NEURAL NETWORKS AND NEURAL FIELDS:
Discrete and Continuous Space Neural Models

by

Roderick Edwards
B.A., University of Victoria, 1980
B.Sc., University of Victoria, 1988
M.Sc., Heriot-Watt University, 1990

A Dissertation Submitted in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Mathematics and Statistics

We accept this dissertation as conforming
to the required standard

Dr. R. Illner, Supervisor (Department of Mathematics and Statistics)

Dr. A. Hurd, Departmental Member (Department of Mathematics and Statistics)

Dr. P. van den Driessche, Departmental Member (Department of Mathematics and Statistics)

Dr. N. Dimopoulos, Outside Member (Department of Electrical and Computer Engineering)

Dr. R. Snell, Outside Member (Royal Roads)

Dr. Jacques Bélair, External Examiner, (University of Montreal)

Supervisor: Dr. Reinhard Illner

Abstract

'Attractor' .al network models have useful properties, but biology suggests that more varied dynamics may be significant. Even the equations of the Hopfield network, without the constraint of symmetry, can have complex behaviours which have been little studied. Several new ideas or approaches to neural network theory are examined here, focussing on the distinction between discrete and continuous space neural models. First, simple chaotic dynamical systems are examined, as candidates for more natural neural network models, including coupled systems of Lorenz equations and a Hopfield equation model with a balance of inhibitory and excitatory neurons. Also, continuous space models with a structure like that of the Hopfield network are briefly explored, with interesting training possibilities.

The main results deal with the approximation of Hopfield network equations with a particular class of connection structures (allowing asymmetry), by a reaction-diffusion equation, using techniques borrowed from particle methods used in the numerical solution of fluid-dynamical equations. It is shown that the approximation holds rigorously only in certain spatial regions but the small regions where it fails, namely within transition layers between regions of high and low activity, are not likely to be critical. The result serves to classify connectivities in Hopfield-type models and sheds light on the limiting behaviour of networks as the number of neurons goes to infinity. Standard discretizations of the reaction diffusion equations are analyzed to clarify the effects which can arise in the limiting process. The discrete space systems can have stable patterned equilibria which must be close to metastable patterns of the continuous systems.

Our results also suggest that the fine structure of neural connections is important, and to obtain complex behaviour in the Hopfield network equations, a predominance of inhibition or wildly oscillating connection matrix entries are indicated.

Examiners:

Dr. R. Illner, Supervisor (Department of Mathematics and Statistics)

Dr. A. Hurd, Departmental Member (Department of Mathematics and Statistics)

Dr. P. van den Driessche, Departmental Member (Department of Mathematics and Statistics)

Dr. N. Dimopoulos, Outside Member (Department of Electrical and Computer Engineering)

Dr. R. Snell, Outside Member (Royal Roads)

Dr. Jacques Bélair, External Examiner (University of Montreal)

Table of Contents

List of Figures

Acknowledgements

I am grateful for the generous financial support which enabled me to pursue this research and continue to help support a family. The work was partially supported by an NSERC Postgraduate Scholarship, a University of Victoria Fellowship, and research grants of Dr. Reinhard Illner (NSERC research grant A-7847) and Dr. Pauline van den Driessche (NSERC research grant A-8965).

The experience would not have been as productive or enjoyable were it not for less concrete (but not less significant) forms of assistance from several people. In particular, I thank Reinhard for inspiration and encouragement, Cathy for patience and support, and Tom for empathizing. I also thank Phil at the computer help desk for saving Chapter 1 from magnetic oblivion.

To my father,

who was always interested.

1. Introduction

1.1 Context of neural network research.

Neural networks, as models of biological neural activity and especially as models of computation, have shown great promise. It is clear that biological brains, even very simple ones, are capable of easily performing certain tasks that are extremely difficult to implement in a standard sequential program on a standard sequential computer, despite the great power and complexity of such machines. Such tasks include, for example: recognition of an object in an image; recall of a memorized pattern from a partial or distorted one; and control of limb movements to manoeuvre smoothly around objects. Furthermore, biological brains are capable of learning. The field of neural network modelling developed as an attempt to understand what it is about a large system of relatively simple interconnected units such as biological neurons that allows them to perform such processing tasks so well.

Certain properties of biological neural systems are evident even from a relatively naïve physiological perspective. They are certainly massively parallel systems, rather than single, powerful sequential processors like conventional computers. They are also remarkably fault-tolerant. That is, although their operation is degraded by substantial damage, and functions may be entirely lost with sufficient damage, a moderate loss of processing units or connections does not seriously affect the functioning of the system. Also, and this is a related fact, information is 'stored' in a distributed manner. Memories, for example, are not located in specific memory 'registers' as in a computer, where damage to these specific registers would obliterate the memory.

The physiological study of neural activity has shown that the essential mechanism is the accumulation of electrical potential in the body or 'soma' of a neuron

from the incoming signals from other neurons, leading to the 'firing' of a spike along the neuron's axon when the potential exceeds a threshold. Axons split into many branches and the signals transmitted along an axon are propagated into each branch. These are connected to other neurons via synapses (either directly onto the soma or onto dendrites of the other neuron). Through these synaptic connections, a signal transmitted by one neuron influences the activity of others. The synapse transmits a signal via the emission of neurotransmitters across the synaptic cleft between the synapse and the receiving dendrite or soma. But the signal is modified in the transmission (the signal reaching the synapse from the axon of the sending or pre-synaptic neuron is called the pre-synaptic potential; the signal received by the soma of the receiving or post-synaptic neuron is called the post-synaptic potential). Depending on what type of transmitters the synapse employs, the signal may inhibit or excite the accumulation of potential in the post-synaptic neuron. Also, the synapse may respond to a given pre-synaptic potential with more or less efficiency. The degree to which it amplifies or mutes the pre-synaptic potential is called its 'synaptic efficacy'. The accumulation or integration of the post-synaptic potentials appears to be more or less a straight summation. This summed potential, called the 'membrane' potential, is a state variable for the neuron. After firing, there is a 'refractory period' in which the neuron recovers and regains the ability to fire; it will fire again as long as the membrane potential is still above the threshold. Thus, there is a maximum rate at which the neuron can fire regardless of the strength of the signals it receives. The firing rate may be considered a function of the membrane potential and this function appears to be a sigmoid, with no response from consistently below-threshold membrane potential but a saturated, maximum response for very large membrane potentials. Thus, in a sense, a neuron is a nonlinear adder. (Much of the above description is taken from [5]).

It was originally suggested by Donald Hebb [29], and has since been generally accepted, that the learning process involves modification of synaptic efficacies. Essentially, when pre-synaptic and post-synaptic potentials remain high for sufficient time, that is when both pre- and post-synaptic neurons are very active, a synapse's efficacy is enhanced. This process occurs on a longer time scale than the fast evolution of the neural activities and it is also more permanent in effect. This allows a modification in the system's mode of operation in response to experience.

The above properties of neural systems are reproducible in models using many interconnected simplified 'neurons'. While the details of the electro-chemical processes occurring in biological neurons are suppressed, what should be the essential features of the process are retained to make a relatively simple mathematical model. Nevertheless, the resulting models involve nonlinearity and feedback and realistically consist of (at the very least) many thousands of differential equations (see equations 2.3).

1.2 Development of neural network models.

To design an artificial neural network that performs a useful function requires more than writing down equations for the above described process. These simply describe the lowest level of neural network activity, namely the manner in which individual neurons receive, respond to and distribute signals from and to other neurons. This is essentially a framework in which to work. Higher levels of design (i.e. how to connect the neurons to perform tasks) must involve new principles. Progress in the neural network field is exactly the discovery of such principles.

First, we may ask what functions neural networks can or should perform. Current artificial neural network models perform fairly simple tasks related to per-

ception, memory and motor control. One of the standard applications of neural networks is to function as a 'content-addressable' memory. Rather than locating a 'memory', which may be any meaningful piece of information (we might just consider it a specific pattern of bits), by knowing the address of the memory 'register' where it was stored, a content-addressable memory locates it by content, i.e. by the specific pattern required. A nice example of this process is given by Denker [18]. An incomplete or distorted version of the pattern is used to locate the original complete and undistorted pattern. This makes it clear that the same process can be used for pattern recognition and image restoration tasks, which are inherently similar. Note that all these image related tasks require prior knowledge, i.e., a previously stored pattern to recall.

A related task performed by current neural networks is that of association or classification. It may be required to generate one of a limited set of appropriate responses given a wide range of possible stimuli. Thus, the network must associate the 'appropriate' response with any given input stimulus. What is meant by 'appropriate' depends on the problem but in any case the network can learn (i.e. modify its dynamics) to make responses that are desired either by an external trainer or by internal considerations. That is, a designer 'operating' the network can nudge the dynamics towards making what she considers an appropriate response (supervised learning) or the system can have a built-in way of deciding when and how to modify its dynamics (unsupervised learning). An example of the former is the multi-layer perceptron (see e.g. [58]) and a simple example of the latter is the Carpenter-Grossberg classifier (as described in [45]). If the number of responses is much smaller than the number of possible inputs or stimuli, then the associator acts as a classifier; a class of inputs is mapped to each response.

Aside from these functions, artificial neural networks have been used to perform

optimization tasks (such as finding good solutions to the travelling salesman problem [36]) and for motor control (see, e.g. [11,43]).

Among the most significant simplifying principles or analytic insights that have enabled neural network models to carry out the above functions, are the perceptron convergence theorem, the back-propagation algorithm, and the use of an energy functional and the Hebb 'learning' rule to fix patterns as attractors in network dynamics. The perceptron theorem elucidates the classification properties of certain layered networks of simple additive units called perceptrons, making them prototypical associators/classifiers [56, pp. 109ff]. With the aid of the back-propagation algorithm these networks allow supervised training to produce the desired response to an input when it is known (at least with a good success rate), and then to continue producing responses to inputs when the appropriate response is not known *a priori*. Since our research does not involve these models, we do not describe them further here, but refer the reader to [49,59].

For equations (like our equations (2.2) or (2.3)) modelling the additive neural processes described above, when the matrix of synaptic efficacies (the 'connection matrix') is symmetric, there is a Lyapunov functional, representing the 'free energy' of the system. This observation, apparently made independently by Cohen and Grossberg [12] and by Hopfield [34,35], initiated the study of 'attractor neural networks', ensuring that the behaviour of the network will always be convergent. Moreover, Hopfield showed that a connection matrix may be constructed, in a natural way using something like Hebb's 'learning' rule, so that a set of patterns become fixed point attactors of the dynamics. Thus, given an initial state of the network where the pattern of neural activity is similar to one of the fixed patterns (i.e. is in its basin of attraction), the network will evolve towards the fixed pattern. This can be interpreted as recall of a memory. There are restrictions on the patterns

for effective storage in Hopfield's formulation: they must be reasonably orthogonal (uncorrelated) and consist of roughly half 'on' and half 'off' bits; and there must not be too many of them.

These methods have made possible the performance of the functions described above and a great deal of analysis has been done to assess their capabilities and limitations and to improve their performance. Alternate learning rules remove the restriction of orthogonality of patterns, enhance the storage capacity of the network and reduce the problems of 'spurious' memories, or extra fixed points, which arise in Hopfield's original formulation [18,48]. These alternate learning rules are less biological, however, requiring a synapse to 'know' the states of many remote synapses at once. Additional features have been introduced to extend the capabilities of the first models but still relying on the same essential principles. For example, the use of stochastic effects (simulated annealing) allows the system to evolve not just to the nearest local minimum of the energy functional, but to bounce around until it finds the global minimum or at least a relatively low one. This is useful in optimization problems, for instance [5, pp.89-91]. Similar stochastic methods are applied in a classifier known as the 'Boltzmann machine' [32]. Other techniques allow temporal sequences of patterns to be made attractors of the dynamics, rather than single fixed patterns [39]. Surveys of neural network models may be found for instance in [5,18,28,45,48,65].

