nomics, Discussion Paper, (1992). recognizers," *Machine Learning*, 7 (1991), 227-252.

[6] Marcus C.M. and Westervelt R.M., "Dynamics of Iterated Map Neural Networks," *Physical Review A*, Vol. 40, (1989), 501-504.

[7] Pollack J.B., "The induction of dynamic recognizers," *Machine Learning*, 7 (1991), 227-252.

[8] Smale S., *The Mathematics of Time*, Springer-Verlag, New York, New York, 1980.

[9] Smith A.W., and Zipser D., "Learning Sequential Structure With The Real-Time Recurrent Learning Algorithm," *International Journal of Neural Systems*, Vol. 1, No. 2, (1989) 125-131.

[10] Wang X., "Period Doublings to Chaos in a Simple Neural Network: An Analytical Proof," *Complex Systems*, 5(4), (1991) 425-441.

[11] Williams R.J. and Zipser D., "A learning algorithm for continually running fully recurrent networks," *Neural Computation*, Vol. 1 (1989), 270-280.

6 Acknowledgments

I would like to acknowledge financial support from a National Defense Science and Engineering Fellowship, and computer support from the San Diego Supercomputer Center. I would like to thank Fu-Sheng Tsung for getting me interested in a special case of theorem 1. I would also like to thank Michael Freedman and Oded Schramm of the UCSD mathematics department for helpful discussions of earlier approaches to the problem. Finally, I would like to thank Arnold Mandell for encouraging me to study mathematics.

References

[1] Cleermans A., McClelland J.L, and Servan-Schreiber D., (1989) "Encoding sequential structure in simple recurrent networks," Tech. Rept. CMU-CS-88-183, Carnegie-Mellon University, Pittsburgh, November 1988.

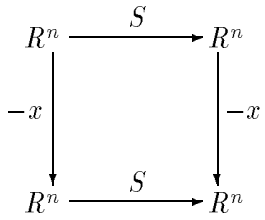
[2] Elman J.L., (1991) "Representation and Structure in Connectionist Models," Tech. Rept. 8903, Center for Research in Language, University of California, San Diego, La Jolla, CA, August 1989.

[3] Hertz J., Krough A., and Palmer R.G., *Introduction to The Theory of Neural Computation*, Addison-Wesley Publishing Company, Menlo Park, California, 1991.

[4] Hopfield J.J., "Neural Networks and Physical Systems With Emergent Collective Computational Abilities," *Proceedings of the National Academy of Sciences, USA*, Vol. 79, (1982) 2554-2558.

[5] Kuan C.M., Hornik K., White H., "Some Convergence Results for Learning in Recurrent Neural Networks," *UCSD, Department of Eco-*

diagram commutes:



Where $-x$ represents the function $x \mapsto -x$. If we now consider $S(Wx)$, where a is a scalar, then the same calculation shows that $-S(Wx) = S(-IWx)$, where I is the identity matrix for R^n. Finally, carefully adding the bias term b yields $-S(Wx + B) = S(-IWx - B)$. Thus we have that S(Wx + B) has the same dynamics as $S(Wx - B)$. Hence, we may identify these points in parameter space.

5 Discussion

We have seen that it is sometimes possible to exploit symmetries in neural network models to reduce the size of their parameter spaces. Taking this process to its limit, it may be possible to carve up the parameter spaces of given models into some skeleton which supports the seemingly finite number of topologically different behaviors (e.g. fixed points and limit cycles) or collections of parameter sensitive behaviors (e.g. the period-doubling route to chaos). This kind of work may be important for at least two reasons. First, it would show that systems of the same nature (possibly actual neural networks) work with a finite number of distinct behaviors and behavioral regimes (Casey and Mandell - in progress). Secondly, it would enumerate the entire range of dynamic possibilities, which would benefit those who use neural network models to solve AI-type problems by allowing them to choose input-output relationships which can be supported by admissible dynamics.

have modulus less than one is equivalent to showing that the zeros of $P(z)$, a complex polynomial of degree n for an n-unit network, all lie in the unit disk. Rouche's theorem says that if $|z^n - P(z)| < 1$ for all z in the unit circle (S^1), then z^n and $P(z)$ have the same number of zeros in the unit disk (counting multiplicities). Thus, if this inequality holds, then the network is a contraction mapping, and so, by the contraction mapping theorem, the network will iterate to a unique fixed point, independent of initial conditions (which is stronger than saying that the dynamics are tame).

