. The novel notion of a "chaotic well" bifurcating into limit cycle attractors in response to inputs is intriguing but imprecise. Most studies of bifurcation go the other way: they deal with bifurcation from limit cycles to chaos as energy (or temperature, etc.) increases. There is clearly a need for much research here.

Biology apart, there are good engineering reasons for investigating nonconvergent nets. The dynamical possibilities are much richer for a given number of units. What we don't yet know are useful ways of exploiting this wealth of dynamic behavior.

5. CONVERGENT DYNAMICS

Suppose we restrict attention to convergent nets. Why not stay with the old reliable feed-forward architectures, since these are guaranteed to converge? In part the answer has already been given. The brain is highly recurrent, and the repertoire of dynamic behavior is richer for recurrent nets. Feed-forward nets can not do competitive learning, for example, nor adaptive resonance. There is also some evidence that recurrent nets can learn more quickly (Almeida, 1987). Williams and Zipser (1988), and Servan-Schreiber and Cleeremans (1988) present learning algorithms for general recurrent nets (not necessarily convergent) that can accomplish rather complex temporal tasks.[7]

Thus it is highly desirable to have at hand criteria for activation dynamics to be convergent. It turns out that slightly weaker conditions are often easier to verify and practically as useful.

- A system is *almost convergent* if the set of initial values whose trajectories do not converge has Lebesgue measure zero—in other words a point picked at random has a convergent trajectory with probability 1. This does not exclude nonconvergent orbits, but it means they are exceptional, and we probably won't observe them since they cannot be stable.
- *Quasiconvergence* means that every trajectory approaches the set $\mathscr{E}$ of equilibria. Since the tangent vector to the trajectory approaches zero, this means that the velocity of every trajectory tends to zero. Any trajectory, after sufficient time has elapsed, will change only imperceptibly. To the observer the trajectory will appear to converge, although mathematically it does not necessarily do

[6] More recently Freeman has stated that he has been unable to find experimental evidence for such limit cycle attractors.

[7] Other studies of learning in recurrent nets are due to Pineda (1987), Rowher and Forrest (1987), Jordan (1987), Stornetta, Hogg, and Huberman (1987), Mozer (1988), Pearlmutter (1988), Elman (1988), Robinson and Fallside (1987), Bachrach (1988), and Rumelhart, Hinton, and Williams (1986).

so. In a quasiconvergent system there cannot be cycles or recurrent trajectories.

- *Almost quasiconvergence* means that the set of initial values whose trajectories are not quasiconvergent has measure zero. This is a combination of the last two conditions. It means that we are unlikely to observe a trajectory that does not appear to converge. There may be cycles or other kinds of nonconvergent orbits, but they cannot be stable.

We discuss below several types of dynamical systems relevant to neural nets, that can be shown to have these properties. For simplicity we shall tacitly assume that any system under consideration is *bounded*, that is, there is a bounded set A which attracts all trajectories: For any trajectory $x(t)$ there exists t_0 such that $x(t) \in A$ for all $t > t_0$. This is a natural requirement for applications, and can usually be proved without difficulty for specific models.

In a bounded system every (forward) trajectory $x(t)$ approaches a nonempty, closed, bounded, connected set of limit points. By a *limit point* p of the trajectory $x(t)$ we mean a point of the form $p = \lim_{i\to\infty} x(t_i)$ for some sequence of times $t_i \to \infty$. (More precisely, p is an *omega limit point*.) If $x(0) = q$ the set of limit points is called the *(omega) limit set* of the point q, denoted $\omega(q)$. All points on the orbit of q have the same limit set. The limit set is *invariant* under the dynamics, that is, if $y(t)$ is a trajectory that starts at a point of $y(0) \in \omega(q)$, then $y(t) \in \omega(q)$ for all t such that $y(t)$ is defined.

To say that a trajectory converges is equivalent to saying its limit set consists of a single equilibrium. When the limit set of a trajectory is a cycle, the orbit appears to be eventually periodic—in the long run it is indistinguishable from the cycle (although mathematically it is disjoint from the cycle, except in the case that the trajectory itself is periodic).

A fundamental dynamical concept is that of a stable equilibrium. An equilibrium p for a vector field H is characterized by $H(p) = 0$; it is *stable* if every eigenvalue of the linearized field $DH(p)$ has negative real part. This implies that trajectories starting near p converge to p at an exponential rate. It also implies that p is robust,[2] that is, any sufficiently small perturbation of H will have a stable equilibrium near p. When p is stable the *basin* of p is the union of the trajectories tending to p.

Another important type of equilibrium is a *hyperbolic equilibrium* p: this means that the eigenvalues of $DH(p)$ have nonzero real parts. This is a generic condition on vector fields. If H has a nonhyperbolic equilibrium, there are arbitrarily small perturbations of H whose equilibria are hyperbolic; while if H has only hyperbolic equilibria, so do all sufficiently small perturbations of H (see e.g., Hirsch & Smale, 1974). Of course this by itself guarantees nothing about a particular vector field. If p is a hyperbolic equilibrium then either p is stable, or else the set of trajectories tending to p has *measure zero*, and forms a smooth manifold of lower dimension than the state space. A hyperbolic equilibrium p is robust in the sense that any vector field sufficiently close to H must have a hyperbolic equilibrium near p.

An equilibrium p is *simple* if $DH(p)$ is invertible, that is 0 is not an eigenvalue. Hyperbolic equilibria are simple and robust. It is a generic condition for all equilibria to be simple. In that case equilibria are isolated, and since we always assume bounded dynamics, it follows that the equilibrium set $\mathcal{E}$ is finite.

In view of the fact that it is a generic condition for equilibria to be isolated, and under our assumptions, finite in number, it is a reasonable assumption, in the *absence* of contrary evidence, that any particular vector field we are dealing with has finite $\mathcal{E}$. This common dogma is less persuasive, however, if we are dealing with vector fields having a particular form, such as equation (1). It is in fact an interesting unresolved problem (probably not difficult) to prove or disprove for our general network equations (1), or for the more restricted additive network equations (4), that generically $\mathcal{E}$ is finite.

6. LIAPUNOV FUNCTIONS

One of the commonest ways to guarantee convergence is to find a *Liapunov function*, that is, a continuous function V on the state space which is nonincreasing along trajectories. Such a function is constant on the set of limit points of a trajectory. If V is a *strict* Liapunov function, meaning that V is strictly decreasing on nonstationary trajectories, then all limit points of any trajectory are equilibria (see e.g., Hirsch & Smale, 1974).

If V is a Liapunov function then any strictly increasing function of V is also a Liapunov function. Because we assume bounded dynamics, any Liapunov function V for our systems is necessarily bounded below; and we can obtain a bounded Liapunov function by composition, for example, $\arctan(V)$. If V is a bounded Liapunov function and $K > 0$ is a sufficiently large constant then $V + K$ will be a bounded positive Liapunov function which is strict if V is strict.

If F is a vector field on R^n and V is a continuously

[2] Robustness of a dynamic phenomenon is considered desirable in mathematical models of natural processes, since physical constants can never be measured with mathematical exactness, and consequently there is always uncertainty in the dynamical equations. Nonrobust phenomena are often thought of as being unobservable, not physically meaningful, and so forth.

differentiable real valued function on $\mathbb{R}^n$, then the chain rule shows that if $x(t)$ is a trajectory then

$$\frac{d}{dt}V(x(t)) = \nabla V(x(t)) \cdot F(x(t)),$$

where ∇V is the gradient vector field of V and the dot means the usual inner product. Therefore V is a Liapunov function if and only if $\nabla V \cdot F \leq 0$ everywhere; and V is strict if an only if $\nabla V(z) \cdot F(z) < 0$ at every point z such that $F(z) \neq 0$. The geometric interpretation of this last condition is the following. Suppose $F(z) \neq 0$; set $V(z) = c$. Then the vector $F(z)$ is transverse to the level surface $V^{-1}(c)$ at z, and points toward the set where $V < c$.

A strict Liapunov function forces every trajectory to approach asymptotically a set of equilibria; the system is thus quasiconvergent. If the S is finite or countably infinite (or more generally, totally disconnected), then a strict Liapunov function causes every trajectory to have a unique limit point (necessarily an equilibrium); in other words the system is convergent.[*]

Even with non-strict Liapunov functions it is often possible to guarantee quasiconvergence. This method, called *LaSalle's invariance principle* (LaSalle, 1968), is based on the fact that the limit set of a trajectory is contained in the largest invariant set S in which the Liapunov function is constant on orbits. Sometimes one can show that this set is composed of equilibria; then the system is quasiconvergent. If S is discrete then the system is convergent. LaSalle's invariance principle (for discrete-time systems) has been used by Golden (1986) to prove convergence for the Brain-State-In-A-Box. Cohen and Grossberg (1983) have applied the invariance principle in connection with their study of Liapunov functions, discussed below.

There is unfortunately no general method for constructing Liapunov functions, or for recognizing systems that have one. The following remarks describe some common situations where Liapunov functions are known.