1.3 Motivation for new approaches.

The simplifying principles or analytic tools discussed above have proven useful in allowing the creation of artificial neural networks that perform certain tasks reasonably well. The performance of these models is not always as good as might be hoped but there is no doubt that they comprise a new and effective tool for

these tasks that is fundamentally different from other methods. They are sometimes very much faster than other methods and have the fault-tolerance and distributed memory characteristic of biological neural systems.

However, it is not at all clear that these simplifying principles are the ones that nature uses. Although the general features of the underlying dynamical equations (additive inputs, sigmoid response function...) are plausible reflections of nature, the way the models operate and their particular structures are not. There is no natural reason for symmetric connections as required by the Hopfield network (and even though the analysis of Hopfield nets sometimes allows some relaxation of symmetry, in nature connections are manifestly non-symmetric). This is discussed further in Chapter 3. Also, the necessity to supervise learning, at least for primitive brain functions (e.g. perception of objects), is contrary to our observations of nature. It has been suggested that inter-cellular chemical activity in the brain could act as a kind of 'supervisor', for example turning on or off the capacity to learn in response to need. It is also possible that for some functions, one subsystem of a brain could supervise another. However, the kind of supervision needed in the artificial neural network models is too dependent on the knowledge and intervention of the designer to be natural. (Research on unsupervised learning in neural network models has recently been summarized in a paper by Becker [6]). Furthermore, most models require artificially stopping and starting the dynamics, resetting initial conditions or at least presenting inputs at fixed moments to be used for training. In a natural neural system, the neurons are continually active and external inputs, when present, must simply alter their dynamics. There is a fundamental problem in training such an unsupervised system in deciding when an input should be used for retrieving an existing memory and when it should itself be learned. The supervised artificial models are operated in learning mode and retrieval mode separately. Nature

evidently has another solution. It is also not natural for a neural system to have convergent dynamics. As Skarda and Freeman put it, "Convergence to a point attractor amounts to 'death' for the system" [60, p.172]. In fact, measurements of neural activity in biological brains show that complex dynamics are typical. Skarda and Freeman [60] and Traub and Miles [66] claim to have demonstrated that the dynamics can be chaotic in the mathematical sense. In any case, nature clearly does not limit herself to fixed point attractors and periodic oscillators.

In fact, neurophysiological measurement of neural activity has typically been not at the level of individual neurons but averaged activity over an area occupied by many neurons (the electroencephalogram, or EEG). Thus, on the neurophysiological side, researchers are often not even working in the same framework as the artificial neural net modellers. There appears to be interesting behaviour on this level of averaged activity, so modelling at this level is of interest [66, pp.191-193; 26, pp.7-10; 60, pp.163,190].

These observations do not detract from the value of existing artificial neural network models. They *do* suggest that while these models mimic some of the functions of biological neural systems, they do not work in the same way. There is necessarily an interplay between the description and understanding of biological brains on the one hand and the development of abstract models and artificial networks on the other. From the point of view of understanding brain function, it is clearly necessary to try to discover the mechanisms used by nature. However, this may involve simplifying the description so that insight into underlying principles may be obtained. From the point of view of developing useful artificial networks, it is still prudent periodically to take whatever inspiration from nature we are currently capable of comprehending. In particular, if we wish to be able to do more with neural network models than is covered by the list of functions discussed above

(or even to do these with nature's effectiveness), it will be necessary to look for new principles.

In attempting to build neural network models that perform useful functions based on the observed behaviour of biological neural systems, the largest obstacle we face is that there is no real understanding of the computational processes occurring in brains, except in the simplest cases of invertebrate motor control functions and perhaps to some extent in the mapping of images in the visual cortex. Traub and Miles [66, pp.xiii-xv,205] point out that although the activity of parts of the brain can be monitored, no-one knows what computation is being performed in most cases. They study the hippocampus, for example, which is known to contribute to the formation of long-term memory, but how this is done is not known, despite all the detailed experiments. In their work, they design a complex mathematical model closely describing the structure of the hippocampus and describe and compare the activity of both the model and the original, but with no real idea of the significance of what these systems are doing. Nevertheless, the information obtained in experiments like these does give us some clues. If we want to *understand* the processes occurring in brains, we can at least explore the mathematics of the behaviours observed and look for insights into their potential information processing capabilities. Then the modeller can attempt to use them as building blocks for information processing tasks. Even if we do not hit upon the exact process occurring in biological brains, there is the potential for new and useful artificial network models.

For these reasons, we consider it important to attempt to stretch the boundaries of conventional approaches to modelling neural networks. It is not so easy to discover new fundamental principles, but unless new approaches are taken and groundwork is laid they will never be discovered. Even if we simply leave the confines of symmetry in the Hopfield network equations, they become capable of

complex activity and analytic tools in this case are scarce. (We will loosely call these equations the 'Hopfield network equations' from now on, although properly the term 'Hopfield network' refers to these equations with the particular structure of symmetric connections and fixed-point memories — even this usage of the name is perhaps not technically correct, as pointed out by Grossberg [28, p.23] but has become common nevertheless).

1.4 Summary of approaches taken and results obtained.

In this research program it was decided to explore several new approaches to neural network models. These each involve mathematical methods different from those usually employed — we need new mathematical tools for new analytic insights. Two basic ideas initiated the research: chaotic dynamics with input-driven bifurcation, and continuous space modelling. The research in the two areas is essentially disjoint, but some interesting ideas emerged, connecting them. The nature of the research is exploratory and the work of Chapters 4 and 5 in particular is preliminary, whereas that of Chapters 6–8 is more fully worked out. Here we summarize the approaches tried and the results achieved.

1. An initial investigation was made into the possibilities of using chaos in neural networks as a background state, in such a way that appropriate inputs are 'recognized' by changes in the dynamics (based on work of Evans, *et al.* [24], as well as that of Priesol, *et al.*, [51] and Kwan [42]). The models of Freeman [27] and Traub and Miles [66] exhibit interesting behaviour (such as chaos) but these models are too complex to be amenable to analysis. Simpler models are needed for insight. We briefly summarize the ideas in references [24,42 and 51] and explore their possibilities a little further. Some new insights obtained make continued work feasible and potentially fruitful. In particular, it is shown that under

certain conditions, many Lorenz systems (which we interpret as systems of three neurons each) coupled together retain bounded but irregular behaviour and respond to particular inputs by converging to a fixed point. We show that the other model suggested by Kwan and his collaborators [42,51] allows the same possibilities in a discrete time setting, and we also show that these can be reformulated as a Hopfield network with a particular connection matrix structure interweaving positive and negative entries.

2. An attempt was made to explore continuous space versions of the Hopfield network equations (i.e. an integro-differential equation model). This provides a different framework and different insights, even though the results could in theory be transferred back to the standard Hopfield network by space discretization. Using very simple connection functions and external inputs, mainly characteristic functions, it was demonstrated (by construction) that such models can 'learn' in a simple sense while operating continuously and responding only to external inputs (no trainer required to stop and start the system, to switch from learning to retrieval mode, etc.).

3. A more complete and mathematically rigorous investigation was made into the approximation of the Hopfield network equations by reaction-diffusion equations. This is a simplifying principle that provides a different classification of connectivities than the standard symmetric/non-symmetric one, and shows how models with one class of connectivities behave. This had been tried by Cottet [14] for a particular simple type of connection matrix but we have found that the method can be extended to a far wider class of connectivities. Theorems are presented demonstrating the approximation formally, and rigorously proving its convergence under appropriate conditions. A problem arose with the approach that had not been noticed by Cottet, limiting the convergence to regions of high or low activity

away from transition layers. But these transition layers are typically very thin and effects are expected to propagate out from these regions extremely slowly, so the approximation is still quite good from the point of view of an essentially binary-state neural network. These theorems apply to a still restricted class of connectivities and thus serve to classify types of connectivity into those that have behaviour like these PDEs and those that do not. For those that do, the theory of these reaction-diffusion equations can be applied to gain insight into behaviour. In particular, the reaction-diffusion equations give insight into the limiting behaviour as the number of neurons goes to infinity. Also, the types of connectivity that do *not* satisfy these theorems (particularly those with a preponderence of inhibition or those with connectivity matrices with wildly fluctuating entries) promise more complex behaviour than that possible for the PDEs. The Hopfield network formulation of Kwan's model has such matrices.

4. Finally, further analysis on the above reaction-diffusion equations was carried out via standard (finite-difference) space discretizations. In fact, the systems of ODEs resulting from these discretizations are very simple neural networks of the Hopfield equation type, which (as a result of the approximation theorems) are representative of a whole class of Hopfield nets (those that are approximated by the particular reaction-diffusion equation). Different parameters in the reaction-diffusion equation will have different discretizations representative of different classes of Hopfield nets. We prove propositions demonstrating the existence of large scale stable patterns for these very simple Hopfield nets. This is contrary to expectations from the reaction-diffusion equations, which typically have no stable states other than those constant in space. This points out the care that must be taken in deducing properties of discrete from continuous systems and vice versa, despite a rigorous (almost) convergence theorem. Previous analytic research suggests that metastable

states of the reaction-diffusion equations are responsible.

Chapter 2 presents background material on the conventional Hopfield network (and some enhancements). A complete summary of the extensive literature on this subject is not attempted; rather the main mathematical properties of the model that have relevance to later chapters are described. Chapter 3 contains a discussion of the limitations of the conventional Hopfield network and related models and outlines an alternate set of features or properties that we consider desirable in a neural system model. In Chapters 4 through 8 we develop and explore models exhibiting some of these features.

2. Background on the conventional Hopfield network model

2.1 The Hopfield network equations.

A good deal of attention has been paid in the last few years to the type of neural network model often referred to as the 'Hopfield network' since Hopfield's important contribution to the study of such systems [34,35]. Of the many discussions of this type of neural network, one of the best and most comprehensive is the book by Amit [5]. The equations describing the dynamics of these networks are of the form

$$\dot{u}_i = -\alpha u_i + \sum_j T_{ij} g(\lambda u_j), \tag{2.1}$$

where i and j are indices over all neurons in the network, u_i represents the membrane potential of the i^{th} neuron, $\alpha > 0$ is a 'leakage' rate (or resistance parameter in Hopfield's formulation), T_{ij} is the 'synaptic efficacy' modulating the effect of neuron j on neuron i, and g is a sigmoid response function with 'gain' $\lambda > 0$, describing how a neuron's firing rate depends on its membrane potential. We have also followed the common practice of assuming that neurons are identical here (i.e. α, g and λ do not depend on i). These equations may also have external input terms, c_i, and threshold terms, θ_i, describing signals to each neuron arriving from outside the net and firing thresholds other than zero. For example,

$$\dot{u}_i = -\alpha u_i + \sum_j T_{ij} g\left(\lambda(u_j - \theta_j)\right) + c_i . \tag{2.2}$$

As a model for biological neural networks this is clearly a great simplification but it nevertheless extracts some features of their design. Artificial neural networks using these dynamics have proven useful in some applications (as mentioned in Section 1.2). There are, of course, many models of neural nets (see, for example,

the survey papers [28,45,48,65]). The dynamics of many of the models have the form of equations (2.2).

The dynamical behaviour of these equations is in general complex and difficult to describe. Much of the literature has concentrated on the special case of a symmetric connection matrix, T (i.e. $T_{ij} = T_{ji}$). This is mathematically convenient as there exists an energy functional in this case which ensures convergent behaviour of the net [35]: Any initial conditions approach a fixed point. Convergence is useful in the application of these neural nets to content-addressable memory or the retrieval of patterns ('memories') from distorted or similar versions of them; it is possible to construct the transition matrix so that selected patterns become fixed points of the dynamics (discussed below).

