Title
Convergence in cascades of neural networks

Author
Hirsch, MW

Publication Date
1989-12-01

Peer reviewed

CONVERGENCE IN CASCADES OF NEURAL NETWORKS

Morris W. Hirsch
Department of Mathematics
University of California
Berkeley, California 94720 USA

Convergence of the activation dynamics of a cascade of neural nets is studied. Several mathematical results are presented which guarantee convergence of the cascade provided each component subnet is convergent.

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.

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 $\mathcal{N}_{m-1}$ is fed only into $\mathcal{N}_m$. A generalization of a layered net is a *cascade*. Let $\mathcal{N}_0$ and $\mathcal{N}_1$ be two separate nets. If some units of $\mathcal{N}_0$ feed their outputs to units in $\mathcal{N}_1$ via new connections, we obtain a larger net $\mathcal{M}_1$, called a cascade of $\mathcal{N}_0$ into $\mathcal{N}_1$. If outputs from $\mathcal{M}_1$ are fed into a third net $\mathcal{N}_2$, separate from $\mathcal{M}_1$, we obtain a net $\mathcal{M}_2$, a cascade of $\mathcal{M}_1$ into $\mathcal{N}_2$. By iterating this process we obtain cascades of any number of nets $\mathcal{N}_0, \mathcal{N}_1, \ldots$ 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* $\mathcal{N}_j$. 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 net is reducible if an only if 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. 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. The following is a crude but useful structure theorem for reducible nets:

Theorem 1 Every reducible net $\mathcal{N}$ is a cascade whose components are the basic subnets of $\mathcal{N}$.

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 consider this question for two dynamical properties: A system is *convergent* if every trajectory converges; it is *globally asymptotically stable* if in addition there is only one equilibrium, and it is asymptotically stable. Globally asymptotically stable nets have been considered by D. G. Kelly [3].

For simplicity we assume that equilibria are hyperbolic, and that some bounded set attracts all trajectories. This implies that the equilibrium set is finite.

A convenient mathematical model of a neural net is a dynamical system with input parameters. We consider systems governed by differential equations, but analagous results hold for discrete time systems.

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

$$\dot{x} = F(x), \ \dot{y} = G(x,y) \tag{1}$$

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

More complex cascades of dynamical systems with parameters can be built by iterating this construction as follows. Let $E^0, \ldots, E^r$ denote Euclidean spaces of various dimensions. Let $F^j: E^0 \times \ldots \times E^j \rightarrow E^j$ be a map, thought of as a vector field on E^j with input parameters from $E^0 \times \ldots \times E^{j-1}$, for each $j = 1, \ldots, r$. The cascade of this family $\{F^0, \ldots, F^r\}$ is the following dynamical system on state space $E^0 \times \ldots \times E^j$, where $\mathbf{x}^j$ denotes a vector in E^j:

$$\dot{\mathbf{x}}^0 = F^0(\mathbf{x}^0),$$

$$\dot{\mathbf{x}}^j = F^j(\mathbf{x}^0, \ldots, \mathbf{x}^j), \ \ j = 1, \ldots, r$$

Theorem 2 A cascade of systems, each of which is globally asymptotically stable for every constant parameter value, is globally asymptotically stable.

Proof It suffices to consider a cascade of two systems, as in (1). Let $z(t) = (x(t), y(t))$ be a trajectory of (1). Since F is globally asymptotically stable, x(t) converges to the unique equilibrium p for F. Therefore the omega limit set of z(t) is a closed bounded set in $p \times \mathbf{R}^n$, 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 (1) is convergent.

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

Suppose for example that the dynamics of F are convergent; that for each fixed equilibrium p for F there is a strict Liapunov function for the system $\dot{y} = G(p,y)$. Then the cascade (1) is convergent. To see this let $z(t) = (x(t), y(t))$ be a solution with $x(t) \rightarrow p$. 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. 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 (1) is convergent.

We now present a general way of constructing a strict Liapunov function[1] for the system (1), assuming that F has one, and that for each fixed ξ the vector field $G(\xi, y)$ has one.

Theorem 3 In system (1) assume that F has a C^1 (continuously differentiable) strict Liapunov function V(x), and that there is a C^1 function U(x,y) such that for each fixed ξ, $V(\xi,y)$ is a strict Liapunov function for the vector field $G(\xi,y)$. Then there is a C^1 strict Liapunov function for system (1).

Proof We may assume V and U are bounded, composing them with arctan otherwise.

Let $\mathcal{E}$ denote the finite equilibrium set of F. Let ρ be a C^1 real-valued function on $\mathbb{R}^n$ taking the value 1 on a neighborhood N of $\mathcal{E}$, and the value 0 outside a larger, bounded neighborhood N′ of $\mathcal{E}$. Pick $\delta>0$, to be specified later.