For n = 2, we have that $|z(W_{11} + W_{22}) + (W_{11}W_{22} - W_{21}W_{12})| < 1$, which holds for all z in S^1 iff $|W_{11} + W_{22}| + |W_{11}W_{22} - W_{21}W_{12}| < 1$. Hence, we must have that $|Tr(W)| + |Det(W)| < 1$.

4 Further Reducing The Size of the Parameter Space

One symmetry in this model which has the effect of identifying points in the parameter space is just the fact that we may relabel our units. Thus, in an n-unit network, each point in the $n^2 + n$ dimensional parameter space gets identified with n! other points. The reason for saying that these points are identified is that we would like to correspond points in the parameter space with the dynamics that the system will have when given those parameter settings. Thus, identification in the parameter space can be accomplished via a topological conjugacy (such as the conjugacy which interchanges variables). Next, we exploit a symmetry in the activation function to make a further identification.

A second, less obvious, symmetry results from the fact that

$S(x) = (2.0)(1.0 + e^{-x})^{-1} - 1.0$

satisfies the condition that $-S(x) = S(-x)$. Equivalently, the following

(e.g. made very simple) so long as the network satisfies the sign symmetric condition. Since, as computer experiments indicate, the sign symmetric regions of weight space contain complicated behavior, this method of control may be more desirable than perturbing to and from symmetric weight regions if complex behavior is necessary at times, since one is likely to have to make more drastic changes in several weights to get to and from the tame dynamics of symmetric weights to complicated behavior.

In two dimensions, these results are essentially sharp (as far as giving a general condition on W_{12} and W_{21} which will insure tame dynamics), since it is possible to have limit cycles and chaos as shown in Wang (ref.) when the sign of W_{12} does not equal the sign of W_{21}. In higher dimensions, numerical searches have found non-tame dynamics in the sign symmetric regions of parameter space where the algebraic condition on the weights is not satisfied.

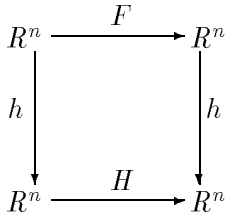
Henceforth, all networks will be assumed to have the activation function given by $S_i(x) = (2.0)(1.0 + e^{-x})^{-1} - 1.0$. Another large class of weights for which networks will have tame dynamics are those weights for which the network is a contraction mapping (i.e. for which there exists a $\lambda < 1$ such that $|f(x) - f(y)| < \lambda|x - y|$ for all x, y in the domain space of the network). This condition has been used by Kuan et. al (1992), to insure convergence of a family of learning algorithms for certain problems, and is equivalent to all of the eigenvalues of the network's derivative having absolute value less than one. Kuan et. al (1992) gave the sufficient condition that the sum of the squares of the entries of W be less than $g = max_{i,x \epsilon R^n} S_i(x)$. Without loss of generality, we may assume that $g = 1$, so long as we remember to rescale W as well. A refinement on this condition may be made by use of Rouche's theorem from complex analysis. Since the eigenvalues of the derivative are the zeros of its characteristic equation ($P(z)$), showing that the eigenvalues

$H(X) = (h_1(x_1, x_2, ..., x_n), ..., h_n(x_1, x_2, ..., x_n))$, then the conjugacy does continue to hold. Since the conjugacy which has been used is a diagonal matrix, it need only be shown that the Lyapunov function which Hopfield found may be generalized so that each unit's activation function may be rescaled. It is easily checked and found to hold, hence the result also holds in the case of asynchronous update.