The Hopfield network equations with external inputs and thresholds may be written as in (2.2) or, if instead we let u_i represent the amount by which the membrane potential of neuron i exceeds its threshold, we may write them alternately as

$$\dot{u}_i = \sum_j T_{ij} g(\lambda u_j) - \alpha u_i + c_i - \alpha\theta_i . \tag{2.3}$$

We require $\alpha > 0$, $\lambda > 0$ and g sigmoidal in shape, increasing on $\mathbf{R}$ and bounded. Typically we use $g : \mathbf{R} \to (0,1)$ or $g : \mathbf{R} \to (-1,1)$ (the effect of this choice is discussed below). To be precise we will assume from now on (except in Chapter 5):

$$g : \mathbf{R} \to (-1,1),\ g \in C^1,$$

$$g'(x) > 0, \tag{2.4}$$

$$g'(x) \leq g'(0) = \frac{1}{\beta}$$

(Fig. 2.1 gives an example of such a function with a gain parameter, $v = g(\lambda u)$).

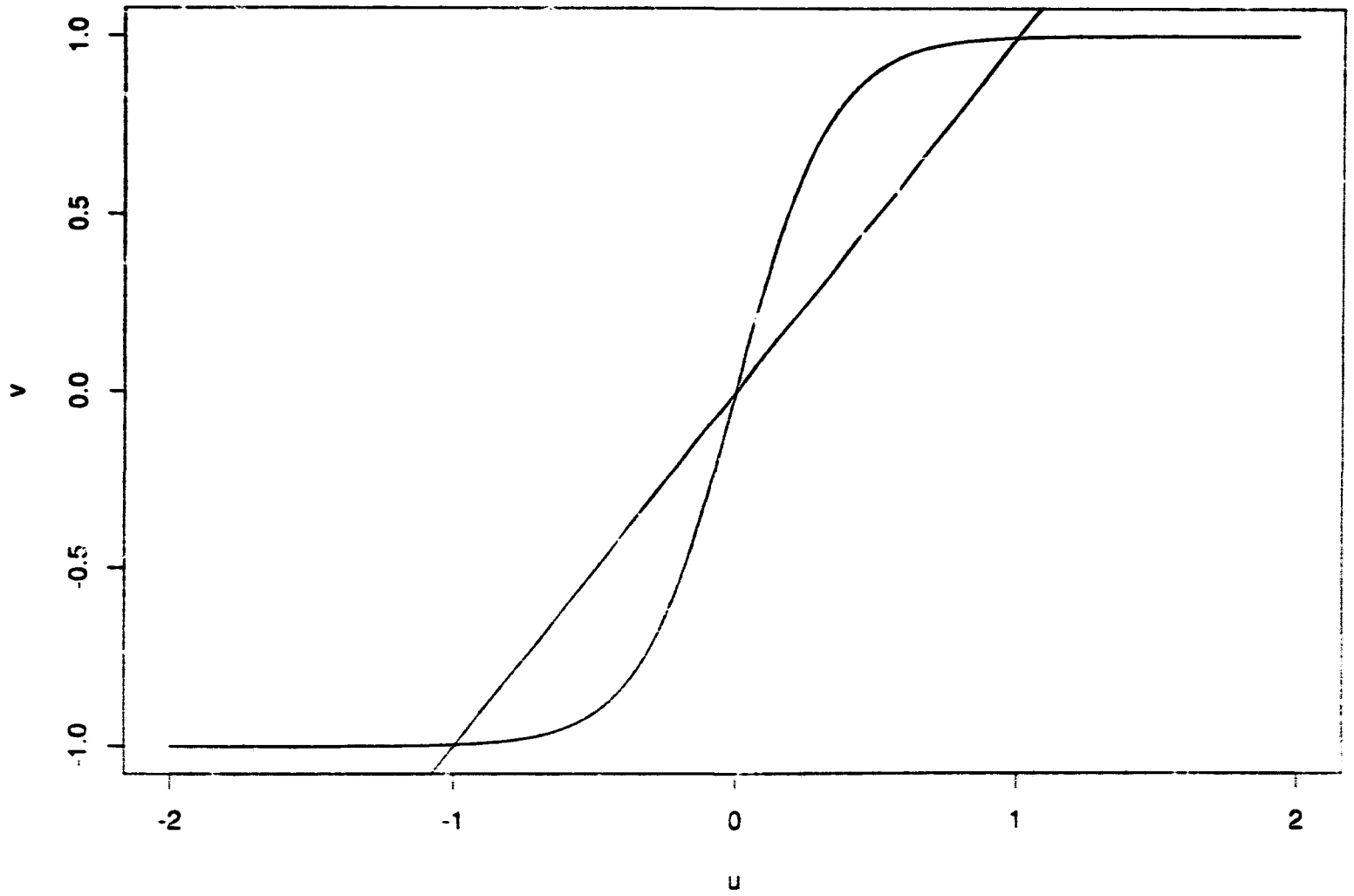

Figure 2.1 An example of a sigmoid response function: $v = g(\lambda u)$, with $g = \tanh$ and $\lambda = 3$. The straight line is $f(u) = u$.

We present below some basic properties of the Hopfield network equations. Although they are well known, they provide a background to the material of subsequent chapters. Indeed, many of these properties carry over directly to other forms of the model explored later.

2.2 Energy functional for the Hopfield network.

Hopfield [35] showed that there exists an energy functional (or Lyapunov functional) for equations (2.3) in the case where T is symmetric:

$$E[u] = -\frac{1}{2}\sum_i\sum_j T_{ij}v_iv_j + \frac{\alpha}{\lambda}\sum_i\int_0^{v_i} G(V)dV - \sum_i(c_i - \alpha\theta_i)v_i\,, \tag{2.5}$$

where $G = g^{-1}$ and

$$v_i = g(\lambda u_i) \tag{2.6}$$

represents the firing rate of neuron i. In fact, along solutions to (2.3),

$$\begin{aligned}
\dot{E} &= -\frac{1}{2}\sum_i\sum_j T_{ij}\left(v_i\dot{v}_j + v_j\dot{v}_i\right) + \frac{\alpha}{\lambda}\sum_i G(v_i)\dot{v}_i - \sum_i(c_i - \alpha\theta_i)\dot{v}_i \\
&= -\frac{1}{2}\sum_i\sum_j \left(T_{ij} + T_{ji}\right)v_j\dot{v}_i + \frac{\alpha}{\lambda}\sum_i G(v_i)\dot{v}_i - \sum_i(c_i - \alpha\theta_i)\dot{v}_i \\
&= -\sum_i\left(\sum_j T_{ij}v_j - \frac{\alpha}{\lambda}G(v_i) + c_i - \alpha\theta_i\right)\dot{v}_i \\
&= -\lambda\sum_i\left(\sum_j T_{ij}g(\lambda u_j) - \alpha u_i + c_i - \alpha\theta_i\right)g'(\lambda u_i)\dot{u}_i \\
&= -\lambda\sum_i g'(\lambda u_i)\dot{u}_i^2 \le 0\,,
\end{aligned}$$

since $g' > 0$, so that energy decreases except at equilibria. Also

$$\frac{\partial E}{\partial u_i} = -\lambda g'(\lambda u_i)\left(\sum_j T_{ij}g(\lambda u_j) - \alpha u_i + c_i - \alpha\theta_i\right) = 0$$

at equilibria. For given parameters, E is clearly bounded below, since v_i is bounded. Thus, solution trajectories for equation (2.3) must descend the energy surface towards a local minimum, which must be an equilibrium point.

If T is not symmetric, then the E above is not a Lyapunov functional for the system. Convergent behaviour appears to be slightly robust in regard to asymmetry (particularly for randomly diluted, but otherwise symmetric connection matrices [5, p.363ff.]) but in general non-convergent behaviour is to be expected (see, e.g., [53,54,61,64]).

We will revisit these energy functionals several times, for analogous integral equation models (Chapter 5), reaction-diffusion equation models (Chapter 7) and their finite-difference discretizations (Chapter 8).

2.3 Membrane potentials, firing rates and the S - Σ exchange.

We can re-express equation (2.3) in terms of firing rates, v_i, as follows:

$$u_i = \frac{1}{\lambda} G(v_i),$$

from (2.6), so

$$\dot{u}_i = \frac{d}{dt}\left(\frac{1}{\lambda}G(v_i)\right) = \sum_j T_{ij} v_j - \alpha\left(\frac{1}{\lambda}G(v_i)\right) + c_i - \alpha\theta_i\,,$$

$$\dot{v}_i = \frac{1}{G'(v_i)}\left[\lambda \sum_j T_{ij} v_j - \alpha G(v_i) + \lambda(c_i - \alpha\theta_i)\right]. \tag{2.7}$$

This is entirely equivalent to (2.3) for initial conditions $v_i(0) \in \text{range}(g)$. It is easy to see that solutions to equations (2.3) and therefore to equations (2.7), are bounded and so exist globally in time and are unique (see Section 2.6).

This technique of switching between different variables proves very useful in later chapters, for integral equations and reaction-diffusion equations corresponding to the Hopfield network equations. In the case of discrete time or equilibrium equations this change of variables takes a particularly simple form. We give an abstract formulation of this idea in Section 5.5.

We note here that discrete time versions of the Hopfield network equations can be obtained via what has been called by Grossberg [28, p.26] the $S - \Sigma$ exchange, from equations describing the actual firing of each neuron. If x_i represents the 'action potential' of a neuron, i.e. the signal being sent down its axon, we can suppose that $x_i(t)$ takes on the values $+1$, meaning that a spike is emitted along the axon at this time, or -1, indicating that no spike is emitted. The activity can then be modelled (with great simplification from the biological reality) as

$$x_i(t+1) = \mathrm{sgn}\left[\sum_{j \neq i} T_{ij} x_j(t) + c_i - \theta_i\right], \tag{2.8}$$

where $\mathrm{sgn}(y) = 1$ if $y > 0$ and $\mathrm{sgn}(y) = -1$ if $y < 0$ (it's value when $y = 0$ is not critical), and θ_i and c_i are the thresholds and external inputs respectively. The evolution is taken to be asynchronous, i.e. only one (randomly selected) neuron can change state at each time step and it fires at the next time step if its accumulated incoming signals exceed its threshold. The lack of connection from a neuron back onto itself ($T_{ii} = 0$) is necessary in the discrete time case for the existence of the energy functional. The $S - \Sigma$ exchange consists of the following transformation. Let

$$u_i(t) = \sum_{j \neq i} T_{ij} x_j(t) + c_i\,, \tag{2.9}$$

which is the i^{th} neuron's membrane potential, and now multiply (2.8) by T_{ki}, sum

over i and add c_k, using (2.9) to get

$$u_k(t+1) = \sum_{i \neq k} T_{ki} \mathrm{sgn}(u_i(t) - \theta_i) + c_k \,. \tag{2.10}$$

This is a discrete time analogue of equation (2.2) with $\alpha = 1$. The $S - \Sigma$ exchange may be reversed via

$$x_i(t+1) = \mathrm{sgn}(u_i(t) - \theta_i) \,. \tag{2.11}$$

The discrete and continuous time Hopfield network models (using equations (2.10) and (2.2) respectively) are formally analogous and both modelling approaches have been taken. Hopfield [35] showed that if the gain in (2.2) is large, the solutions to the two systems have similar behaviour. (Of course, the high gain limit of $g(\lambda u)$ is the 'hard nonlinearity', $\mathrm{sgn}(u)$).

2.4 The sigmoid response function and the high gain condition.

In the literature g is often taken to be an odd function taking values in $(-1, 1)$, such as tanh. In particular, such a function has $g(0) = 0$. Horizontal shifts in the response function may be accounted for by a threshold term. A more realistic sigmoid might, however, take values in (0,1). For example, a logistic function is often used:

$$g(\lambda u) = \frac{1}{1 + e^{-4\lambda u}} \,.$$

However, by a change of coordinates, the resulting Hopfield-type equation can be transformed into the original one with an additional threshold term. Equation (2.3) can still be transformed to equation (2.7) and then we let $w = 2v - 1$ so that $\dot{w} = 2\dot{v}$. Then

$$\dot{w}_i = \frac{2}{G'\left(\frac{w_i+1}{2}\right)} \left[\lambda \sum_j T_{ij} \left(\frac{w_j + 1}{2} \right) - \alpha G \left(\frac{w_i + 1}{2} \right) + \lambda(c_i - \alpha\theta_i) \right] .$$

Now let $F(w) = 2G\left(\frac{w+1}{2}\right)$, so that $F'(w) = G'\left(\frac{w+1}{2}\right)$. Then

$$\dot{w}_i = \frac{1}{F'(w_i)}\left[\lambda \sum_j T_{ij} w_j - \alpha F(w_i) + \lambda\left(2c_i - 2\alpha\theta_i + \sum_j T_{ij}\right)\right],$$

which is of the same form as before with inputs $2c_i$ and thresholds $\left(2\alpha\theta_i - \sum_j T_{ij}\right)$.