In dissipative mechanical systems, energy is (by definition) a strict Liapunov function; hence Liapunov functions are sometimes called energy functions. Entropy is a strict Liapunov function in classical thermodynamical systems. For a gradient system $\dot{x}_i = -\partial U/\partial x_i$ the real valued function U on the state space is a strict Liapunov function: By the chain rule, $dU(x(t))/dt = \Sigma(\partial U/\partial x_i)\dot{x}_i = -\Sigma(\partial U/\partial x_i)^2$; this is negative unless $x(t)$ is an equilibrium. In many adaptive learning systems an error function is constructed so as to be a Liapunov function for the weight dynamics; in fact many algorithms for adapting weights are approximations to gradient descent on the error function. If a vector field F can be written in the form $F(x) = \rho(x)G(x)$ in such a way that ρ is a positive continuous function on the state space, and G is a vector field that admits a Liapunov function $V(x)$, then V is also a Liapunov function for F: this is because the trajectories of G are simply reparameterizations of those of F.

An early use of Liapunov functions in ecological systems is due to MacArthur (1969) for Gause-Lotka-Volterra systems of interacting species having symmetric community matrices. Cohen and Grossberg (1983) greatly extended this results by constructing Liapunov functions for all systems of the form

$$\dot{x}_i = a_i(x_i)[b_i(x_i) - \Sigma_k c_{ik} d_k(x_k)] = F_i(x) \qquad (6)$$

where $a_i \geq 0$, the constant matrix $[c_{ik}]$ is symmetric, and $d_i' \geq 0$. In this system we can assume $c_{ii} = 0$, since the term $c_{ii}d_i(x_i)$ can be absorbed into $b_i(x_i)$.

Special cases of system (6) have often been used to represent neural networks: x_i is the activity level of unit i; $d_k(x_k)$ is the output of unit k; c_{ik} is the strength (weight) of the connection from unit k to unit i; $a_i(x_i)$ is an amplification factor. If we suppose all x_i and d_j are ≥ 0, then the connection from unit k to unit i is inhibitory if $c_{ik} > 0$ and excitatory if $c_{ik} < 0$. By assumption these relationships are symmetric. The sum in (6) represents the net input to unit i. If the amplification factor is positive, equation (6) means that the activity of unit i decreases if and only if the net input to unit i exceeds a certain intrinsic function b_i of the unit's activation. If all connections between different units are inhibitory then we can think of the units as competing among themselves, the competition being modulated by the state-dependent amplification factors a_i, the self-excitement rates b_i, and the inhibitory interactions $c_{ik}d_k$.

The Liapunov function discovered by Cohen and Grossberg for system (6) is

$$V(x) = -\Sigma_i \int_0^{x_i} b_i(\xi)d_i'(\xi)d\xi + \tfrac{1}{2}\Sigma_{j,k} c_{jk}d_j(x_j)d_k(x_k) \qquad (7)$$

They showed that if $a_i > 0$ and $d_i' > 0$, then V is a strict Liapunov function, and therefore the system is quasiconvergent. Using LaSalle's invariance principle they showed this also holds in certain more general circumstances.

Essentially the same Liapunov function for a special case of (3) was given by Hopfield (1984), where

$$F_i(x) = -c_i x_i + \Sigma_j T_{ij} g(x_j); \qquad (8)$$

here $[T_{ij}]$ is a constant symmetric matrix and $g' \geq 0$. The Liapunov function is $-\frac{1}{2}\Sigma_i c_i x_i^2 + \frac{1}{2}\Sigma_{jk} T_{jk} g(x_j)g(x_k)$; in Hopfield's electrical circuit interpretation this is

[*] There is a theorem, for which I do not know a reference, that in an analytic gradient system every bounded trajectory converges, regardless of the nature of ε.

exactly the energy. Cohen and Grossberg also refer to this model.

A generalization of the Cohen–Grossberg theorem to certain nonsymmetric nets is given in Section 10, Theorem 7.

There is a little-known stability property for a dynamical system (3) having a strict Liapunov function and *isolated* equilibria. Not only does every trajectory converge to an equilibrium; but even if we allow arbitrary errors to perturb the trajectory, provided they are small enough, every limit point of the perturbed trajectory will be close to an equilibrium. To state this precisely we use the following definition. A (r, δ)-*perturbed* trajectory is a map (possibly discontinuous!) $y: [0, \infty) \to \mathbf{R}^n$ such that the following conditions hold: There is an increasing sequence $T_j \to \infty$ such that $T_{j+1} - T_j > r > 0$ for all $j = 1, 2, \ldots$, and solutions $x^{(j)}(t)$ such that for all j we have $\|y(t) - x^{(j)}(t)\| < \delta$ for $T_j \le t \le T_{j+1}$.

Theorem 1. Suppose equilibria are isolated and there is a strict Liapunov function V. Then for any $r > 0$, $\varepsilon > 0$ there exists $\delta_0 > 0$, depending on r and ε, such that if $0 < \delta < \delta_0$ then any limit point of a (r, δ)-perturbed trajectory $y(t)$ is within ε of an equilibrium p. Moreover if $y(0)$ is in the basin of a stable equilibrium q, then for sufficiently small δ we can take $p = q$.

This result is false if the assumption of a strict Liapunov function is deleted, and instead it is merely assumed that every trajectory converges. Theorem 1 gives a theoretical robustness to the dynamics of systems to which it applies. The perturbations, subject to the conditions of the theorem, can be otherwise completely arbitrary: for example, due to rounding or truncation error in a numerical simulation, noisy inputs, errors in estimating system parameters, etc. No statistical assumptions are needed.

The proof uses the boundedness of the system to ensure that the image of y is bounded, and that outside any neighborhood N of the equilibria, V decreases by at least some fixed number $\lambda > 0$ along any trajectory on an interval of length $\ge r$. By taking δ small we can ensure that $V(y(t)) - V(x^{(j)}(t))$ is less than $\lambda/2$ for $T_j \le t \le T_{j+1}$. This implies that there is an upper bound to the number of successive intervals $[T_j, T_{j+1}]$ whose images under y is disjoint from N. If N is the union of balls around equilibria of small radius, V cannot change by much along trajectory curves inside the balls. The upshot is that any limit point of $y(t)$ must be inside one of the balls.

Golden (1988) has shown how a broad class of nets with strict Liapunov functions respond to inputs as if they are maximizing *a posteriori* estimates of the probability distribution of the environment. This gives an interesting psychological interpretation of network dynamics.

7. A CONVERGENCE THEOREM WITHOUT LIAPUNOV FUNCTIONS

Grossberg (1978) (reprinted in Grossberg, 1982) proved a remarkable convergence theorem for a class of competitive systems for which no Liapunov functions are known; these are the systems of the form

$$\dot{x}_i = a_i(x)[b_i(x_i) - C(x_1, \ldots, x_n)] \qquad (9)$$

where $a_i > 0$ and $\partial C/\partial x_i > 0$ for all i. (His theorem is also valid if $\partial C/\partial x_i < 0$ for all i.) Notice that each b_i is a function of only the one variable x_i, and that the function $C: \mathbf{R}^n \to \mathbf{R}$ does not depend on i. In this kind of a system the x_i compete indirectly with each other through the medium of a scalar "field" $C(x)$ created by the interaction of all the x_i.

Grossberg showed that if the functions b_i are piecewise monotone then system (9) is convergent. Even without piecewise monotonicity, it can be proved that the system is quasiconvergent.

A simple example of a system (9) is

$$\dot{x}_i = c_i x_i (B_i - x_i - K \textstyle\sum \sigma_j(x_j)), \qquad 0 \le x_i \le B_i \qquad (10)$$

with $\sigma_j' > 0$ and positive constants c_i, B_i and K. This represents a special kind of completely interconnected network in which all weights equal $-K$, thus all connections are inhibitory, including the self-connections. If all connections were severed (i.e., $K = 0$), each nonzero activation would rise to its upper limit B_i, since it would obey $\dot{x}_i = c_i x_i(B_i - x_i)$. The connections serve to inhibit activations by means of the field term $-K\sum \sigma_j(x_j)$, negatively proportional to the total output signal.

Grossberg's result has been extended (Hirsch, 1980) to mildly nonautonomous systems of the form

$$\dot{x}_i = a_i(t, x)[b_i(x_i) - C(t, x_1, \ldots, x_n)] \qquad (11)$$

where $a_i > 0$, a_i and $1/a_i$ are uniformly bounded in t for each x, the partial derivatives of C are all positive or all negative, and $\partial C/\partial t \to 0$. Here the amplification factors are allowed to depend on time in a bounded way. This is one of the few examples of a convergence theorem for nonautonomous systems.

8. GLOBAL ASYMPTOTIC STABILITY

A system is *globally convergent* if there is a unique equilibrium to which everything converges. If in addition the equilibrium is stable, the system is called *globally asymptotically stable*. This concept is interesting for nets whose dynamical equations have the form

$$\dot{x}_i = F_i(x_1, \ldots, x_n; I_1, \ldots, I_n) \qquad (12)$$

where the I_i are clamped external inputs. If the system is globally convergent for each input vector I, then we need not specify initial values of the x_i, since all trajectories end up at the same unique equilibrium tion. This equilibrium depends on I. In this way we obtain a mapping from the space of input vectors I to the space of activation vectors x. Moreover we need not reset the activations when changing inputs, which is convenient for a system running in real time. Additive networks that run in this way have been considered by Kelly (in press).