The class of positively symmetrizable matrices is much larger than the class of symmetric matrices. For example, in two dimensions, this is equivalent to the off diagonal entries having the same sign. In which case, $\lambda_1 = W_{21}$ and $\lambda_2 = W_{12}$. A straight-forward calculation shows that, in dimension n ($n > 2$), we have that the sign(W_{ij}) must equal the sign(W_{ji}), and then that the weights satisfy the further algebraic relation that:

$W_{n1} \prod_{i=2}^{n} W_{(i+1)(i)} = W_{1n} \prod_{i=2}^{n} W_{(i)(i+1)}$

Thus we see that, in dimension two, we have that fully half of the parameter space is occupied by fixed points and period two orbits, and in higher dimensions, these dynamics occupy a subspace which contains a hypersurface of dimension one less than the dimension of the parameter space. There are at least two ways of looking at the higher dimensional condition. The first is to say that since the dimension of the hypersurface is one less than the dimension of the parameter space, it has no volume (i.e. zero Lebesgue measure), and, hence, is not physically relevant. Another point of view notes that this condition is nicer than the condition that the weight matrix must be symmetric, since in this case, one may be fairly careless while choosing the off diagonal entries in the weight matrix (assuring only that sign(W_{ij}) = sign(W_{ji})) for all but one of the parameters, and then one can carefully choose the last weight to assure that the condition is satisfied. If an analogue of this condition exists somewhere in the working space of the nervous system, this says that by adjusting one parameter, behavior may be controlled



Where $H = S'(W'(x)), S'(x) = (S'_1(x), ..., S'_n(x)), S'_i(x) = (2.0)(1.0 + e^{-x})^{-1} - 1.0$, and $W' = WC'(x) + D'$, where C' is a diagonal matrix, with one-halves along the diagonal, and D' is an vector of n one-halves. ($C'(x) + D'$ is the inverse of $C(x) + D$.) Now noting that WC' is positively symmetrizable if and only if W is positively symmetrizable, the corollary follows from Theorem 1.

Of course, tame dynamics do occur in regions of the weight space for which the connectivity matrix is not positively symmetrizable, but these regions have very complicated boundaries, and do not seem likely to yield to such a simple characterization. Thus it is conjectured that the result of theorem 1 is essentially sharp in as far as it is the most powerful, yet very general, condition which can be put on the weights to insure tame dynamics.

The results of Marcus and Westervelt do not apply to networks with asynchronous update, so one might question that this weakening of the symmetric weight condition applies to such networks.

Corollary 2: Hopfield networks with positively symmetrizable connectivity matrices have only fixed point attractors.

Proof: It is not true in general that a topological conjugacy between multidimensional dynamical system holds when the update of these systems is allowed to take place in some subset of the variables at each time step. However, if the topological conjugacy is of the form $H(X) = (h_1(x_1), h_2(x_2), ..., h_n(x_n))$, rather than the more general

Consider the homeomorphism from R^n to R^n given by the linear map whose standard matrix representation is diagonal and has real, strictly positive entries. Call this matrix U, and its diagonal entries λ_i. If A is a matrix such that there exists a matrix U of this form, and AU (matrix multiplication) is a symmetric matrix, then we shall call A positively symmetrizable.

Theorem 1: Let $H(x)$ be an MW-nice network with a positively symmetrizable weight matrix W, then $H(x)$ has tame dynamics.

Proof: Since H(x) = S(Wx + B) be our MW-nice network, and W is positively symmetrizable by a matrix U, then we have the following diagram commutes:

$$\begin{array}{ccc} R^n & \xrightarrow{H} & R^n \\ U \downarrow & & \downarrow U \\ R^n & \xrightarrow{N} & R^n \end{array}$$

Hence $N(x) = U^{-1}(H(Ux))$, and therefore, $N_i = \frac{1.0}{\lambda_i} S_i(WUx + B)$. Hence $N(x)$ is also MW-nice. Since W is positively symmetrizable by the matrix U, WU is symmetric, and, therefore, but the result of Marcus and Westervelt (1989), $N(x)$ has tame dynamics. Since $N(x)$ and $H(x)$ are topologically conjugate, $H(x)$ also has tame dynamics.