Thus, horizontal and vertical shifts in the response function do not significantly alter the model, except in changing the threshold values. If $g(0) = 0$, then the Hopfield equations without inputs or thresholds (2.1) have the steady state solution $u_i \equiv 0$. In biological nets, where firing rates should be strictly positive, this may not make sense, but there is no reason not to create artificial nets with this property if it is desired.

In the absence of input or threshold terms in the Hopfield network equations (2.3), if the gain, λ, is small, then all solutions decay to zero. This may be proved either by a contraction mapping argument or (in the symmetric case) from the energy functional. We will not give a proof of this result here, but we will prove it for analogous equations in Section 5.3, using the contraction mapping theorem and in Section 8.3 using the energy functional. It is a straightforward matter to modify these proofs for the standard Hopfield equations.

2.5 Learning rules for the Hopfield network.

'Learning' in a conventional Hopfield net is accomplished by setting

$$T_{ij} = \sum_{m=1}^{s} v_i^{(m)} v_j^{(m)},$$

for $i \neq j$, and $T_{ii} = 0$, where $v^{(m)}$ represents the m^{th} pattern to be learned with $v_i^{(m)} = \pm 1$, say. If the number of patterns s is not too large and the patterns are

approximately orthogonal then the patterns will be close (in phase space) to fixed points of the dynamics. This is shown in [34,35], though the orthogonality condition takes a slightly different form when $v_i \in \{0, 1\}$ as in Hopfield's original formulation. It is not guaranteed that these will be the only fixed points of the dynamics; in fact, there are usually extraneous fixed points, called 'spurious memories'.

This 'learning rule', loosely referred to as Hebbian learning, is not the only one that has been applied to the Hopfield network equations. Some others (and the 'Hebbian rule') are described by Denker [18] and Michel and Farrell [48], for example. In particular, we mention here the Adaline rule and the 'geometric' or 'pseudo-inverse' rule. Like the 'Hebbian' rule used by Hopfield, both of these can be built up one pattern at a time, so that additional memories can be added without doing the calculation for all patterns from scratch. This incremental 'learning process' is obviously necessary from the biological viewpoint. The Hebb rule expressed incrementally is

$$T_{ij}^{(m+1)} = T_{ij}^{(m)} + v_i v_j \,,$$

where $\mathbf{v} = (v_i)$ here is the new pattern to be stored [18, p.224]. The Adaline rule is

$$T_{ij}^{(m+1)} = T_{ij}^{(m)} + \Gamma v_i^{\perp} v_j \,,$$

where $v_i^{\perp} = v_i - \frac{1}{N} \sum_k T_{ik}^{(m)} v_k$, N is the number of neurons in the network and $\Gamma < 1$ is a positive parameter [18, pp.224-225]. This rule takes the component of $\mathbf{v}$ orthogonal to the span of the previously stored patterns, and thus does not require initially orthogonal patterns. A variant of this idea which retains symmetry is the 'geometric' rule [18, pp.225-227]

$$T_{ij}^{(m+1)} = T_{ij}^{(m)} + \Gamma \frac{v_i^{\perp} v_j^{\perp}}{\mathbf{v}^{\perp} \cdot \mathbf{v}^{\perp}} \,,$$

where $\mathbf{v}^{\perp} \cdot \mathbf{v}^{\perp} = \sum_k \left(v_k^{\perp}\right)^2$. This last rule guarantees exact storage of the desired patterns with no 'spurious memories' and equal depths of the patterns in the energy surface.

From the computational point of view, these have considerable advantages over the Hebb rule suggested by Hopfield. However, the Hebb rule has some basis in biology, while the others above do not. In particular, the calculation of v_i^+ requires knowledge at the i^{th} neuron of the strengths of synaptic connections across the network. This information is not available to biological neurons.

2.6 The role of external inputs.

Little of the literature on the conventional Hopfield network deals with the role of external inputs. Hopfield's formulation of the network does not require them. Rather, initial conditions serve as 'inputs,' and the solution evolves towards a (pre-established) fixed point. Inputs are used effectively in applications of Hopfield nets to optimization problems (see e.g. [36]), but this is a purely mathematical application, not a biological simulation.

We note here that sufficiently strong external inputs can dominate the behaviour of the network equations. This may be significant in the unsupervised operation of a biological neural system, as there must be stimulus from outside to distract it from its meanderings and to provide experience from which to learn. In effect, a strong input can reset the system, the equivalent (without supervision) of artificially restarting the evolution from a new set of initial conditions.

First, we need a maximum principle to ensure boundedness of solutions. Consider equation (2.3) without the input term, c_i. Suppose that u_i is positive and that $\dot{u}_i$ is also positive. Then

$$\alpha u_i < \sum_j T_{ij} g(\lambda u_j) - \alpha\theta_i < \sum_j |T_{ij}| - \alpha\theta_i$$

which is a constant. Similarly for u_i and $\dot{u}_i$ negative,

$$\alpha u_i > \sum_j T_{ij} g(\lambda u_j) - \alpha\theta_i > -\sum_j |T_{ij}| - \alpha\theta_i .$$

These two together show that $u_i\dot{u}_i > 0$ implies

$$\alpha|u_i| < \sum_j |T_{ij}| + \alpha|\theta_i| .$$

Therefore, if

$$\alpha|u_i| \geq \sum_j |T_{ij}| + \alpha|\theta_i| \equiv B_i , \tag{2.12}$$

then $u_i\dot{u}_i < 0$ and $|u_i|$ cannot grow. For initial data within these bounds (2.12), the solution must always remain within them.

If input is added, we still have a maximum principle with the term $|c_i|$ added to the bound B_i, so solutions are still bounded. But while an input is active,

$$\alpha u_i < B_i \Rightarrow \dot{u}_i > c_i - 2B_i ,$$

$$\alpha u_i > -B_i \Rightarrow \dot{u}_i < c_i + 2B_i ,$$

from (2.3). Thus,

$$|c_i| > 2B_i \Rightarrow \dot{u}_i c_i > 0$$

for as long as $\alpha|u_i(t)| < B_i$.

So if u_i is initially within the bounds (2.12) and then a strong input is applied, $|c_i| > 2B_i$, then $\dot{u}_i$ has the sign of c_i, and u_i is pulled toward c_i, at least until it exceeds the bounds in (2.12). Thus, wherever a strong input is applied, the neural activity responds in kind, and the firing rate becomes high (or low) when the input is positive (or negative).

3. Limitations of the conventional Hopfield network

The Hopfield network, properly speaking, consists of the dynamical system given by (2.2), (2.3) or (2.7), with initial conditions representing an input pattern, and a symmetric connection matrix determined beforehand by the 'Hebbian' learning rule to fix chosen patterns into memory. A great many variations on this scheme have been put forward in the literature on neural networks, a great deal of attention has gone into determining the memory capacity of such a network and ways have been developed to remove some of its problems, such as the necessity for 'nearly' orthogonal memories, and the existence of 'spurious memories'. There are learning rules, for instance, that enable any input patterns to be stored (still with a maximum capacity of course) and that prevent the occurrence of spurious memories (as discussed in Section 2.5).

However valuable these neural network models are from the point of view of their capabilities in the abstract, they do not correspond very closely to biology. This is not an objection to their study — on the contrary, it is surely advantageous to follow two lines of research, one that keeps an eye on biology and attempts to understand and model its processes more deeply, and the other that takes insights already gained from biology as a starting point and attempts to reshape them to maximum effect in an abstract or artificial setting. Sometimes insights may arise from surprising directions, such as Hopfield's importing of statistical physics into neural network theory. However, there is, no doubt, a limit to how much can be expected from the standard Hopfield network. It has taken its place as a useful tool for certain tasks, such as 'content addressable memory', image restoration, classification, and will continue to be explored in relation to these tasks. But the study of biological brains requires us to step outside the bounds set by the

standard Hopfield model. The same goes for other models and techniques, such as perceptrons, Boltzmann machines, simulated annealing, and so on, which are subject to many of the same limitations.

The Hopfield network allows storage of patterns in a distributed, fault tolerant way, as attractors of a dynamical system. Biological memory also has these features: They are certainly distributed and fault tolerant and may well be attractors of a dynamical system. The equations of the Hopfield network also characterize features of the operation of biological neural networks: the summation of inputs, the modification at synapses, excitation and inhibition, etc.

However, it is certain that the mechanism used by the Hopfield network to store memory patterns is not one used in biological neural networks. The convergent behaviour of the Hopfield network depends heavily on the symmetry of connections and this certainly has no biological reality. Other learning rules (such as the Adaline and 'geometric' rules discussed briefly in Section 2.5) bring asymmetries into the network, but rely on other non-biological mechanisms (such as use of non-local information at synapses). Multi-layer perceptron networks, once trained and presented with an input, still have the same nonlinear additive elements as a Hopfield network with a particular feed-forward connectivity structure. They are also convergent in their behaviour, in this case converging to a particular set of states of the output neurons. Moreover, the feed-forward structure is not applicable to most areas of the brain (see e.g. [66, p. 210], regarding the hippocampus, and [60, p.171], regarding the olfactory bulb), though it may have some relevance to the visual system.

More generally, there are good reasons why neural nets (perhaps even artificial ones) should not simply converge to fixed points (see e.g. the paper of Skarda and Freeman [60] and the discussion following; also [50]). Biological neural nets, whether we refer to the entire brain of an organism or a subsystem of a brain, do

not simply converge to fixed levels of neural activity. Some neural network models exhibit oscillatory behaviour rather than fixed point attractors (e.g. [23,40]). The biological relevance of weakly coupled oscillators has been questioned by Traub and Miles [66, p.211], since their observations indicate that coupling is very strong (at least in the hippocampus).

It is evident from experimental studies [60,66] (and references in [24,42]) that biological brains commonly exhibit complex dynamical behaviour, that has been characterized as chaotic in some cases. The Hopfield network equations are certainly capable of chaotic activity, but not under the condition of symmetric connectivity. In particular, it has been suggested [24,27,60] that in some subsystems of the brain, particularly those involved in perception, the 'ground state' is chaotic and input causes bifurcations in the dynamics resulting in more regular activity. Traub and Miles also observe chaotic activity interrupted by periods of more regular behaviour (synchronized firing of large numbers of neurons), but no convergence to fixed points, in their work on the hippocampus [66, p.119,175ff,208,210].

Perhaps even more fundamentally, biological networks must have the capacity for unsupervised learning (although chemical conditioning of neural activity, i.e. extracellular effects, could be considered a kind of supervision, and one subsystem of a brain might 'supervise' another). They must also operate in a continuous, unsupervised way, rather than simply evolving from preset initial conditions to converge on preset equilibria and then stop. They must be capable of responding to external inputs by recalling associated, previously stored information as well as by storing the new information, and doing one or the other or both when appropriate.

Hence, it seems likely that something can be learned by studying more general neural network models and in particular non-symmetric ones. When the Hopfield network equations are allowed to have a non-symmetric transition matrix, the range

of possible dynamics is much greater. It appears, on the basis of numerical studies [53,54] and analytic studies (e.g. [61,64]), that their behaviour is often chaotic. We explore the types of dynamics possible and develop analytic techniques to help do this. All of the above references have studied randomness in Hopfield network connection matrices, but there have also been some analyses on connection matrices with very particular structures other than symmetry (see, e.g. [21]).