There is a conceptually simple condition on a dynamical system $\dot{x} = F(x)$ which guarantees global asymptotic stability. Let $\langle \xi, \eta \rangle$ denote the inner (dot) product of vectors ξ, η; the square of the Euclidean norm of ξ is $\|\xi\|^2 = \langle \xi, \xi \rangle$.

Theorem 2. Assume there is a constant $\mu > 0$ such that each Jacobian matrix $A = DF(y)$ has the property that $\langle A\xi, \xi \rangle \le -\mu\|\xi\|^2$ for all $\xi \in \mathbf{R}^n$. Then the dynamical system $\dot{x} = F(x)$ is globally asymptotically stable.

The idea of the proof is to use the Taylor expansion of F to get the following estimate for the distance between solutions.

$$\frac{1}{2}\frac{d}{dt}\|x(t) - y(t)\|^2 = \langle F(x) - F(y), x - y \rangle$$
$$= \langle DF(y)(x - y) + R(x, y), x - y \rangle.$$

If $x(0)$, $y(0)$ are close and $t \ge 0$ is bounded we can assume $\|R(x, y)\| \le \varepsilon\|x - y\|$ for any given $\varepsilon > 0$. We take $\varepsilon < \frac{1}{2}\mu$ and get

$$\frac{1}{2}\frac{d}{dt}\|x(t) - y(t)\|^2 \le \langle DF(y)(x - y), x - y \rangle$$
$$+ \frac{1}{2}\mu\|x - y\|^2 \le -\mu\|x - y\|^2$$
$$+ \frac{1}{2}\mu\|x - y\|^2 = -\frac{1}{2}\mu\|x - y\|^2.$$

This implies that $\|x(t) - y(t)\|$ decreases exponentially if $x(0)$ and $y(0)$ are close. Closeness is in fact not necessary, since there is a finite sequence from $x(0)$ to $y(0)$ in which successive pairs are close. Taking $y(0) = x(s)$ for some $s > 0$ leads to a proof that $x(t)$ converges to an equilibrium p. This equilibrium must be unique, since any two trajectories are mutually asymptotic. It also follows that p is asymptotically stable. Thus there is global asymptotic stability.

The condition $\langle A\xi, \xi \rangle \le -\mu\|\xi\|^2$ for all $\xi \in \mathbf{R}^n$ on a matrix A is equivalent to the largest eigenvalue of $\frac{1}{2}(A + A^T)$ being $\le -\mu$, where A^T denotes the transpose of A. By Gerschgorin's circle theorem (Noble & Daniel, 1988) this is implied by the condition

$$A_{ii} + \frac{1}{2}\sum_{j \ne i} |A_{ij} + A_{ji}| \le -\mu, \quad i = 1, \ldots, n. \tag{13}$$

As an example we consider an additive net

$$\dot{x}_i = -c_i x_i + \sum_j W_{ij}\sigma_j(x_j) + I_i = F_i(x_1, \ldots, x_n). \tag{14}$$

Assume that $c_i > 0$ and

$$0 < \sigma_j' \le \gamma \quad \text{for all } j.$$

Now

$$\partial F_i/\partial x_i = -c_i + W_{ii}\sigma_i'(x_i) \le -c_i + \gamma W_{ii}$$

and

$$|\partial F_i/\partial x_j| = |W_{ij}\sigma_j'(x_j)| \le \gamma|W_{ij}| \quad \text{for } i \ne j.$$

Therefore we see from Gerschgorin's condition (13) that the inequality in Theorem 2 will hold for all Jacobian matrices $DF(x)$ provided

$$-c_i + \gamma W_{ii} + \frac{1}{2}\sum_{j \ne i}\gamma(|W_{ij}| + |W_{ji}|) \le -\mu \quad \text{for all } i.$$

This proves:

Theorem 3. System (14) is globally asymptotically stable, for any inputs I_i, provided there is a constant $\gamma \ge 0$ such that for all i

$$0 < \sigma_i' \le \gamma \quad \text{and} \quad \gamma\Big(W_{ii} + \frac{1}{2}\sum_{j \ne i}(|W_{ij}| + |W_{ji}|)\Big) < c_i. \tag{15}$$

Thus global asymptotic stability can be guaranteed by choosing transfer functions σ_i having gains σ_i' sufficiently small relative to the self-inhibitions c_i, or alternatively, by making each self-weight W_{ii} sufficiently negative relative to the absolute values of the other weights on lines connected to unit i. Of course in any specific case it may turn out that these conditions conflict with other constraints on the network, or with algorithms for choosing the weights.

One way of using condition (15) is to assume uniform bounds on the weights, the self-inhibitions c_i, and on the *connectivity* m = maximum number of other units any unit is connected to. Assume that for all distinct i, j with $W_{ij} \ne 0$, we have in (14)

$$0 < \sigma_i' \le \gamma, \quad W_{ii} \le \beta, \quad |W_{ij}| \le \alpha, \quad c_i \ge \delta > 0;$$

then (15) holds provided

$$\gamma(\beta + m\alpha) < \delta. \tag{16}$$

Thus condition (16) implies global asymptotic stability of the additive net (14), and it depends only on local properties of the network. Therefore it has the important virtue of being *independent of the number of units*.

A globally asymptotically stable system has a strict Liapunov function—but we construct it after the fact: In order to know there is a Liapunov function we need to know the system is globally asymptotically stable. Nevertheless the existence of a strict Liapunov function may be useful in determining convergence of a cascade in which the net is a component; see section 10. For each initial value x, let $V(x)$ denote the length of the trajectory $y(t)$ which starts at $x = y(0)$:

$$V(x) = \int_0^\infty \|F(y(s))\|ds$$

It is not hard to show that V is finite[14], and that

$$\frac{d}{dt} V(x(t)) = -|F(x(t))|,$$

proving that V is a strict Liapunov function.

Kelly (1988) gives a different criterion for global asymptotic stability of (14). Define the *operator norm* $\|W\|$ of the weight matrix W to be maximum of $|Wx|$ taken over all unit vectors x. Kelly shows that *if all $c_i = 1$ and $\gamma\|W\| < 1$, then* (14) *is globally asymptotically stable,*[15] *and the function $|x - p|^2$ is a strict Liapunov function for system* (14), where p is the unique equilibrium.

9. CASCADES

In studying the activation dynamics of a net it is often useful to decompose it into simpler subnets, and then try to understand the qualitative dynamics of the full net in terms of the dynamics of the subnets. The dynamics of feed-forward nets, for example, can be analyzed in terms of the dynamics of the individual units.

A recurrent net may be known to have convergent dynamics; for example it may have a strict Liapunov function. Consider a *layered* net $\mathcal{N}$. It is built up from a collection of subnets $\mathcal{N}_0, \mathcal{N}_1, \ldots$ in such a way that the output of $\mathcal{N}_{k-1}$ is fed only into $\mathcal{N}_k$. Suppose it is known that each $\mathcal{N}_k$ has convergent dynamics. Does this imply that the full net $\mathcal{N}$ also has convergent dynamics? Not in full generality. But we shall see several cases where this can be proved.

A generalization of a layered net is a *cascade*. Let $\mathcal{N}_1$ and $\mathcal{N}_2$ be two separate nets. If some units of $\mathcal{N}_1$ feed their outputs to units in $\mathcal{N}_2$ via new connections, we obtain a larger net $\mathcal{N}_{(2)}$, called a cascade of $\mathcal{N}_1$ into $\mathcal{N}_2$. If outputs from $\mathcal{N}_{(2)}$ are fed into a third net $\mathcal{N}_3$, separate from $\mathcal{N}_{(2)}$, we obtain a net $\mathcal{N}_{(3)}$, a cascade of $\mathcal{N}_{(2)}$ into $\mathcal{N}_3$. By iterating this process we obtain cascades of any number of nets $\mathcal{N}_1, \mathcal{N}_2, \mathcal{N}_3, \mathcal{N}_4, \ldots$. (We may think of a cascade as a feed-forward supernet whose superunits are nets.) For example each $\mathcal{N}_j$ might be a recurrent net doing competitive learning, feeding its output to $\mathcal{N}_k$, $k > j$. A net $\mathcal{N}$ obtained in this way is called the cascade of the *components* $\mathcal{N}_k$. A basic problem is to understand the behavior of a cascade in terms of the behavior of its component subnets.

We call a net *irreducible* if every pair of distinct units belongs to a loop of directed transmission lines, or in other words, if every unit can directly or indirectly influence the output of every other unit. A net that is not irreducible is called *reducible*. A feed-forward net with more than one unit is reducible to one-unit nets. Every cascade is by definition reducible.

A maximal irreducible subnet of a given net is called a *basic subnet*. It is easy to see that every irreducible subnet of a given net is contained in a unique basic subnet. In Appendix A we prove the following result.

Theorem 4. Every reducible net $\mathcal{N}$ is a cascade whose components are the basic subnets of $\mathcal{N}$.

The irreducibility of a net represented by equations (2) can be expressed in terms of the weight matrix $W = W_{ij}$. The net is irreducible if and only if W is an *irreducible matrix* in the following sense: A square matrix $[M_{ij}]$ is *irreducible* if for any pair of distinct indices i and j we can find a chain of indices $i = k_0, \ldots, k_m = j$ such that, if $k_r = a$ and $k_{r+1} = b$ then $M_{ab} \neq 0$. Equivalently, the linear map determined by M does not have any proper, nontrivial invariant subspace obtained by equating some set of coordinates to zero. Another equivalent formulation is that there is no way of permuting the coordinates to give this linear transformation a matrix with a square block in the upper left corner with only zeroes beneath it.