Define the function

$$L(x,y)=V(x)+\delta\rho(x)U(x,y).$$

Clearly L is C^1. We show L is a strict Liapunov function provided δ is small enough.

Let $H(x,y)\equiv(F(x),G(x,y))$ denote the vector field defined by the right hand side of (1). Then the derivative of L along a trajectory of H is

$$\dot{L}=\nabla L\cdot H=\nabla_x V\cdot F \;+(\;\delta\nabla_x\rho\cdot F)U \;+\delta\rho\nabla_y U\cdot G, \qquad (2)$$

where $\nabla_y U$ means the gradient of U with respect to the y coordinates, etc. 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 (2) drops out and the each of the other terms is ≤ 0. Moreover if $a\notin N'$ then the first term is negative; and if $a\in N$ then $\dot{L}(a,b)\;=\nabla_x V\cdot F+\delta\nabla_y U\cdot G$, which is negative unless $H(a,b)=0$. Therefore it suffices to prove $\dot{L}(a,b)<0$ for $a\in N'\backslash N$, the set where $0<\rho(a)<1$. Since V is a strict Liapunov function and a is outside the neighborhood N of $\mathcal{E}$, it follows that $\nabla_x V\cdot F(a)\;\leq\;-K\;<\;0$ for some constant K and all $a\in N'\backslash N$. Now the third term on the right hand side of (2) 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 $|U|$. By taking δ small enough we ensure $\dot{L}<0$. QED

One can iterate Theorem 3 for certain ***additive cascades*** of networks that individually admit strict C^1 Liapunov functions for their activation dynamics. 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 $\mathcal{N}_J$ each satisfy the hypotheses of the Cohen-Grossberg theorem[2]. Fix J and let y be the vector of activations of $\mathcal{N}_J$. The activation dynamics of $\mathcal{N}_J$ are assumed to be

$$\dot{y}_i\;=\;a_i(y_i)[b_i(y_i)\;-\;\sum_k c_{ik}d_k(y_k)]\;+\;h_i(z^J) \qquad (3)$$

where z^J is a vector whose components are the activations of the units in the nets $\mathcal{N}_1,\ldots,\mathcal{N}_{J-1}$. Assume $a_i>0$, $d_k'>0$ and $c_{ik}=c_{ki}$. Denote $h_i(z^J)$ by ξ. We recast (3) as

$$\dot{y}_i\;=\;a_i(y_i)[\;B_i(y_i)-\;\sum_k c_{ik}d_k(y_k)]\equiv G_i(\xi,y) \qquad (4)$$

where $B_i(y_i)=b_i(y_i)+(\xi/a_i(y_i))$. This is in the form required by the Cohen-Grossberg theorem, for each fixed ξ. Therefore the Cohen-Grossberg Liapunov function gives a function $U(\xi,y)$ which for each ξ is a strict Liapunov function for $G_i(\xi,y)$. We need one more hypothesis in order to apply Theorem (3): the vector fields (4) 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_J are C^1.

This gives a generalization of the Cohen-Grossberg Theorem: There is a Liapunov function for an additive cascade of nets, each component of which separately satisfies the hypothesis of the Cohen-Grossberg theorem. More precisely, we can 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 4 Consider a network

$$\dot{x}_i\;=\;a_i(x_i)[b_i(x_i)\;-\;\sum_k c_{ik}d_k(x_k)] \qquad (5)$$

with C^1 functions a_i, b_i, d_i. Assume $a_i>0$ and $d_i'>0$. Assume the constant matrix $[c_{i,j}]$ 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 and hence has a strict Liapunov function. The preceding discussion shows that Theorem 3 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 trajectories in earlier components converge to equilibria at faster exponential rates than equilibria in later stages.

Let us assume about the cascade (1) that every equilibrium is hyperbolic and that every trajectory of $\dot{x}=F(x)$ converges. Assume also that for each equilibrium p of F, every trajectory of G(p,y) converges to an equilibrium q of (1). 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, if the real part of λ is negative, then it is less than the real part of μ.

Theorem 5 With the assumptions of the preceding paragraph, every trajectory of the cascade (1) converges.

Proofs, details and examples will appear in a forthcoming article in *Neural Networks*.

References

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

[2] S. Grossberg and M. Cohen, "Absolute stability of global pattern formation and parallel memory storage by competitive neural networks," *IEEE Trans. Systems, Man and Cybernetics*. vol. SMC-13, pp. 815–826.

[3] D. G. Kelly, "Stability in contractive nonlinear neural networks. IEEE Transactions in Biomed. Eng., in press.