Finally, there is a modelling decision that must be made as to whether to take a microscopic or macroscopic approach. Few neural network models keep track of the internal electro-chemical processes involved in the functioning of neurons and synapses. There is clearly a difference of scale here, or a hierarchy, in which the details of operation of lower levels are not critical to the operation of higher levels. Salient features of the lower level processes must be extracted to be used by higher level processes. While this is not a clear-cut task, it is unlikely that the details of the chemistry of neuro-transmitters, for example, play a critical role in the large scale movement of information amongst large clusters of neurons in a network.

The same question can be asked on a still higher level. While most neural network models keep track of the activities of individual neurons (or at least claim to model either individual neurons or groups of neurons that act as a unit like an individual neuron), some have modelled neural activity on a higher level. Freeman describes a hierarchy of levels of neural information processing, claiming that the higher levels are the most relevant for behaviour [26]. Another possibility, one used by Amari [3,4] for instance (see also [13,57]), is to simply treat neural activity as a quantity depending on a continuous space variable. Thus, individual neurons are blurred out, and we have a model of neural activity averaged over an area. This is something like what is measured in an EEG reading (see e.g. [60, pp.163,190; 66, pp.191-193]).

It is a question of some interest as to whether there are essential differences in continuous space versus discrete space models. Does the blurring of activity of individual neurons preclude important dynamical behaviours? Do individual neurons make a difference? One reason that the continuous space approach is attractive is that we may want to know what happens to the overall behaviour of a neural network as the number of neurons approaches infinity (the numbers involved in biological brains are, of course, enormous). We can examine a continuous space neural activity variable as a limit of an increasingly dense discrete space variable. However, there is a possibility that assuming the neural activity variable is a continuous (or otherwise smooth) function of space may limit the dynamical possibilities. A mathematical analysis of the limiting process may tell us something about the differences (see Chapters 6 and 8).

To summarize the above discussion, we explore possible models with some or all of the following features:

1. Asymmetry.
2. Non-convergent (for example chaotic) behaviour, at least as a background state.
3. Unsupervised activity (as opposed to the setting and resetting of initial conditions).
4. Response to external inputs.
5. Unsupervised learning.
6. Ability to retrieve and/or learn as appropriate.
7. Neural activity a function of a continuous space versus discrete space variable.

In the following chapters, we attempt to explore some of these possibilities.

Chapter 4 deals with chaotic dynamical systems that may be interpreted as components of neural networks and respond to significant inputs by becoming regular (converging to equilibria). Chapter 5 deals with continuous space versions of the Hopfield network, an integro-differential equation model and discrete time versions of it. Though the capabilities of the model developed are rather limited, it does provide an example of continuous, unsupervised activity and unsupervised learning with the ability to be coaxed into retrieval or learning, depending on the input. The new training rule suggested could also be transferred back to the discrete space Hopfield network. Chapter 6 deals with a rigorous approximation of solutions to the system of Hopfield net equations by solutions to a single quasi-linear partial differential equation of reaction-diffusion type, via the integro-differential equation, and thus gives some insight into the discrete versus continuous space question and the limiting process. This derivation of partial differential equations as macroscopic laws ignoring the distinction between individual neurons is in the same spirit as the laws of thermodynamics and statistical mechanics where important properties do not depend on keeping track of the motions of individual particles. Chapter 7 deals with further analysis of the PDEs that arise from the approximation in Chapter 6 and Chapter 8 deals with finite-difference discretizations of them.

4. Chaotic neural networks

4.1 Chaos in neural networks: why and how?

The coupling of 'chaos' with neural networks, both popular topics, has been examined by a number of investigators. Interesting discussions of the place of mathematical chaos in neural nets can be found in [24,42,50,51,53,54,60,61,64].

There appear to be three main reasons for the presence of chaotic activity in neural networks, that have been suggested in the literature:

1. A background chaotic state allows access to many parts of phase space [60, p.168]. External inputs can cause bifurcations to produce stable fixed point or periodic attractors, and potentially many of these can arise from bifurcation when the ground state is chaotic.

2. A chaotic state may indicate an 'unrecognized' input [50, p.L677; 60, p.168]. Many neural network models will converge for any input. Thus, even if an input is not near any memorized pattern, it will either converge to one, or to a spurious memory. It is desirable for a network to have the capability of indicating when it does not recognize an input (i.e. indicating that an input is not near any of its memorized patterns). One way to do this would be for a chaotic network to remain chaotic unless the input was near a known pattern.

3. Chaotic or otherwise ergodic wandering may effectively prevent learning when there is nothing to be learned, which is important if the system must be constantly active and ready to learn when appropriate [50]. Random patterns of activity should not be allowed to alter synaptic efficacies. If learning results from sustained activity patterns of long duration, then when an activity pattern is not sustained, and particularly when the time-averaged activity is essentially zero,

positive and negative influences cancelling each other out, no significant 'learning' (synaptic changes) can occur. Parisi [50], for example, has proposed a model including an equation for the evolution of the synaptic efficacies. They evolve on a slower time scale than the neural activities and changes depend on the combination of pre- and post-synaptic potentials (Hebbian rule) but time-averaged potentials are used, which should prevent synaptic change as long as the neural activity wanders ergodically.

It should be pointed out here that there is nothing in the above reasons that requires chaos in a strict sense. All that is required is activity that wanders over large regions of phase space in such a way as to cancel out net changes to synapses caused by neural activity levels. Random behaviour or even very complex long-period behaviour could do as well. However, it may be easier for a truly chaotic dynamical system to reach many other types of behaviour by bifurcation.

Among the mechanisms suggested in the literature for introducing chaos into simple neural network models are

1. High gain in a Hopfield network with random connection matrix.

2. Asymmetry in the Hopfield network connection matrix.

3. Dynamic thresholds.

4. Quadratic nonlinearities.

We discuss each of these below.

1. High gain in a Hopfield network with random connection matrices [61]. In the low gain regime solutions simply decay to zero; with medium gain, stable patterns occur; and with high gain, chaos results. As the number of neurons in the system, N, approaches ∞, the transition from the low gain to high gain regimes becomes sharp.

2. Asymmetry in a Hopfield connection matrix. Renals [53] and Renals and Rohwer [54] simulate (continuous time) Hopfield nets with random connection matrices having varying degrees of asymmetry and find more or less chaotic behaviour (following Hopfield's similar experiments for the discrete time net [34]).

Parisi [50] suggests introducing asymmetry into the conventional Hopfield network (with Hebbian learning) by random dilution of the connection matrix (setting random entries to zero). This leaves open the question of whether memories can still be retained but one expects that if the dilution is not too great, such stable fixed points should remain (even though co-existing with chaotic activity in other regions of phase space). This has since been demonstrated by Derrida, Gardner and Zippelius [19], and explored further by Tirozzi and Tsodyks [64].

It can be shown that as long as memories (equilibria) are patterns with extreme values (blacks and whites rather than greys) then local stability is guaranteed, despite possible chaotic activity elsewhere in phase space. This can be seen easily if we write the Hopfield net equations (2.2) with zero thresholds in matrix form

$$\dot{\mathbf{u}} = -\alpha \mathbf{u} + Tg(\lambda \mathbf{u}) + \mathbf{c}\,,$$

where $\mathbf{u}$ is the vector of membrane potentials, T is the transition matrix, $\mathbf{c}$ is the input vector, and g is applied to $\lambda\mathbf{u}$ componentwise. For an equilibrium $\mathbf{u}^*$, linearized stability depends on the eigenvalues of the Jacobian:

$$TF - \alpha I\,,$$

where I is the identity matrix and F is a diagonal matrix with $F_{ii} = \lambda g'(\lambda u_i^*)$. If $|u_i^*|$ is large enough for all i, then $g'(\lambda u_i^*)$ is small (see Fig. 2.1) and the eigenvalues are all negative. (This criterion for stability has been pointed out by Sudharsanan and Sundareshan [63], for example).

Kwan and his collaborators [42,51] have proposed a model with a particular combination of inhibitory and excitatory neurons designed to produce chaotic activity. It turns out that this model can be reformulated as an asymmetric Hopfield net (see Section 4.4).

The asymmetric learning rules for the Hopfield net discussed in Section 2.5 are not designed to generate chaos but are designed to retain convergent behaviour, or at least to ensure that desired patterns are stable fixed points.

3. Dynamic thresholds have been used by Horn and Usher [37,38] and Hendin, Horn and Usher [30] to produce chaos in a different way. They simulate the phenomenon of neural fatigue, the gradual raising of the firing threshold of a neuron that is consistently firing at peak rate. If thresholds move towards the activity levels of their neurons, they can destabilize a fixed point after the network has rested there a while. This forces an otherwise convergent net to eventually go off to another fixed point and ultimately to wander chaotically among the set of available fixed points (suggesting free association).

4. The introduction of quadratic terms into the network equations (and thus departing from the standard Hopfield-type equations) can be used in judicious ways to generate chaos. In fact, one may start with a chaotic model and attempt to perform information processing tasks with it regardless of biological neural mechanisms. The simplest such chaotic dynamical system is the well-known Lorenz system of equations which can be given a rough neural network interpretation, and can be used in conjunction with external forcing terms for simple information processing tasks. This idea is discussed in Sections 4.2 and 4.3. Note that although some of Grossberg's models have quadratic terms they are not used to generate chaos [28].

For a chaotic neural network to be useful, it should be possible for the dynamics to become more regular when information is being processed. Skarda and Freeman

[60], as well as Evans, *et al.* [24], argue the case for input-driven bifurcation in the dynamics. The idea is that the network has chaotic idling dynamics, and inputs (at least some inputs) cause a bifurcation stabilizing a fixed point attractor or even a periodic attractor. Different inputs should cause convergence to different fixed point (or periodic) attractors, and there should, ideally, be a way to make use of new attractors as new inputs are learned. Whether the stable patterns should be simply reflections of the input or internal representations (i.e. internally 'meaningful' patterns consistently produced by the same input) is an interesting question; in the work of Freeman and his collaborators (discussed in [60]), which is based closely on the biology of the mammalian olfactory bulb and cortex, internal representations are produced.

Two attempts at very simple models with some of these features are discussed below.

4.2 The Lorenz equations as a neural network.

Evans, Illner and Kwan [24] suggest a framework for input-driven bifurcation to stable dynamics with a chaotic 'idling' state. They propose a dynamical system describing neural activity, with an external force that can be turned on and off in the form

$$\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x}) + \alpha \mathbf{y} \chi_{[t_1,t_2]},$$

where $\mathbf{x}$ is the neural activity vector, $\mathbf{F}$ is a vector valued function describing the effects of neurons on each other, $\mathbf{y}$ is a unit input vector, α is an input strength, and χ denotes a characteristic function on a set. They propose that the system without the input should be chaotic and that inputs should have the ability to change the dynamics in a way that depends on the input.

They provide a concrete example of such a system, namely the Lorenz system, which is well known to be one of the simplest systems of differential equations displaying chaotic activity (with appropriate choices of parameters) [62]. These equations are

$$\begin{aligned} \dot{x} &= -\sigma(x - y) \\ \dot{y} &= -y + rx - xz \\ \dot{z} &= -bz + xy \end{aligned} \tag{4.1}$$

and the usual choice of parameters (following Lorenz) is $\sigma = 10$, $r = 28$, and $b = 8/3$, but we now interpret x, y and z as the activity levels of three neurons connected by both the linear and quadratic terms. With these parameters there exist three unstable fixed points of the dynamics, namely at $(x, y, z) = (0, 0, 0)$, and $(\pm\sqrt{b(r-1)}, \pm\sqrt{b(r-1)}, r-1)$. These are sometimes labelled Q, P_1 and P_2, respectively. Solution trajectories are bounded but there are no stable fixed points or periodic attractors. It is believed that the trajectories approach a 'strange attractor' (see e.g. [62]). The definition of chaotic dynamics and the question of whether or not the Lorenz system really displays such behaviour are not critical for our purposes. We need only be clear that typical trajectories wander over large regions of phase space and that the fixed points are unstable.