Corollary 1: $F(x) = S(Wx + B)$, where $S_i(x) = (1.0 + e^{-x_i})^{-1}$ has tame dynamics if W is positively symmetrizable.

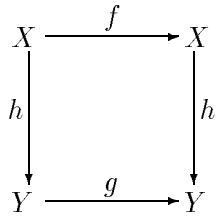
Proof: Now use the topological conjugacy $h(x)$, where $h(x) = C(x) + D$, where C is a diagonal matrix, with twos along the diagonal, and D is a vector of n negative ones, to transform our networks into models which have the same dynamics, but the functions are centered about the origin (thus satisfying the conditions of Marcus and Westervelt (1989)). We have the following diagram:

which is equivalent (in a sense described below), but which more plainly embodies the principle which constrains the dynamics. In the case of additive neural network models with sigmoidal activation functions, this principle seems to be symmetry in the connectivity matrix.

Definition: A network will be said to be MW-nice if it satisfies the following conditions (i) the state variables are continuous and updated in parallel (ii) the activation functions (which may be different for each unit) increase slower than linearly (i.e. that $|S_i(y)| \leq k_i|y|^{\alpha_i}$ for all y in R, and for some constants $0 \leq \alpha_i < 1, k_i < \infty$) and (iii) the activation functions are monotone increasing.

The precise results of Marcus and Westervelt are that MW-nice networks have tame dynamics.

The equivalence mentioned above is the equivalence relation in the category of continuous maps between topological spaces - called topological conjugacy. Two maps are said to be topologically conjugate if there exists an invertible continuous map from some topological space X onto another topological space Y such that the inverse function is also continuous (such a map is called a homeomorphism). One may think of such a map as a change of coordinates. More graphically, we have that the following diagram commutes:



In other words, $g = h(f(h^{-1}))$. (h^{-1} exists and is continuous since h is a homeomorphism.) It follows that $g^n = h(f^n(h^{-1}))$, so, in particular, there is a $1-1$ correspondence between periodic and fixed points of f and those of g.

3 Dynamics

One dimensional networks' range of behaviors can be completely enumerated. If the self-weight is zero, then the network is a constant function, so assume the self-weight is non-zero. Up to small perturbations, the network can have one attracting fixed point, two attracting and one repelling fixed point, or an attracting period two oscillation and an unstable fixed point. To see that the network cannot have more attracting fixed points, observe that the second derivative must pass through zero at some point between two stable fixed points, and the second derivative has only one zero. It follows from the fact that the network is differentiable and injective that the periodic points cannot have period greater than two (by application of the mean value theorem).

In two and more dimensions, we might hope for such a simple characterization of the entire range of behaviors which a network may possess, but this is not the case. As the weights are smoothly varied, the network can be seen to undergo various local and global bifurcations, producing limit sets which include fixed points, periodic points of any given period, limit cycles, at least two different general types of fractal structures which support chaotic dynamics, and combinations of the above (in the sense that which limit set one iterates to depends upon initial conditions). See Wang (1990) for a rigorous proof of the existence of chaotic dynamics in a two dimensional network. In networks of dimension two and higher, a reasonable first attempt to characterize behaviors consists of eliminating large ranges of parameters as prospects for containing "interesting" dynamics. The main result of this paper is meant to be taken in this spirit.

In dynamical systems theory, it is common practice to say something about the dynamics of a system by first understanding them for a system

himself, after noting the existence of the Lyapunov function given symmetric weights, went on to do numerical work to test the necessity of this restriction (1982).