To test a matrix M for irreducibility, draw a directed graph with one vertex for each row of M, and an arrow from vertex j to vertex i if and only if $M_{ij} \neq 0$. If M is the weight matrix of a net then this graph is just the flow chart of the net. M is irreducible if and only if for pair i, j of distinct vertices there is a directed path of edges from i to j; or equivalently, i and j belong to a loop of directed edges.

If a net is reducible its units can be ordered so that the weight matrix is in *lower block triangular form*, square submatrices down the diagonal, zeroes above them, arbitrary entries below.

10. CONVERGENT CASCADES

It is frequently useful to know whether some particular property shared by all the components of the cascade $\mathcal{N}$ is also true for $\mathcal{N}$ itself. Here we inquire: If each net in a cascade has convergent dynamics, does the whole cascade have convergent dynamics? If each net in the cascade is globally asymptotically stable, is this true of the whole cascade?

The second question has an easy answer: Yes. It is convenient to formulate this result more generally, for vector fields with parameters (which are math-

[14] This uses the assumption that the Jacobian matrix at the equilibrium has eigenvalues with negative real parts, which prevents the orbit from wiggling too much as it approaches the equilibrium.

[15] This result is closely related to Theorem 3. It is easy to see that every eigenvalue of $\frac{1}{2}(W + W^T)$ has absolute value $\leq \|W\|$; if $c_i = 1$ for all i, calculation of $DF(x)$ shows that Kelly's assumption implies the hypothesis of Theorem 3.

ematical models of systems with inputs). An analogous result holds for discrete-time systems.

Let F be a vector field on $\mathbb{R}^m$, G a map from $\mathbb{R}^m \times \mathbb{R}^n$ to $\mathbb{R}^n$, and consider the dynamical system

$$\dot{x} = F(x), \quad \dot{y} = G(x, y) \tag{17}$$

This is the cascade of the two systems $\dot{x} = F(x)$ and $\dot{u} = G(\xi, u)$, where ξ is a parameter for the latter system.

More complex cascades of dynamical systems with parameters can be built by iterating this construction. Let $E^1, \ldots, E^r$ denote Euclidean spaces of various dimensions. For each $i = 1, \ldots, r$ let $F^i: E^1 \times \ldots \times E^i \to E^i$ be a map, thought of as a vector field on E^i with input parameters from $E^1 \times \ldots \times E^{i-1}$. The cascade of this family $\{F^1, \ldots, F^r\}$ is the following dynamical system on state space $E^1 \times \ldots \times E^r$, where $\mathbf{x}^i$ denotes a vector in E^i:

$$\dot{\mathbf{x}}^1 = F^1(\mathbf{x}^1),$$

$$\dot{\mathbf{x}}^i = F^i(\mathbf{x}^1, \ldots, \mathbf{x}^i), \quad i = 1, \ldots, r.$$

Theorem 5. A cascade of systems, each of which is globally asymptotically stable for every parameter value, is globally asymptotically stable.

Proof. It suffices to consider a cascade of two systems, as in (17). Let $z(t) = (x(t), y(t))$ be a trajectory of (17). Since F is globally asymptotically stable, $x(t)$ converges to the unique equilibrium p for F. Therefore the limit set of $z(t)$ is a closed bounded set in $p \times \mathbb{R}^n$, invariant under the dynamics of $\dot{y} = G(p, y)$. Since this system is globally asymptotically stable it has a unique equilibrium q. Now the only compact nonempty invariant set of a globally asymptotically stable system is the equilibrium. Therefore the limit set of $z(t)$ is (p, q). Since p and q are asymptotically stable, so is (p, q). QED

This proof also shows that if F is merely convergent, while G is globally asymptotically stable for every parameter value, then the cascade (17) is convergent. A similar result holds for almost convergence.

It is not true that a cascade of convergent systems is necessarily convergent, or even almost convergent. To achieve convergence we need special assumptions.

As an example of what can be proved, suppose that the dynamics of F are convergent; that for each fixed equilibrium p for F there is a strict Liapunov function for the system $\dot{y} = G(p, y)$, and that $G(p, y)$ has only a finite number of equilibria. Then the cascade (17) is convergent. To see this let $z(t) = (x(t), y(t))$ be a solution with $x(t) \to p$. The limit set K of $z(t)$ is an invariant set for the dynamics of $G(p, y)$ which has the property of being chain recurrent (Conley, 1978). The definition of chain recurrent is not needed here, but only the fact that a strict Liapunov function is constant on any chain recurrent set whose equilibrium set is finite. Therefore K must consist entirely of equilibria, and hence (being connected) of a single equilibrium. Thus (17) is convergent. If instead of being convergent F is merely almost convergent, the same argument shows that (17) is almost convergent.

This last result can be iterated and applied to cascades of arbitrary networks whose activation dynamics are of the form (1), provided that each component of the cascade has a finite equilibrium set, and admits a strict Liapunov function for every set of values of the inputs I_i. Under these conditions any such cascade will be convergent.

We now present a general way of constructing a strict Liapunov function for the system (17), assuming that F has one, and that for each equilibrium p of F, the vector field $G_p(y) = G(p, y)$ has one, provided equilibria are hyperbolic.

Theorem 6. In system (17) assume equilibria of F are isolated, and equilibria of G_p are hyperbolic for each equilibrium p of F. Suppose that F has a C^1 (continuously differentiable) strict Liapunov function $V(x)$, and that there is a C^2 strict Liapunov function for each G_p. Then there is a C^1 strict Liapunov function for (17).

Proof (with help from Michael Cohen). Let (p, q) be an equilibrium, taken to be $(0, 0)$ for simplicity. By the implicit function theorem we may assume $G(x, q) = 0$ for all x near p. Using hyperbolicity it is possible to linearly change coordinates so that $D_yG(p, q) = T + R$, where T is a nonsingular diagonal matrix, and there are constants $\tau > 0$ and arbitrarily small $\beta > 0$ such that

$$|\langle Ty, y\rangle| \geq \tau|y|^2, \quad |\langle Ry, y\rangle| \leq \beta|y|^2$$

Let $U(y)$ be a bounded strict Liapunov function for G_ξ. For each ξ near enough to p, the function $W(y) = U(q) - \langle Ty, y\rangle$ is another strict Liapunov function for G_ξ in a neighborhood of q. Now define

$$\tilde{U}(y) = (1 - \gamma r(y))W(y) + \gamma r(y)U(y)$$

where $\gamma > 0$ is a small constant, and r is 1 outside a neighborhood of q and 0 inside a smaller neighborhood of q. A calculation shows that if β and γ are small enough and U is C^2 then $\tilde{U}(y)$ is a strict Liapunov function for G_ξ, for each ξ near p.

Let the solutions to $G(p, y) = 0$ be $q_1, \ldots, q_k$. For each q_i let $\tilde{U}_i$ be defined similarly to $\tilde{U}$, using q_i in place of q. Then the function $W(y) = \Sigma_i \tilde{U}_i(y)$ is a strict Liapunov function for G_ξ for each ξ near p.

Let $V(x)$ be a bounded strict Liapunov function for F. Let ρ be a C^1 real-valued function on $\mathbb{R}^n$ taking the value 1 on a neighborhood N of p, and the value 0 outside a larger, bounded neighborhood N' of p containing no other equilibrium. Pick $\alpha > 0$, to be specified later, and define the function $L(x, y) = V(x) + \delta\rho(x)W(y)$. We show that if δ is small enough then L is a Liapunov function, strict for x near p.

Let $H(x, y) = (F(x), G(x, y))$ denote the right hand side of (17). The derivative of L along a trajectory of H is

$$\dot L = \nabla L \cdot H = \nabla_x V \cdot F + \delta(\nabla_x \rho \cdot F)W + \delta\rho \nabla_y W \cdot G \qquad (18)$$

If we evaluate $\dot L$ at a point (a, b) such that a belongs to a region where ρ is constant, then the middle term of (18) drops out and the other terms are ≤ 0. Moreover, if $a \notin N'$ then the first term is ≤ 0; and if $a \in N$ then $\dot L(a, b) = \nabla_x V \cdot F + \delta\nabla_y W \cdot G$, which is negative unless $H(a, b) = 0$. Therefore it suffices to prove $\dot L(a, b) < 0$ for $a \in N' \setminus N$. Since V is a strict Liapunov function, $\nabla_x V \cdot F(a) \leq -K < 0$ for some constant K and all $a \in N' \setminus N$. Now the third term on the right hand side of (18) is always ≤ 0, so we have $\dot L \leq -K + \delta MB$ where M is an upper bound for $|\nabla_x \rho|$ and B is an upper bound for $|W|$. By taking δ small enough we ensure $\dot L(a, b) < 0$. Constructing a function like L for each equilibrium p of F and adding them up we obtain a strict Liapunov function for (18). QED

One can iterate Theorem 6 for certain *additive cascades* of networks. In an additive cascade, functions of the outputs of the component nets are added to the input units of later nets in the cascade.