The important observation of [24] is that when a sufficiently strong input is added to the y equation, then one of the two equilibrium points, P_1 or P_2 (moved slightly), becomes stable and the trajectories spiral in toward it. Which of these two equilibria becomes stable depends on the sign of the input. (Although Evans, *et al.* only deal with input to the y equation, the same effect occurs in numerical experiments when input is added to the x equation). Using the y equation for input

the system becomes

$$\begin{aligned} \dot{x} &= -\sigma(x-y) \\ \dot{y} &= -y + rx - xz + \alpha\chi_{[t_1,t_2]} \\ \dot{z} &= -bz + xy\,. \end{aligned} \tag{4.2}$$

The system effectively distinguishes between positive and negative inputs (of sufficient magnitude).

The properties of this system of equations that allow it to perform this simple information-processing task are as follows:

1. Solutions are bounded. The symmetries in the quadratic terms ensure this even though there is no saturating response function like the sigmoid in the Hopfield network. Boundedness is proved by showing that for $u = x^2 + y^2 + [z - (r+\sigma)]^2$, there exist constants $c_1, c_2 > 0$ such that $\dot{u} \leq -c_1 u + c_2$.

2. All fixed points (or periodic orbits) are unstable, ensuring complex behaviour.

3. It has two unstable fixed points aside from the origin which allow for some choice as to which is to become stable, and therefore allow for the ability to respond in more than one way to input.

4.3 Attempts to extend the Lorenz equations.

The information-processing ability of the Lorenz system with input considered as a simple network of three neurons suggests that it might be possible to extend such capabilities by working with larger (higher dimensional) chaotic systems. We would want to retain the three properties of the Lorenz system listed above, but to obtain systems with larger numbers of fixed points to allow for more complex responses to inputs.

It is not straightforward, however, to add equations to the Lorenz system in such a way as to retain these features. *Ad hoc* attempts to do this typically make the technique for proving boundedness fail or do not add extra fixed points. Ensuring instability makes it even more difficult.

The easiest way to extend this idea into higher dimensions is to make use of what we know about the Lorenz system by considering it as a processing unit and then connecting many such units together. An uncoupled pair of Lorenz systems can be considered as a six-dimensional dynamical system with bounded solutions and nine unstable fixed points. Inputs to the y component of each system can be used to make one of four of these fixed points stable. Of course, any number of Lorenz systems could be considered together and the number of stabilizable fixed points grows rapidly, but as long as the systems are uncoupled, they are essentially just displaying binary information. An input simply causes the entire system to converge to a fixed point which reflects that input.

For more complex behaviour, we should couple the Lorenz units, making a real network. However, this changes the dynamics and analysis becomes difficult again. To show the idea, suppose we take the y component of one unit (x_2, y_2, z_2) and feed it into the y equation of another unit (x_1, y_1, z_1):

$$\begin{aligned}
\dot{x}_1 &= -\sigma(x_1 - y_1) \\
\dot{y}_1 &= -y_1 + r x_1 - x_1 z_1 + \alpha y_2 \\
\dot{z}_1 &= -b z_1 + x_1 y_1 \\
\dot{x}_2 &= -\sigma(x_2 - y_2) \\
\dot{y}_2 &= -y_2 + r x_2 - x_2 z_2 \\
\dot{z}_2 &= -b z_2 + x_2 y_2 \,,
\end{aligned} \tag{4.4}$$

where the parameters are all the same as before (equation 4.1). If α is small, then the two systems operate almost as before. We know from [24] that small inputs

to the y component do not disrupt the chaotic behaviour of the system. If α is larger, however, then the input depends on the sign of y_2. The second Lorenz unit is unaffected by the first, so y_2 oscillates chaotically between positive and negative values. When y_2 is positive, the first Lorenz unit is driven to one of its two non-trivial equilibria (the positive one if $\alpha > 0$, the negative one if $\alpha < 0$), and when y_2 is negative, the first Lorenz unit is driven to the other equilibrium. Thus, the second unit controls the first but the first still behaves chaotically because the second does. Now if an external input to the second unit causes it to stabilize at one or the other equilibrium, it will force the first to stabilize also, and the entire system will approach equilibrium. However, an external input of sufficient strength to the first unit could override the forcing from the second unit.

Proposition 4.1 *The system of equations* (4.4) *has bounded solutions if* $\alpha \leq 2$.

Proof Let

$$u = x_1^2 + y_1^2 + [z_1 - (\sigma + r)]^2 + x_2^2 + y_2^2 + [z_2 - (\sigma + r)]^2 ,$$

then

$$\dot{u} = -2\sigma x_1^2 - 2y_1^2 - 2bz_1^2 + 2bz_1(r + \sigma) + 2\alpha y_1 y_2$$

$$-2\sigma x_2^2 - 2y_2^2 - 2bz_2^2 + 2bz_2(r + \sigma) .$$

Using the fact that

$$-2bz^2 + 2bz(r + \sigma) \leq -\frac{b}{2}[z - (r + \sigma)]^2 + \frac{2}{3}b(r + \sigma)^2 ,$$

we get

$$\dot{u} \leq -2\sigma(x_1^2 + x_2^2) - 2(y_1^2 + y_2^2) - \frac{b}{2}\left([z_1 - (r + \sigma)]^2 + [z_2 - (r + \sigma)]^2\right)$$

$$+\frac{4}{3}b(r + \sigma)^2 + 2\alpha y_1 y_2 .$$

Now, noting that $2y_1y_2 \leq y_1^2 + y_2^2$, we have that

$$-2(y_1^2 + y_2^2) + 2\alpha y_1 y_2 \leq (\alpha - 2)(y_1^2 + y_2^2)$$

and for $\alpha \leq 2$, we can find positive constants c_1 and c_2 such that

$$\dot{u} \leq -c_1 u + c_2 .$$

□

Thus, we have found a 6-dimensional system with bounded solutions and nine fixed points, including the trivial one. We have guaranteed boundedness only in the case of negative or weakly positive connections, α, whereas Evans, *et al.* use large positive and negative inputs to stabilize the equilibria. We get a more generous bound if we use the same technique applying the connection to the x_1 variable instead under the condition $\alpha \leq 2\sigma$. Also, we could apply the connections in both directions, i.e. include an x_1 term on the x_2 equation and an x_2 term on the x_1 equation, as long as the sum of the connection strengths was less than 2σ. This is easily generalizable to the coupling of any number of Lorenz units, but we are sure of the boundedness only when the sum of connection strengths is ≤ 2 (for y connections) or $\leq 2\sigma$ (for x connections).

All the fixed points are retained for fairly sizable connection strengths α on the y equation, for example. This can be seen as follows. The second, undisturbed, Lorenz system has the usual three fixed points. In one of them the y value is 0, and the first system is also undisturbed at equilibrium, so the usual three exist in this case. In the other two equilibria for the second system, $y = \pm\sqrt{b(r-1)}$. Then at equilibrium for the first system, $x_1 = y_1$, $z_1 = \frac{1}{b}x_1^3$, and

$$(r-1)x_1 - \frac{1}{b}x_1^3 \pm \alpha\sqrt{b(r-1)} = 0 .$$

Letting $f(x) = (r-1)x - \frac{1}{b}x^3$, it is easy to see that the maximum vertical shift it can be given without losing two of its roots is the magnitude of its local extrema, which is $\frac{2}{3}(r-1)^{\frac{3}{2}}\sqrt{\frac{b}{3}}$. Thus, we retain the three roots as long as

$$|\alpha|\sqrt{b(r-1)} \leq \frac{2}{3}(r-1)^{\frac{3}{2}}\sqrt{\frac{b}{3}}$$

or $|\alpha| \leq \frac{2}{3\sqrt{3}}(r-1)$. For the usual parameter values (see equation 4.1), this is about 10.39.

Some numerical simulations of equations (4.4) are given in Figs. 4.1a–d, where solution trajectories for the first Lorenz system (x_1, y_1, z_1) are displayed on the left and for the second on the right. These are projections onto the $x = y$ plane in each case and show the region of phase space occupied by the Lorenz attractor [62]. Time is marked on the trajectories for both systems (which evolve together) at significant moments. We have used values of the connection parameter α larger than 2 despite the lack of proof of boundedness in this case. This did not appear to be a problem for the parameters chosen.

It is difficult to derive general conditions under which such systems with many or larger connection parameters have bounded solutions. The boundedness must come from the special structure of the quadratic terms. Another way to guarantee boundedness of solutions, however, is to put the activities of neighbouring neurons through a sigmoid function before summing them, as in the Hopfield network. With a leakage term and a bound on the size of the effects from other neurons, boundedness is automatic. We know that it is also possible for Hopfield nets to be chaotic, so if we could be sure that we had chaotic activity in a Hopfield net, then we would still need to ensure instability of equilibria, sufficient numbers of equilibria and the possibility of stabilizing them with input, to carry out the program suggested above for the Lorenz-type systems, but without any worry of losing boundedness. One method of doing this is given in the next section.

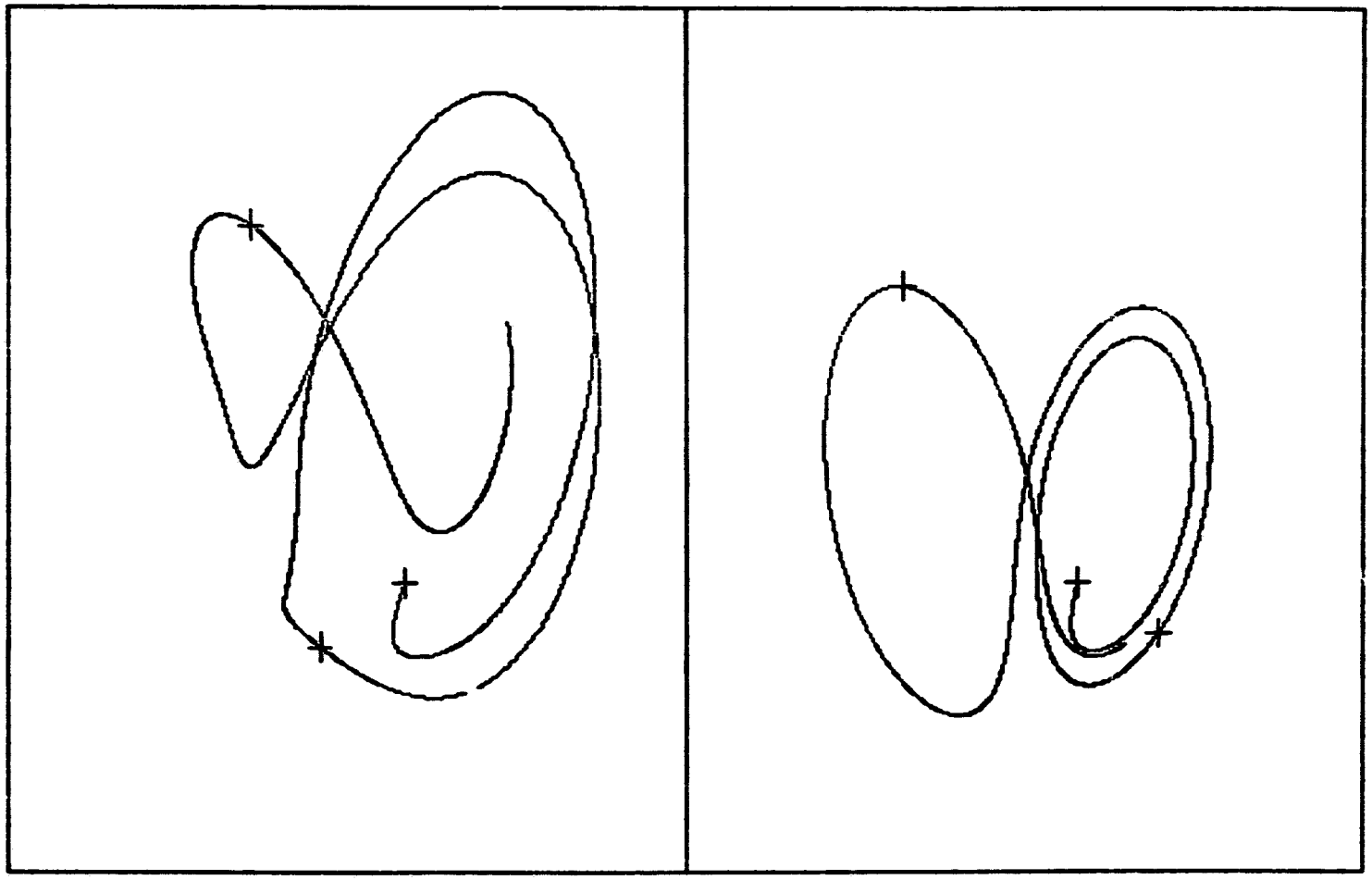

Figure 4.1a Trajectories of coupled Lorenz systems (4.4) with $\alpha = 9$. The second system evolves undisturbed; the first is strongly pulled to the same side (of the $x = y$ plane) as the second. Crosses mark times $t = 0, 1$ and 2.