2 The Model

To understand what dynamical systems theory can tell us, we must first put the model into a more formal, mathematical framework. Unless stated otherwise, an n-dimensional or n-unit network will be taken to mean an endomorphism $F(x)$, from real n-dimensional euclidean space (R^n) into itself, given by:

$F(x) = S(Wx + B)$

$x = (x_1, x_2, ..., x_n)$

W is the linear transformation given by the weight matrix, Wx denotes the vector $(\sum_{j=1}^{n} W_{1j}x_j, ..., \sum_{j=1}^{n} W_{nj}x_j)$ and B is a translation (the vector of bias weights).

$S(x) = (S_1(x_1), S_2(x_2), ..., S_n(x_n))$

The activation function, $S_i(x)$ can be any sigmoidal function, but will usually be taken to be the function:

$S_i(x) = (1.0 + e^{-x_i})^{-1}$

It is clear that the range of F is contained in the unit n-cube (I^n), and that F is a C^∞ diffeomorphism from the unit n-cube to all points in the range of F if and only if the linear transformation, W, is nonsingular. [1]

[1] The general problem of understanding the space of diffeomorphisms of a space (or, more generally, a manifold) to itself has been an object of mathematical interest since the time of Poincare. Birkhoff continued the work in the 1920's, but relatively little was accomplished before Smale's program in the 1960's, and the explosion of results that followed in the next fifteen or so years. The subject is, however, far from exhausted, and many important problems remain open (See, for example, Smale 1980).

1 General Discussion

In recent years, attention has been given to models of networks of neurons with recurrent connections (for example: Cleermans, et al. 1988; Elman 1991; Pollack 1991; Smith and Zipser 1989; Williams and Zipser 1989). Such models are more realistic than those lacking recurrent connections, since such networks have the capacity for temporal structure which is not a static function of the temporal structure of input to the system. Along with this increased power, however, there are deep problems which arise when one attempts to make sense of the inner workings of such networks. It will be shown that dynamical systems theory and its tools shed some light on this matter, though what we would like to know soon outstrips what the existing theory can tell us.

Hopfield (1982) showed that networks which use asynchronous parallel update which further have a symmetric connectivity matrix (i.e. $W_{ij} = W_{ji}$, where W_{ij} is the weight from unit j to unit i) have a monotonically decreasing Lyapunov function (also called an energy function) which insures that the network will iterate to a fixed point. Hopfield then went on to suggest that attracting fixed points may be useful for performing abstract computation. Marcus and Westervelt (1989) showed that there also exists a Lyapunov function for synchronous parallel update networks with continuous state variables when the connectivity matrix (also called the weight matrix) is symmetric, which insures that the network will iterate to either a fixed point or to a periodic orbit of period two. Such dynamics will be called tame. They gave further conditions on the weight matrix that ensured that the network would iterate to a fixed point. The condition that the connectivity matrix be symmetric has been criticized as removing the work from the realm of biological importance (p.22, Hertz, et. al.). Hopfield

Relaxing The Symmetric Weight Condition for Convergent Dynamics in Discrete-Time Recurrent Networks

Mike Casey*

Department of Mathematics 0112
UC San Diego
9500 Gilman Dr.
La Jolla, CA 92093-0112

April 7, 1995

Abstract

Hopfield and others have shown that for a fairly broad class of neural network models there exist Lyapunov functions (which insure that the dynamics of the network will remain tame), so long as the coupling matrix is assumed to be symmetric. It will be shown that this condition may be replaced by a more easily satisfied condition. I will then present a few other results which further restrict the complexity of the parameter space of some network models.

*Supported by NDSEG Graduate Fellowship, INCOR grant under Professor Michael Freedman and the San Diego Supercomputer Center. e-mail: mcasey@ucsd.edu

Relaxing The Symmetric Weight Condition for Convergent Dynamics in Discrete-Time Recurrent Networks

Mike Casey

Department of Mathematics 0112
UC San Diego
9500 Gilman Dr.
La Jolla, CA 92093-0112
E-mail: **mcasey@ucsd.edu**

Technical Report INC-9504
Institute for Neural Computation
University of California, San Diego
9500 Gilman Drive DEPT 0523