Consider for example a cascade whose component nets S_k each satisfy the hypotheses of a special form of the Cohen–Grossberg theorem (see eqn (6)). Fix k and let y be the vector of activations of S_k. The activation dynamics of S_k are assumed to be

$$\dot y_i = a_i(y_i)\left[b_i(y_i) - \sum_j c_{ij} d_j(y_j)\right] + h_i(z) \qquad (19)$$

where z is a vector whose components are the activations of the units in the nets $S_1, \ldots, S_{k-1}$. Notice that a_i is a function of y_i only. We assume $a_i > 0$, $d_j' > 0$ and $c_{ij} = c_{ji}$. Denote $h_i(z)$ by ξ_i. We recast (19) as

$$\dot y_i = a_i(y_i)\left[B_i(y_i) - \sum_j c_{ij} d_j(y_j)\right] = G_k(\xi, y) \qquad (20)$$

where $B_i(y_i) = b_i(y_i) + (\xi_i / a_i(y_i))$. This is in the form required by the Cohen–Grossberg theorem, for each fixed ξ. Therefore the Cohen–Grossberg Liapunov function (7) gives a function $U(\xi, y)$ which for each ξ is a strict Liapunov function for $G_k(\xi, y)$. To apply Theorem 6 the vector fields (20) and the functions $U(\xi, y)$ must be C^1. To achieve this it suffices to assume that the functions a_i, b_i, d_i and h_i are C^1.

This gives a generalization of the Cohen–Grossberg Theorem: There is a Liapunov function for a cascade of nets, each of which separately satisfies the hypothesis of the Cohen–Grossberg theorem in a slightly stronger form. More precisely, we weaken the requirement of symmetry of the weight matrix, assuming instead that it is in triangular block form with symmetric diagonal blocks, provided we restrict the amplification factors to be functions of one variable.

Theorem 7. Consider a network

$$\dot x_i = a_i(x_i)\left[b_i(x_i) - \sum_j c_{ij} d_j(x_j)\right] \qquad (21)$$

with C^1 functions a_i, b_i, d_i. Assume $a_i > 0$ and $d_i' > 0$. Assume hyperbolic equilibria. Assume the matrix $[c_{ij}]$ is in lower (or upper) block triangular form, and that the diagonal blocks are symmetric. Then the activation dynamics has a strict C^1 Liapunov function.

Proof. The block triangular form allows us to represent the net as an additive cascade, of which each component satisfies the requirements of the Cohen–Grossberg Theorem for a strict Liapunov function. The preceding discussion shows that Theorem 5 can be applied to the successive stages of this cascade. QED

It is more difficult to obtain convergence for cascades of systems that are merely assumed to be convergent, but without benefit of Liapunov functions or global asymptotic stability. One way of doing this is to place strong restrictions on the rates of convergence. Roughly speaking, the cascade will be convergent provided the stable equilibria in the earlier stages in the cascade have faster convergence rates in their basins than equilibria in the later stages.

Let us assume about the cascade (20) that *almost every initial value for $\dot x = F(x)$ belongs to the basin of a stable equilibrium p.* (This holds, for example, if F has simple equilibria and there is a strict Liapunov function; it also holds for certain cooperative or competitive systems described below.) Assume also *that for each stable equilibrium p of F, every trajectory of $\dot y = G(p, y)$ converges to a hyperbolic equilibrium q of* (20). The key assumption is: *For any such equilibria p and q, trajectories of $F(x)$ approach p at a faster exponential rate than trajectories of $G(p, y)$ approach q.* The technical formulation of this rate condition is the following. For any eigenvalues λ, μ of the linearizations of $F(x)$ at $x = p$ and of $G(p, y)$ at $y = q$ respectively, the real part of λ, denoted by $\Re(\lambda)$, is less than $\Re(\mu)$. Note that these real parts are negative by the assumption of stability of p and q.

Theorem 8. With the assumptions of the preceding paragraph, almost every initial state of the cascade (20) belongs to the basin of a stable equilibrium.

The proof is outlined in Appendix B.

There are examples of cascades for which Theorem 8 holds, but which have nonconstant periodic orbits: these of course cannot be stable. A variant of Theorem 8, assumes instead: (a) every equilibrium is hyperbolic; (b) every trajectory of $\dot{x} = F(x)$ converges to an equilibrium p; (c) for each such p, every trajectory of $\dot{y} = G(p, y)$ converges; and (d) for any eigenvalues λ, μ of the linearizations of $F(x)$ at $x = p$ and of $G(p, y)$ at $y = q$ respectively, if $\Re(\lambda) < 0$ then $\Re(\lambda) < \Re(\mu)$. The conclusion is then that every trajectory of the cascade converges. It then follows from hyperbolicity that the conclusion of Theorem 8 also holds. The proof of this result is outlined in Appendix B.

To illustrate a possible application of Theorem 8, consider a two layer net $\mathcal{N}$, with each layer an additive recurrent net, such that the second layer $\mathcal{N}_2$ does not send signals to the first layer $\mathcal{N}_1$. Thus $\mathcal{N}$ is a cascade of $\mathcal{N}_1$ into $\mathcal{N}_2$. The dynamics is represented by

$$\dot{x}_i = -c_i x_i + \sum W_{ij}\sigma_j(x_j) + I_i = F_i(x) \tag{22}$$

$$\dot{y}_k = -b_k y_k + \sum U_{kl}\tau_l(y_l) + \sum_j V_{kj}\sigma_j(x_j) + J_k = G_k(x, y) \tag{23}$$

where the weights in the first layer are W_{ij}, those in the second layer are U_{kl}, and the weights from the first to the second layer are V_{kj}. The activation functions in the first layer are $\sigma_j(x_j)$, and in the second layer $\tau_l(y_l)$. Suppose it is known that almost every initial value of the x-dynamics (22) is in the basin of a stable equilibrium, and that the real part of the eigenvalues of DF at such an equilibrium are $\leq -\mu < 0$. (These eigenvalues may be estimated with Gerschgorin's theorem; see section 8.) Assume further that for every such equilibrium p of the x-dynamics (22), the y-dynamics (23) with $x = p$ is convergent, with all equilibria simple. Suppose also that for each k (indexing the units in the second layer), the following inequalities holds:

$$0 < \mu < b_k - \beta_k\left(U_{kk} + \tfrac{1}{2}\sum_{l \neq k} |U_{kl}| + |U_{lk}|\right)$$

Then it can be shown using Gerschgorin's theorem that at any equilibrium q of the y-dynamics (23), with x held constant at a stable equilibrium p of the x-dynamics, the eigenvalues of $D_y G(p, q)$ have real parts $> -\mu$ (compare the discussion preceding Theorem 2).

Therefore by Theorem 8 these assumptions imply that almost every initial value of the activation dynamics of the net $\mathcal{N}$ lies in the basin of a stable equilibrium. Notice that this result is independent of the weights and connections *between* layers.

Further convergence results for cascades will be discussed below in connection with even loop systems.

11. EXCITATORY, INHIBITORY AND SIGN-SYMMETRIC NETS

Consider a net with fixed inputs; we suppress notation for biases and inputs. Assume nonnegative activations x_i, and let the activation dynamics be represented by our standard differential equation

$$\dot{x}_i = G_i(x_i, W_{i1}\sigma_1(x_1), \ldots, W_{in}\sigma_n(x_n)) = F_i(x), \quad i = 1, \ldots, n \tag{24}$$

Suppose the net is *inhibitory*, meaning that all connections between distinct units are inhibitory. We interpret this as $W_{ij} \leq 0$ for $i \neq j$. It then follows from our standing assumptions about equation (2) that $\partial F_i/\partial x_j \leq 0$ for $i \neq j$. Any vector field F with this property is called *competitive*.[2]

When all connections between distinct units are excitatory we call the net *excitatory*. In this case $\partial F_i/\partial x_j \geq 0$ for $i \neq j$. Any vector field F with this property is called *cooperative*.[3]

Inhibitory nets are often used for competitive learning (Grossberg, 1976, 1972; Kohonen, 1984; Malsburg, 1973). Usually the dynamics are designed so that the system is convergent, and for almost all initial conditions, the limiting equilibrium has only one unit with nonzero activation—this kind of activation dynamics is called "choice" or "winner-take-all" competition.[4] This arrangement seems wasteful, since such a net can have only as many responses as it has units. Is there a useful type of competitive learning where the ratio of the number of stable equilibria to the number of units scales at a greater than linear rate? Using the theorem of Smale referred to below one can construct competitive systems in R^n that are convergent and have any number of stable equilibria; but most of these systems do not resemble nets.

[2] This rather abstract formulation of competition is mathematically elegant but hard to verify from real biological or economic data, or to use for predictive purposes; it is more useful to mathematicians than to biologists or economists. Many other mathematical models of competition have been devised, some of which have even been experimentally validated; see e.g., Hsu, Hubbell, and Waltman (1978a, 1978b). Their use in neural networks is unexplored.

[3] A competitive system becomes cooperative under time-reversal. This is a useful trick in investigating attractors and other compact invariant sets, since cooperative systems enjoy special properties derived from the Kamke-Müller comparison principle (see Coppel, 1965): if $x(t)$ and $y(t)$ are solutions to a cooperative system and $x_i(0) \leq y_i(0)$ for all i then $x_i(t) \leq y_i(t)$ for all $t \geq 0$.