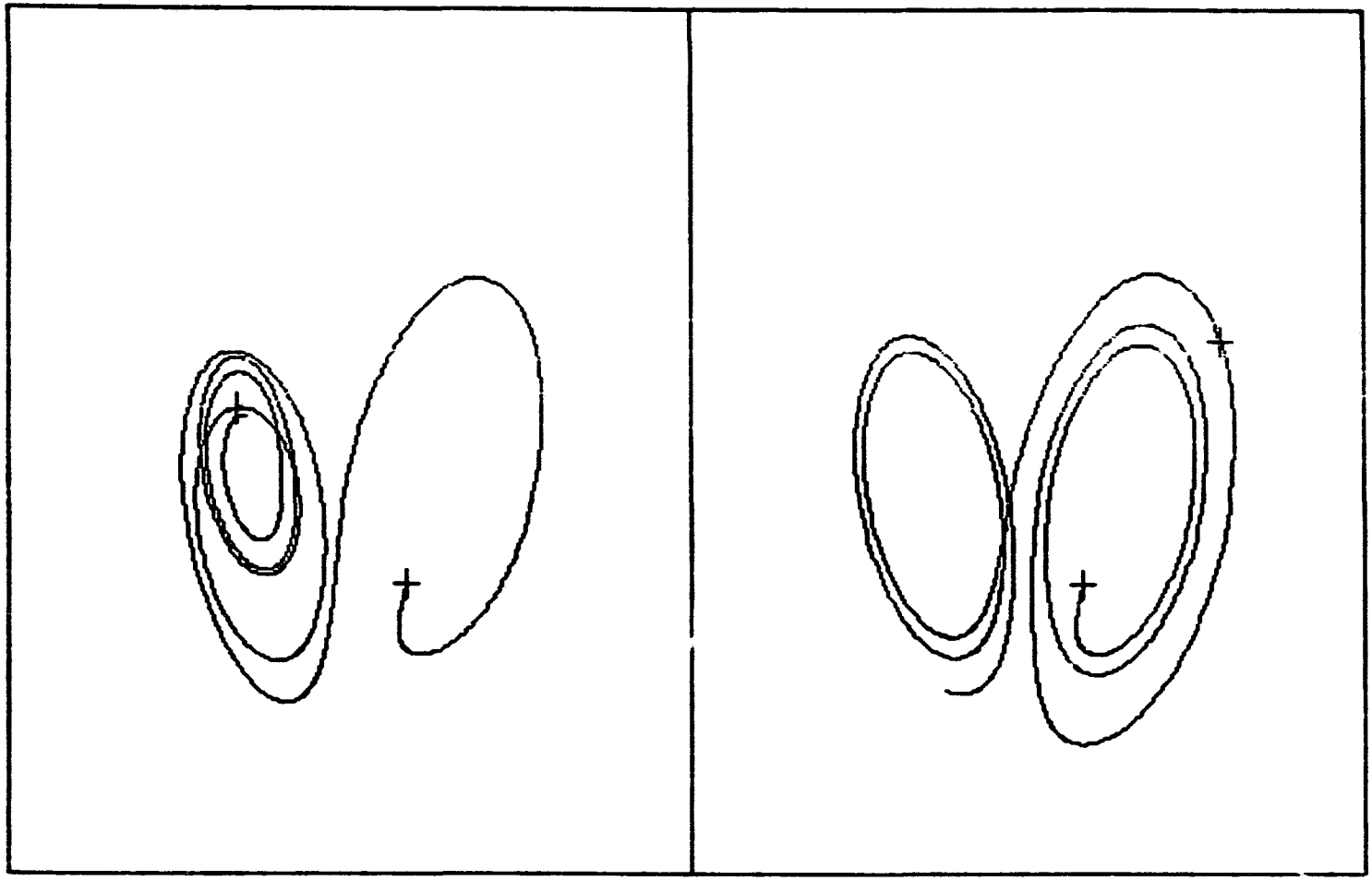

Figure 4.1b Trajectories of coupled Lorenz systems (4.4) with $\alpha = 2$. While the second system spirals outwards on the left side of the plane, the first spirals inwards, suggesting temporary 'stability.' Crosses mark times $t = 0$ and 2.

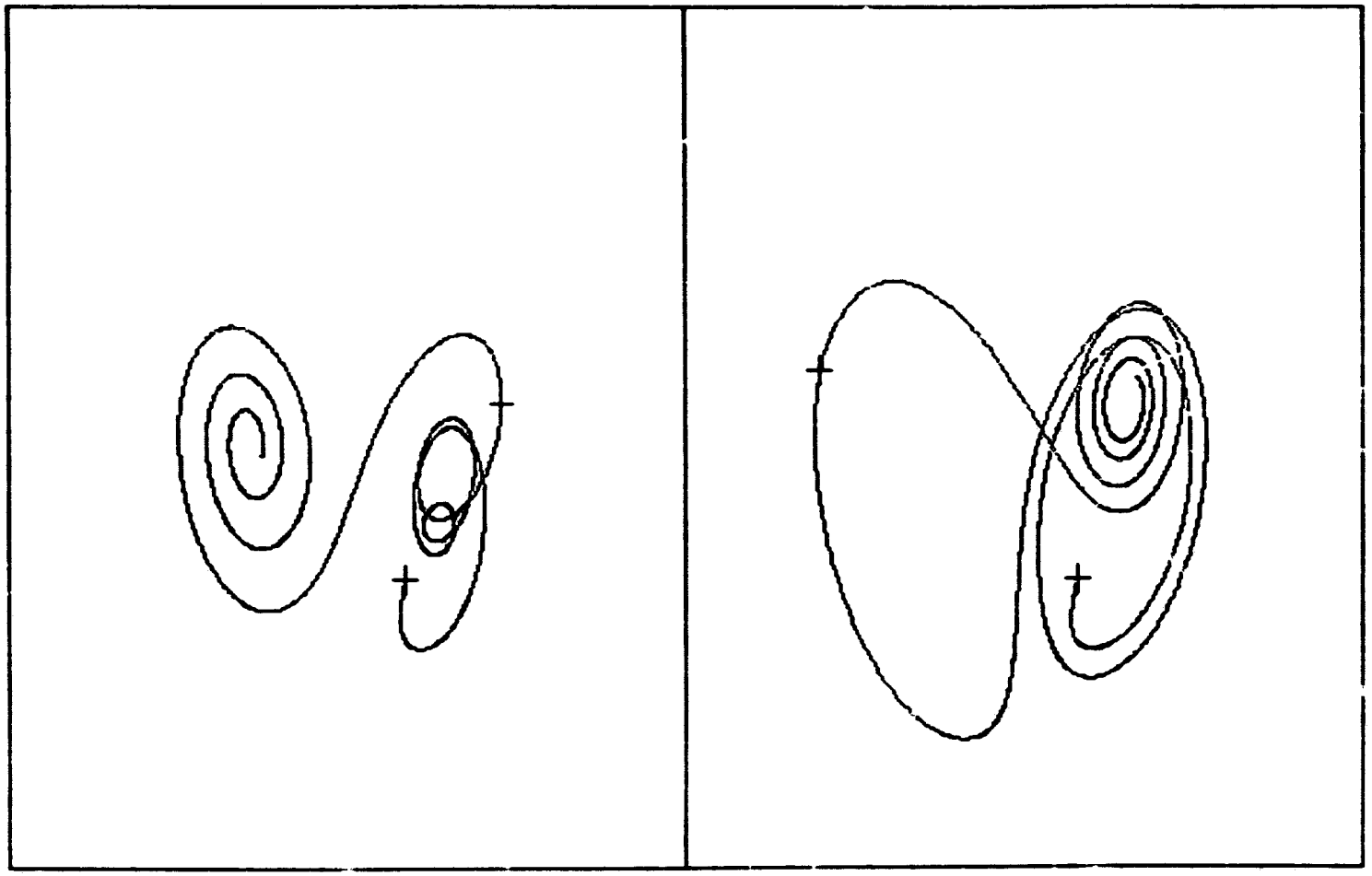

Figure 4.1c Trajectories of coupled Lorenz systems (4.4) with $\alpha = -2.5$ and external input of strength 50 applied to the y_2 equation from $t = 2$ to $t = 4$. The input causes the second system to converge to its now stable fixed point on the right, and the consistent (negative) forcing by the second system causes the first system to converge on the left. Crosses mark times $t = 0$ and 2.

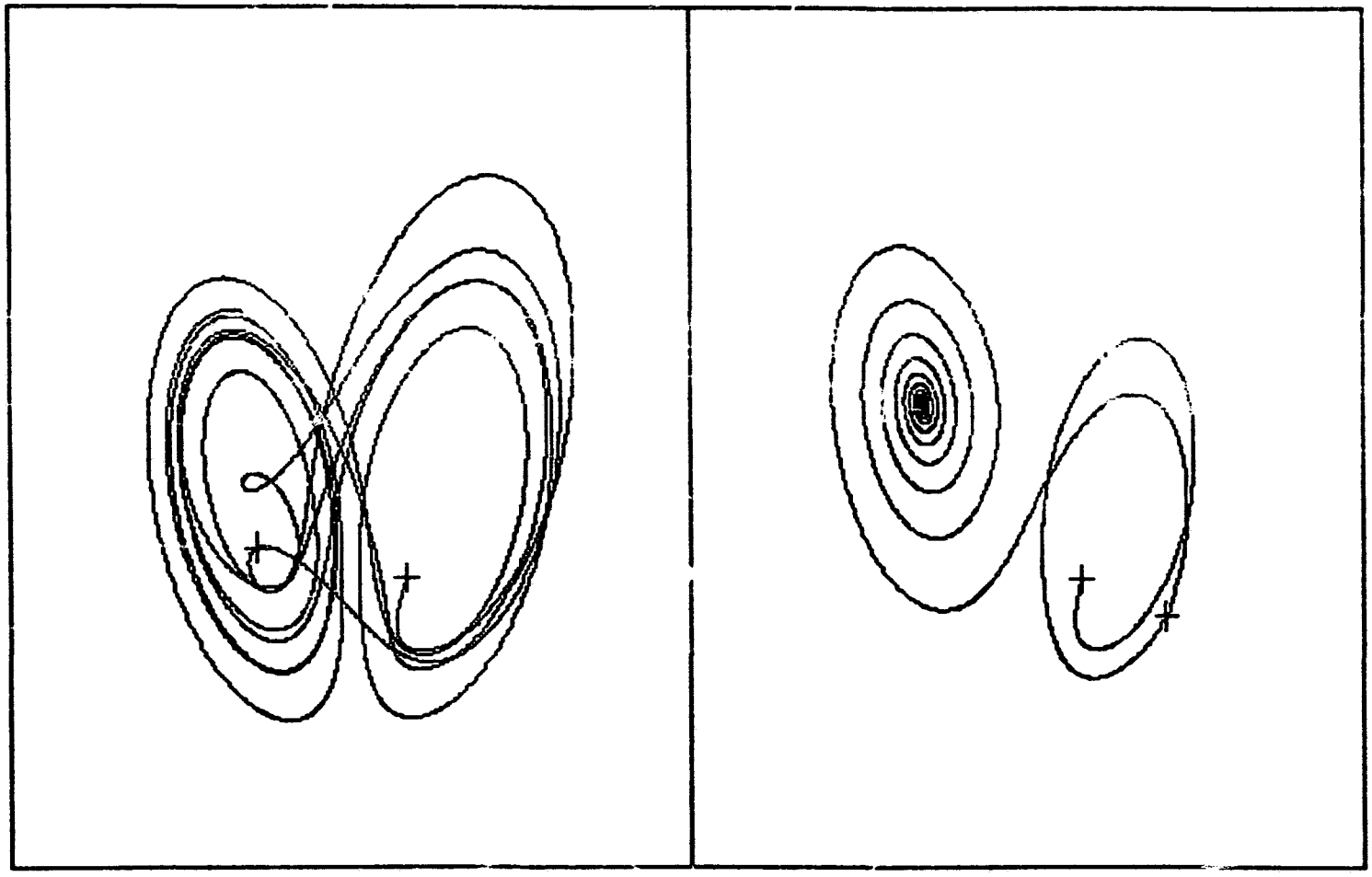

Figure 4.1d Trajectories of coupled Lorenz systems (4.4) with $\alpha = 5.5$ and external input of strengths 50 and -50 applied to the y_1 and y_2 equations, respectively, from $t = 1$ to $t = 9$. The negative input causes the second system to converge on the left. The first system then receives negative forcing from the second system counteracted by positive forcing from its own input, so it remains chaotic. Crosses mark times $t = 0$ and 1.