[4] It is sometimes assumed that this holds for all initial values. But this can only hold when there is only one stable equilibrium, an uninteresting property for competitive learning nets. When there are $c \geq 2$ stable equilibria in a winner-take-all net (as defined here) with convergent dynamics, then there must exist at least $c - 1$ unstable equilibria at each of which two or more units are activated.

Smale showed in 1976 that any $(n-1)$-dimensional system can be embedded as an attractor in a system of n competing species. This unexpected result shows that only special kinds of competitive systems can be convergent, for example the symmetric Lotka-Volterra-Gause systems studied by MacArthur. Convergence theorems for other special classes of competitive systems have been proved by Chenciner (1977), Coste, Peyraud, Coullet, and Chenciner (1978), Grossberg (1978), Cohen and Grossberg (1983), Hirsch (1988). Competitive and more general systems of differential equations were intensively studied as ecological models by Lotka (1924), Volterra (1931), and Gause (1934).

These results imply that *only special kinds of inhibitory networks can be expected to have convergent activation dynamics.* In the following section we will describe some networks of this type.

Excitatory nets, which can be used for "cooperative learning," are an interesting class of nets which deserve more attention. They have very good convergence properties:

Theorem 9 An irreducible excitatory net represented by eqn (24) has almost quasiconvergent activation dynamics. If all equilibria are simple, then almost every activation state tends toward a stable equilibrium.

This is an immediate consequence of the fact that *every cooperative irreducible system is almost quasiconvergent* [9] (Hirsch, 1984, 1985, 1988a).

In view of this result and others (Hirsch, 1982), cooperative systems cannot have very exotic dynamics. While there are examples of cooperative systems that are not convergent because they contain nonconstant periodic orbits, and even chaotic orbits, these orbits cannot be stable. We would observe them only in special circumstances.

We shall see that some nets having both excitatory and inhibitory connections, and whose weights are sign-symmetric, can be represented by cooperative systems after changing the signs of certain variables. Such systems are therefore almost quasiconvergent.

In Hirsch (1987) the convergence results for cooperative irreducible systems were applied to obtain convergence theorems for certain kinds of neural networks having irreducible activation dynamics, but not necessarily excitatory. The hypothesis of irreducibility is a serious restriction on the network architecture, however. It turns out that similar results apply to many reducible networks, as we now explain.

Consider a net $\mathcal{N}$ represented by our standard differential equation. Besides assuming $a_i > 0$, we assume $\sigma_j' > 0$ for all j (rather than our standing hypothesis $\sigma_j' \ge 0$). Since we allow these derivatives to be arbitrarily close to zero, this assumption does not seem unduly restrictive.

We also impose the very restrictive requirement that the weights are *sign-symmetric*: $W_{ij}W_{ji} \ge 0$ for all i, j. This includes both excitatory and inhibitory nets, and many others as well.

To the net $\mathcal{N}$ (or the system (24)) we associate a *labeled directed interaction graph* Γ: the nodes are the indices $1, \ldots, n$ of the units, with an arrow pointing from j to i only if $W_{ij} \ne 0$. This arrow is labeled with the sign of W_{ij}. Thus Γ is simply a picture of the network with signs of weights attached to the transmission lines. (We adopt the usual convention that if $W_{ij} = 0$ then there is no transmission line from unit j to unit i.)

To a sign-symmetric net we associate another labeled graph, which is not directed, by joining node j to node i only if either W_{ij} or W_{ji} is $\ne 0$. We call this the *reduced graph* Γ^* of the network.

In section 9 we observed that every net breaks up into a cascade of maximal irreducible subnets $\mathcal{N}_k$, called the basic subnets, connected to each other in a feed-forward fashion (see Appendix A). It is tempting to conjecture, but false, that if each basic subnet has convergent activation dynamics, then so has the whole net. The following result, however, can be proved using the results of Hirsch (1985):

Theorem 10 The net $\mathcal{N}$ has almost quasiconvergent dynamics provided each basic subnet $\mathcal{N}_k$ is represented in some coordinate system by a cooperative system of differential equations having isolated equilibria.

It is therefore of interest to determine conditions guaranteeing such a representation.[10]

Now the *standard* differential equations for a subnet $\mathcal{N}_k$ are obtained from (24) by deleting the variables corresponding to units outside the subnet, and setting the corresponding weights to zero. This system is irreducible; and it will be cooperative precisely when all connections between distinct vertices in the subnet are excitatory. Even if the standard differential equations do not give a cooperative system, however, it is sometimes possible to find a change of variables rendering the system cooperative in the new variables.

A simple change of variables effecting this can be made for a basic subnet $\mathcal{N}_k$ in case its interaction

[9] It is unfortunately not known whether this result, and its consequences such as Theorem 9, are valid for systems operating in discrete time.

[10] I don't know whether the assumption of isolated equilibria is really needed. Having isolated equilibria is a generic property of vector fields. It is very likely a generic property of systems like (24).

graph Γ_i satisfies the following *even directed loop condition*:

- Every directed loop contains an even number of minus signs.[17]

Any sign-symmetric net for which this condition holds is called *consistent*. For an irreducible net the even directed loop condition is equivalent to the even-loop condition in Hirsch (1987). In that paper I showed that the activation dynamics of an even-loop net is cooperative if coordinates are changed by reversing the sign of certain activation variables. (See Smith, 1988 for a systematic approach to this process.)

Assume now that $\mathcal{N}$ is consistent. It is easy to see that each $\mathcal{N}_i$ is consistent.[18] It follows that each $\mathcal{N}_i$, being irreducible, is almost quasiconvergent; and therefore Theorem 10 implies that $\mathcal{N}$ is almost quasiconvergent provided it has finitely many equilibria.

In summary:

Theorem 11 Let $\mathcal{N}$ be a net whose activation dynamics are represented by equation (24), with $\sigma_i \geq 0$ and $\sigma_i' > 0$.

(a) If $\mathcal{N}$ is excitatory and irreducible, then system (24) is almost quasiconvergent.
(b) If the activation dynamics in each basic subnet can be represented by a cooperative system of differential equations (allowing an arbitrary change of variables in each basic subnet), then system (24) is almost quasiconvergent provided the equilibrium set is finite.
(c) Suppose $\mathcal{N}$ is sign-symmetric and consistent. Then (24) is almost quasiconvergent provided the equilibrium set is finite.

As an example, consider an irreducible inhibitory net whose reduced graph is embedded in a cubical or hexagonal lattice with nodes at lattice points. Thus every directed edge of the interaction graph is negative and every loop has an even number of edges. Therefore the system is irreducible and consistent, so it transforms into a cooperative system by changing the sign of some variables. Hence (a) implies it is almost quasiconvergent. Even if the net is reducible, provided there are only finitely many equilibria, (c) implies almost quasiconvergence.

Another example of a consistent net is a sign-symmetric net whose reduced graph is embedded in the plane with each node having integer coordinates, with each edge vertical, horizontal, or diagonal of slope ± 1, and with positive weighs on the diagonal edges and negative weights on the vertical and horizontal edges. This represents inhibition between nearest neighbors and excitation between immediate diagonal neighbors. It is easy to see that by changing the signs of the activation variables of each unit whose horizontal and vertical coordinates differ by an even number, we obtain a cooperative dynamical system. If the net is irreducible, or if the equilibrium set is finite, Theorem 11 shows that the activation dynamics are almost quasiconvergent.

An inhibitory net whose reduced graph is embedded in a triangular lattice is not necessarily consistent. There is a completely connected three-dimensional competitive system (representing an inhibitory net with three units) which has an attracting limit cycle; thus it is not almost quasiconvergent. Of course such a net is not consistent.

A convenient property of the class of consistent nets is that it is closed under arbitrary cascading, with *arbitrary* signs for the weights on the new transmission lines, because no new loops are introduced by cascading. In this way quite complex neural nets can be built up, which are guaranteed to have almost quasiconvergent activation dynamics provided the number of equilibria is known to be finite. These are not biologically plausible as models of the nervous system, but they may be useful as designs for artificial networks where convergence is desired.

In Hirsch (1984) it is shown that *sufficiently small perturbations of irreducible cooperative systems are almost quasiconvergent*. Thus almost quasiconvergence is a robust property of these systems. The size of the allowable perturbations can in principle be estimated.

It is not known if the same applies to perturbations of cascades of irreducible cooperative systems. But it can be shown to hold for such cascades (and hence for consistent systems) provided that the perturbed system introduces no new connections between different irreducible components; or more generally, if they are introduced, then they join only components that were originally connected, with the same direction as the original connections.

The following result applies only to nets with a very special architecture, but it yields convergence for all initial states.

Theorem 12. Let $\mathcal{N}$ be a sign-symmetric irreducible net represented by eqn (24). Assume the reduced graph embeds in a straight line, and that the vector field F has continuous partial derivatives of order $n - 1$. Then the activation dynamics are convergent.[19]

Proof By changing the signs of certain activation variables the dynamics can be made cooperative (or competitive). The conclusion now follows from the theorem of Smillie (1984) on tridiagonal systems. QED

[17] Of course zero is an even number.

[18] But this does not imply $\mathcal{N}$ can be made cooperative by a change of variables: $\mathcal{N}$ may be reducible.

[19] The same result probably holds even if F is only continuously differentiable.

It is unfortunately not known whether analogues of the convergence results in this section are valid for discrete-time systems.

It is known that in dimension $n = 2$ competitive systems are always convergent (Albrecht, Gatzke, Haddad, & Wax, 1974), and their orbit structure is completely analyzed (Hale, 1987), but in dimension 3 there can be periodic orbits (Coste, Peyraud, & Coullet, 1979; Gilpin, 1975; Zeeman, 1989) and nonperiodic oscillations (May & Leonard, 1975; Schuster, Sigmund, & Wolff, 1979), but there cannot be so-called "strange attractors" or any kind of chaotic dynamics (Hirsch, 1982, in press-b). In higher dimensions there can be numerically chaotic dynamics (Arneodo et al., 1982); see also the papers by Kerner (1967), Levin (1970), Coste et al. (1978). See also the valuable survey by Freedman (1980), which treats many related types of systems arising in biological modeling.

REFERENCES

Albrecht, F., Gatzke, H., Haddad, A., & Wax, N. (1974). The dynamics of two interacting populations. *Journal of Mathematical Analysis and Applications*, **46**, 658–670.

Almeida, L. B. (1987). A learning rule for asynchronous perceptrons with feedback in a combinatorial environment. In M. Caudill & C. Butler (Eds.), *Proceedings of IEEE International Conference on Neural Networks* (Vol. 2). San Diego, CA: SOS Printing.

Amari, S.-A. (1982). Competitive and cooperative aspects in dynamics of neural excitation and self-organization. In S. Amari & M. Arbib (Eds.), *Competition and cooperation in neural nets* (Springer Lecture Notes in Biomath, 45). New York: Springer.

Amari, S. (1972). Characteristics of random nets of analog neuron-like elements. *IEEE Transactions on Systems, Man and Cybernetics*, **SMC-2**, 643–653.

Aplevich, J. D. (1968). Models of certain nonlinear systems. In E. R. Caianiello (Ed.), *Neural Networks* (pp. 110–115). Berlin: Springer-Verlag.

Arneodo, A., Coullet, P., Peyraud, J., & Tresser, C. (1982). Strange attractors in Volterra equations for species in competition. *Journal of Mathematical Biology*, **14**, 153–157.

Baird, W. (1986). Nonlinear dynamics of pattern formation and pattern recognition in rabbit olfactory bulb. *Physica*, **22D**, 150–175.

Baird, W. (1989). A bifurcation theory approach to vector field programming for periodic attractors. *Proceedings IEEE-INNS International Joint Conference Neural Networks, Washington, DC*. San Diego, CA: SOS Printing.

Bachrach, J. (1988). *Learning to represent state*. Unpublished master's thesis, University of Massachusetts, Amherst.

Carpenter, G. A., & Grossberg, S. (1987). A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphics, and Image Processing*, **37**, 54–115.

Chenciner, A. (1977). Comportement asymptotique des systèmes différentiels du type "compétition d'espèces." *Comptes Rendus de l'Académie des Sciences, Paris*, **284**, 313–315.

Cohen, M., & Grossberg, S. (1983). Absolute stability of global pattern formation and parallel memory storage by competitive neural networks. *IEEE Transactions on Systems, Man and Cybernetics*, **SMC-13**, 815–826.

Coppel, W. (1965). *Stability and asymptotic behavior of differential equations*. Boston: D.C. Heath.

Conley, C. (1978). *Isolated invariant sets and the Morse index* (Regional conference series in mathematics, No. 38). Providence, RI: American Mathematical Society.

Coste, J., Peyraud, J., Coullet, P., & Chenciner, A. (1978). About the theory of competing species. *Theoretical Population Biology*, **14**, 165–184.

Coste, J., Peyraud, J., & Coullet, P. (1979). Asymptotic behaviors in the dynamics of competing species. *SIAM Journal of Applied Mathematics*, **36**, 516–543.

Cowan, J. D. (1967). *A mathematical theory of central nervous activity*. Unpublished dissertation, Imperial College, University of London.

Ellias, S. A., & Grossberg, S. (1975). Pattern formation, contrast control, and oscillations in the short term memory of shunting on-center off-surround networks. *Biological Cybernetics*, **20**, 69–98.

Elman, J. L. (1988). *Finding structure in time* (Tech. Rep. No. 8801). University of California at San Diego, Center for Research in Language.

Freeman, W. (1987). Simulation of chaotic EEG patterns with a dynamic model of the olfactory system. *Biological Cybernetics*, **56**, 139–150.

Freeman, W. J., & Viana Di Prisco, G. (1986a). EEG spatial pattern differences with discriminated odors manifest chaotic and limit cycle attractors in olfactory bulb of rabbits. In G. Palm (Ed.), *Brain theory*. New York: Springer-Verlag.

Freeman, W. J., & Viana Di Prisco, G. (1986b). Correlation of olfactory EEG with behavior: Time series analysis. *Behavioral Neuroscience*, **100**, 753–763.

Freedman, H. I. (1980). *Deterministic mathematical models in population ecology*. New York: Marcel Dekker.

Gause, G. F. (1934). *The struggle for existence*. New York: Hafner.

Gilpin, M. E. (1975). Limit cycles in competition communities. *American Naturalist*, **109**, 51–60.

Golden, R. M. (1986). The "brain-state-in-a-box" neural model is a gradient descent algorithm. *Journal of Mathematical Psychology*, **30**, 73–80.

Golden, R. M. (1988). A unified framework for connectionist systems. *Biological Cybernetics*, **59**, 109–120.

Grossberg, S. (1969). On learning and energy-entropy dependence in recurrent and nonrecurrent signed networks. *Journal of Statistical Physics*, **1**, 319–350.

Grossberg, S. (1972). Neural expectation: Cerebellar and retinal analogs of cells fired by learnable or unlearned pattern classes. *Kybernetik*, **10**, 49–57.

Grossberg, S. (1976). Adaptive pattern classification and universal recoding. I. Parallel development and coding of neural feature detectors. *Biological Cybernetics*, **23**, 121–134.

Grossberg, S. (1978). Competition, decision, and consensus. *Journal of Mathematical Analysis and Applications*, **66**, 470–493.

Grossberg, S. (1982). *Studies of mind and brain*. Boston: Reidel.

Hirsch, M. W. (1980). Unpublished manuscript.

Hirsch, M. W. (1982). Systems of differential equations that are competitive or cooperative. I: Limit sets. *SIAM Journal Mathematical Analysis*, **13**, 167–179.

Hirsch, M. W. (1984). The dynamical systems approach to differential equations. *Bulletin of the American Mathematical Society*, **11**, 1–64.

Hirsch, M. W. (1985). Systems of differential equations that are competitive or cooperative. II: Convergence almost everywhere. *SIAM Journal of Mathematical Analysis*, **16**, 423–439.

Hirsch, M. W. (1987). Convergence in Neural Nets. In M. Caudill & C. Butler (Eds.), *Proceedings of the IEEE International Conference on Neural Networks* (Vol. 2). San Diego, CA: SOS Printing.

Hirsch, M. W. (in press). Stability and convergence in strongly monotone dynamical systems. *Journal für Reine und Angewandte Mathematik*, **383**, 1–53.

Hirsch, M. W. (1988b). Systems of differential equations that are competitive or cooperative. III: Competing species. *Nonlinearity*, 1, 51–71.

Hirsch, M. W. (in press-a). Systems of differential equations that are competitive or cooperative. IV: Structural stability in 3-dimensional systems.

Hirsch, M. W. (in press-b). Systems of differential equations that are competitive or cooperative. V: Convergence in 3-dimensional systems. *Journal of Differential Equations*.

Hirsch, M. W., & Smale, S. (1974). *Differential equations, dynamical systems, and linear algebra*. New York: Academic Press.

Hirsch, M. W., & Pugh, C. (1970). Stable manifolds and hyperbolic sets. *Proceedings of Symposia in Pure Mathematics*, **14**, 133–164.

Holz, M. (1987). Dissertation, University of California at Berkeley.

Hopfield, J. J. (1984). Neurons with graded response have collective computational properties like those of two-state neurons. *Proceedings of the National Academy of Sciences, USA*, **81**, 3088–3092.

Hsu, S. B., Hubbell, S. P., & Waltman, P. (1978a). A contribution to the mathematical theory of competing predators. *Ecological Monographs*, **48**, 337–349.

Hsu, S. B., Hubbell, S. P., & Waltman, P. (1978b). Competing predators. *SIAM Journal of Applied Mathematics*, **35**, 617–625.

Jordan, M. I. (1987). Attractor dynamics and parallelism in a connectionist sequential machine. *Proceedings of the Eighth Annual Conference of the Cognitive Science Society*, 531–546.

Kelly, D. G. (in press). Stability in contractive nonlinear neural networks. *IEEE Transactions on Biomedical Engineering*.

Kerner, E. H. (1961). On the Volterra-Lotka principle. *Bulletin of Mathematical Biophysics*, **23**, 141–157.

Kohonen, T. (1984). *Self-organization and associative memory*. New York: Springer Verlag.

LaSalle, J. P. (1968). Stability theory for ordinary differential equations. *Journal of Differential Equations*, **4**, 57–65.

Levin, S. A. (1970). Community equilibria and stability, and an extension of the competitive exclusion principle. *American Naturalist*, **104**, 413–423.

Lotka, A. (1956). *Elements of mathematical biology*. New York: Dover (Reprinted from *Elements of physical biology*, Williams and Wilkins Co.)

Li, Z., & Hopfield, J. J. (in press). Modeling the olfactory bulb. *Biological Cybernetics*.

MacArthur, R. H. (1969). Species packing, or what competition minimizes. *Proceedings of the National Academy of Sciences USA*, **54**, 1369–1375.

Malsburg, C. van der (1973). Self organization of orientation sensitive cells in the striate cortex. *Kybernetik*, **14**, 85–100.

May, R. M., & Leonard, W. J. (1975). Nonlinear aspects of competition between three species. *SIAM Journal of Applied Mathematics*, **29**, 243–253.

Mozer, M. C. (1988). *A focused back-propagation algorithm for temporal pattern recognition* (Tech. Rep.). University of Toronto, Departments of Psychology and Computer Science.

Noble, B., & Daniel, J. (1988). *Applied linear algebra*. Englewood Cliffs, NJ: Prentice Hall.

Pearlmutter, B. A. (1988). *Learning state space trajectories in recurrent neural networks: A preliminary report* (Tech. Rep. No. AIP-54). Carnegie Mellon University, Department of Computer Science.

Pineda, F. J. (1987). Generalization of back propagation to recurrent and higher order neural networks. *Proceedings of IEEE Conference on Neural Information Processing Systems*. San Diego, CA: SOS Printing.

Robinson, A. J., & Fallside, F. (1987). The utility driven dynamic error propagation network (Tech. Rep. No. CUED/F-INFENG/TR 1). Cambridge University, Engineering Department.

Rowher, R., & Forrest, B. (1987). Training time-dependence in neural networks. In M. Caudill & C. Butler (Eds.), *Proceedings of the IEEE First International Conference on Neural Networks*. San Diego, CA: SOS Printing.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1). Cambridge: MIT Press.

Schuster, P., Sigmund, K., & Wolff, R. (1979). On ω-limits for competition between three species. *SIAM Journal of Applied Mathematics*, **37**, 49–54.

Sejnowski, T. J. (1977). Storing covariance with nonlinearly interacting neurons. *Journal of Mathematical Biology*, 4, 303–321.

Servan-Schreiber, D., & Cleeremans, A. (1988). Learning and internal representations in recurrent networks (Tech. Rep. No. CMU-CS-88-183). Carnegie Mellon University, Computer Science Department.

Skarda, C. A., & Freeman, W. (1987). How brains make chaos in order to make sense of the world. *Behavioral and Brain Sciences*, **10**, 161–195.

Smale, S. (1976). On the differential equations of species in competition. *Journal of Mathematical Biology*, 3, 5–7.

Smillie, J. (1984). Competitive and cooperative tridiagonal systems of differential equations. *SIAM Journal of Mathematical Analysis*, **15**, 530–534.

Smith, H. (1988). Systems of differential equations which generate a monotone flow: A survey of results. *SIAM Review*, **30**, 87–113.

Stornetta, W. S., Hogg, T., & Huberman, B. A. (1987). A dynamical approach to temporal pattern processing. *Proceedings of the IEEE Conference on Neural Information Processing Systems*. San Diego, CA: SOS Printing.

Volterra, V. (1931). *Leçons sur la Théorie Mathématique de la Lutte pour la Vie*. Paris: Gauthier-Villars et Cie.

Williams, R. J., & Zipser, D. (1988). A learning algorithm for continually running fully recurrent neural networks (Tech. Rep. No. ICS 8805). University of California at San Diego, Institute of Cognitive Science.

Zeeman, M. L. (1989). Hopf bifurcations in three dimensional competitive Volterra-Lotka systems. Unpublished dissertation, University of California at Berkeley.

APPENDIX A: CASCADE DECOMPOSITION

We prove Theorem 4, which states that we can decompose any reducible net $\mathcal{N}$ into a cascade whose components are the basic subnets of $\mathcal{N}$.

Call two units U_i and U_j *cocyclic* if they are the same or they belong to a loop of directed transmission lines. This is an equivalence relation on the set of units; the equivalence class of a unit is the set of all units cocyclic with it. An equivalence class is a maximal set of units of which each pair is cocyclic. A net is irreducible if there is only one equivalence class. Since we are supposing $\mathcal{N}$ is reducible, the set of units breaks up into at least two equivalence classes $\mathcal{E}_1, \ldots, \mathcal{E}_r$.

Let $\mathcal{N}_i$ be the subnet of $\mathcal{N}$ whose units are those in $\mathcal{E}_i$, with all the connections between them. Each $\mathcal{N}_i$ is a maximal irreducible subnet of $\mathcal{N}$; the $\mathcal{N}_i$ are the basic subnets of $\mathcal{N}$. It is easy to see that if there are connections between two basic subnets $\mathcal{N}_i$ and $\mathcal{N}_j$, they can be in only one direction: all from $\mathcal{N}_i$ to $\mathcal{N}_j$ or all from $\mathcal{N}_j$ to $\mathcal{N}_i$.

There must be some basic subnet with no lines coming into it; call these the *level 0* subnets. Let $\mathcal{N}_0$ be the union of all these. Since $\mathcal{N}$ is reducible then there must be some basic subnet receiving signals from $\mathcal{N}_0$ but not from any other basic subnet except itself. These are *level 1* subnets. Let N_1 be the subnet comprising the units of the level 1 subnets and all connections between them. Continuing in this way we recursively define a basic subnet to be *level $k+1$* if it receives input from units in one or more basic subnets at a lower (previously defined) level, but not from any other units of [illegible]; then N_{k+1} is defined to be the subnet comprising the units of the level $k+1$ subnets and all connections between them.

The net comprising [illegible] and all the connections between them (which go only from [illegible] to [illegible]) is a subnet [illegible] which is a cascade of [illegible] into [illegible]. If [illegible], then we define [illegible] to be the net comprising [illegible] and all connections from [illegible] to [illegible]. Then [illegible] is the cascade of [illegible] into [illegible]. In this way we build up the original net [illegible] as an iterated cascade whose components are the maximal irreducible subnets [illegible].

APPENDIX B: PROOF OF THEOREM 8

We prove Theorem 8.

Let F be a vector field on R^m, G a map from $R^m \times R^n$ to R^n, and consider the dynamical system

$$\dot{x} = F(x), \quad \dot{y} = G(x, y). \qquad (25)$$

This is the cascade of the two systems $\dot{x} = F(x)$ and $\dot{y} = G(p, y)$, where p is a parameter for the latter system. We make the following assumptions:

(a) Almost every initial value for $\dot{x} = F(x)$ belongs to the basin of a stable equilibrium p.

(b) For each stable equilibrium p of F, every trajectory of $G(p, \cdot)$ converges to a hyperbolic equilibrium q of system (25).

(c) Let p be a stable equilibrium of $F(x)$ and q a stable equilibrium of $G(p, \cdot)$ (with p held fixed). Suppose λ is an eigenvalue of the linearization of F at p, and μ is an eigenvalue of the linearization of $G(p, \cdot)$ at q. Then the real part of λ is less than the real part of μ.

Theorem 8. With the preceding assumptions, almost every initial state of the cascade (25) belongs to the basin of a stable equilibrium.

Proof. Let $S \subset R^m \times R^n$ be a set of positive measure; we must show that some point of S lies in the basin of a stable equilibrium for (25). By (a) we assume that there is a point $(x, y) \in S$ such that $x \in R^m$ is in the basin $B(p)$ of a stable equilibrium p for F. Fix such a p, and let $T \subset S$ be subset of all points (x, y) in S such that $x \in B(p)$. It can be shown using (c) that every trajectory starting in T is asymptotic with some trajectory in $p \times R^n$ (the proof uses the methods of Hirsch and Pugh, 1970). Therefore by (b) every trajectory starting in T is in the stable manifold $W(q)$ of an equilibrium q; thus T is contained in the union of stable manifolds. Since the equilibrium set is countable by hyperbolicity, at least one of the stable manifolds $W(q)$ meeting T must have positive measure. But this means $W(q)$ has dimension $m + n$, so q must be a stable equilibrium. QED

The variant of Theorem 8 assumes:

(a′) every equilibrium is hyperbolic;

(b′) every trajectory of $\dot{x} = F(x)$ converges to an equilibrium p, and for each such p every trajectory of $\dot{y} = G(p, y)$ converges; and

(c′) for any eigenvalues λ, μ of the linearizations of $F(x)$ at $x = p$ and of $G(p, y)$ at $y = q$ respectively, if [illegible], then [illegible].

The conclusion is then that every trajectory of the cascade converges. To prove this, consider a single trajectory $(x(t), y(t))$ with $x(t) \to p$. Then $x(0)$ lies in the stable manifold of p for the vector field F; denote this stable manifold by $V \subset R^m$. Then $(x(t), y(t))$ lies in the submanifold $M = V \times R^n \subset R^m \times R^n$, and M is invariant under the flow of the full cascade (25). Assumption (c′) can be used to show that every trajectory in M is asymptotic to a trajectory in $p \times R^n$. Therefore the trajectory converges. QED