4.4 Kwan's model and a new chaotic Hopfield network.

An alternate model for chaotic neural networks has been suggested by Kwan [42] and Priesol, *et al.* [51]. They propose a unit composed of four neurons, that has the potential of acting chaotically, by itself or in conjunction with other such units (see Fig. 4.2).

The model operates in discrete time. Each unit receives input from itself, from other units and from external signals to its 'input' neuron. The input neuron has a linear response, so that the membrane potential accumulated is essentially the firing rate produced. Its activity is denoted by $X_i(t)$ and is given by

$$X_i(t) = \sum_j w_{ij} a_j(t) + I_i, \tag{4.5}$$

where a_j is the output of a connected unit modulated by the synaptic weighting, w_{ij}, and I_i is the external input to the i^{th} unit. This activity is then fed into an excitatory and an inhibitory neuron, each with sigmoid responses (logistic in this case) and a threshold term. These two neurons receive no other input, so their firing rates are given by $g(X_i(t) - \theta_E)$ and $g(X_i(t) - \theta_I)$ for the excitatory and inhibitory neuron respectively, where

$$g(y) = \frac{1}{1 + e^{-y/s}},$$

s a constant. These two signals are then passed into the linear 'output' neuron of the unit, weighted by different amounts, so the activity of the output unit is

$$a_i(t+1) = w_E g(X_i(t) - \theta_E) - w_I g(X_i(t) - \theta_I). \tag{4.6}$$

This is then passed on to the other units for the next time step. All weighting parameters, w, in the model are taken to be positive.

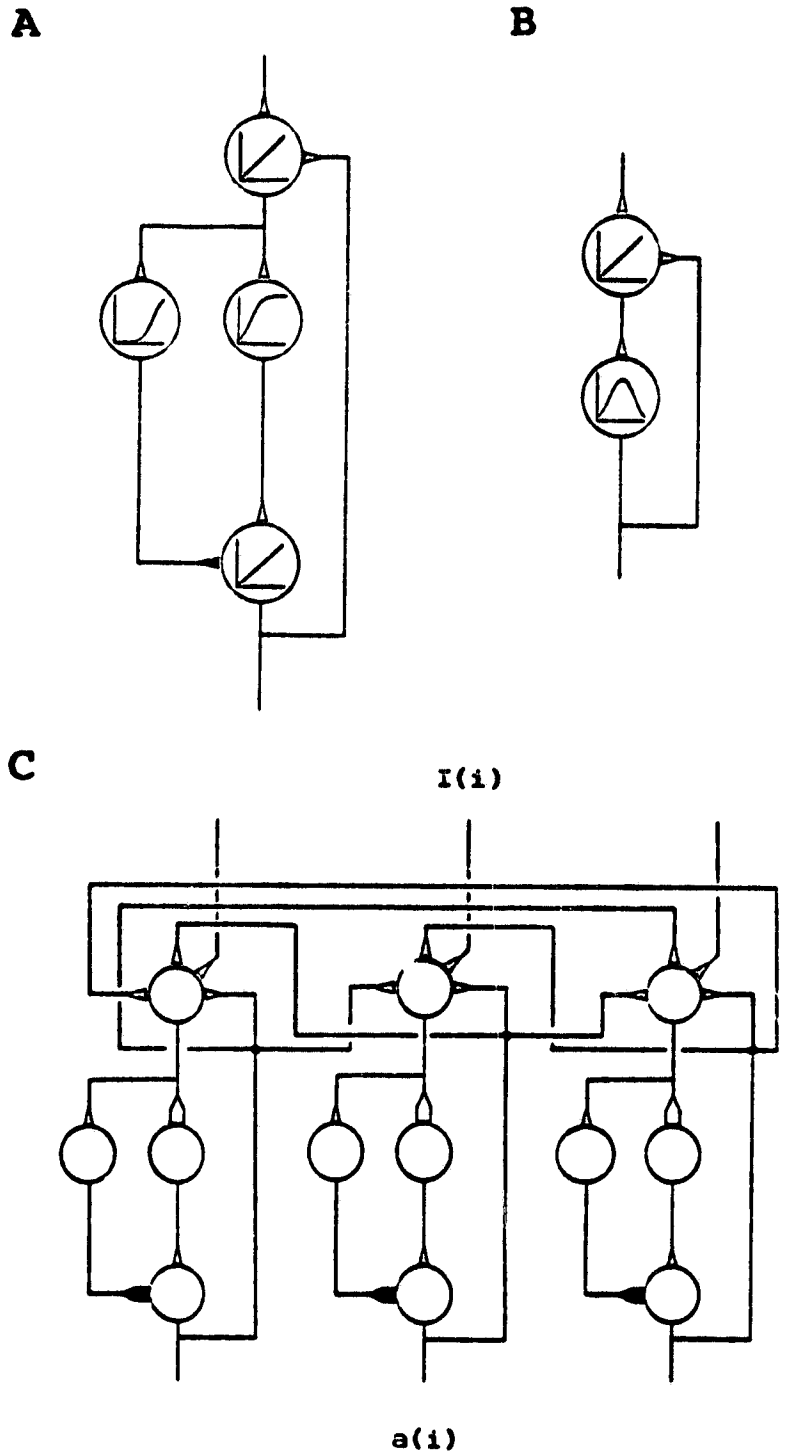

Figure 4.2 The 'Kwan' model. A. A unit of 4 neurons, two linear, and an inhibitory–excitatory pair. B. An equivalent conceptualization of a unit with a 'hump'-shaped response, the combination of the inhibitory and excitatory neurons' effects. C. A system of connected units with external inputs. Dark connection symbols indicate inhibition. (From Preisol, *et al.*, 1991, with permission).

Priesol, *et al.* [51] demonstrate by numerical experiment that a network of this type, consisting of three units and with appropriate choice of parameters, behaves chaotically for some sets of inputs and not for others. They do not give an analysis of why this should be so and under what conditions, but they point out that the difference of the two logistic functions in (4.6) gives a 'hump' function similiar to the quadratic in the simple and well known dynamical system:

$$x(t+1) = rx(t)(1 - x(t)),$$

at least with appropriate choices for the parameters w_E, w_I, θ_E, θ_I, and s. In particular, we should take $\theta_E < \theta_I$. Here we analyze the model a little further and show that it can at least reproduce the result of the Lorenz system in Section 4.2 and can be extended to a system of many interconnecting units without the problems of boundedness encountered before.

First we observe that the dynamics of their model are equivalent to the dynamics of a discrete time Hopfield network (equation 2.8) with a particular connection structure. The linear response 'input' and 'output' neurons in each unit simply transfer the signals they accumulate directly to the neurons they influence. Therefore, the activity of the network is essentially unchanged if we consider these neurons to be absent and channel their incoming signals directly to the neurons they influence. Then we need only keep track of the activity of the inhibitory and excitatory neurons within each unit and a unit now consists of just these two neurons. We denote the activity of these neurons by $x_k(t)$, where odd k refer to the excitatory neurons and even k refer to the inhibitory neurons. A unit, then, is an odd-even pair, so that neuron k belongs to the i^{th} unit where $i = \lceil \frac{k+1}{2} \rceil$ (the greatest integer $\leq \frac{k+1}{2}$). The system becomes

$$x_k(t+1) = g\left[\sum_l T_{kl}x_l(t) + J_k - \theta_k\right], \tag{4.7}$$

where

$$T_{kl} = \begin{cases} w_{ij}w_E \text{ for } l \text{ odd} \\ -w_{ij}w_I \text{ for } l \text{ even} \end{cases}, \text{ with } i = \left\lceil \frac{k+1}{2} \right\rceil, j = \left\lceil \frac{l+1}{2} \right\rceil$$

$$J_k = I_i, \text{ with } i = \left\lceil \frac{k+1}{2} \right\rceil,$$

$$\theta_k = \begin{cases} \theta_E \text{ for } k \text{ odd} \\ \theta_I \text{ for } k \text{ even.} \end{cases}$$

This is a standard Hopfield network equation with a particular structure. Inputs to the two neurons of a unit are the same, $J_{2i-1} = J_{2i}, i = 1, 2, ...N$. Thresholds have one of two values, alternating between θ_E and θ_I, both being positive. The connection matrix has alternating columns of positive and negative entries composed as in this example:

$$T = \begin{bmatrix} w_{11}w_E & -w_{11}w_I & w_{12}w_E & -w_{12}w_I & w_{13}w_E & -w_{13}w_I \\ w_{11}w_E & -w_{11}w_I & w_{12}w_E & -w_{12}w_I & w_{13}w_E & -w_{13}w_I \\ w_{21}w_E & -w_{21}w_I & w_{22}w_E & -w_{22}w_I & w_{23}w_E & -w_{23}w_I \\ w_{21}w_E & -w_{21}w_I & w_{22}w_E & -w_{22}w_I & w_{23}w_E & -w_{23}w_I \\ w_{31}w_E & -w_{31}w_I & w_{32}w_E & -w_{32}w_I & w_{33}w_E & -w_{33}w_I \\ w_{31}w_E & -w_{31}w_I & w_{32}w_E & -w_{32}w_I & w_{33}w_E & -w_{33}w_I \end{bmatrix}.$$

We could also write system (4.7) in terms of membrane potentials as

$$y_k(t+1) = \sum_l T_{kl} g(y_l(t)) + J_k - \theta_k. \tag{4.8}$$

Yet another formulation is to consider each unit as a 'neuron' with a hump-shaped response function, rather than a sigmoid. Then $X_i(t)$ represents the membrane potential of the 'neuron', $a_i(t)$ represents the firing rate of the 'neuron', and we have either

$$X_i(t+1) = \sum_j w_{ij} h(X_j(t)) + I_i \tag{4.9}$$

or

$$a_i(t+1) = h\left[\sum_j w_{ij} a_j(t) + I_i\right], \tag{4.10}$$

where

$$h(y) = w_E g(y - \theta_E) - w_I g(y - \theta_I),$$

the hump function.

Consider first a single unit, using equation (4.7):

$$x_1(t+1) = g\left[w_{11}\left(w_E x_1(t) - w_I x_2(t)\right) + I_1 - \theta_E\right],$$

$$x_2(t+1) = g\left[w_{11}\left(w_E x_1(t) - w_I x_2(t)\right) + I_1 - \theta_I\right],$$

where $w_{11}\left(w_E x_1(t) - w_I x_2(t)\right) + I_1 = X_1(t+1)$ from (4.9). Note that the input may be considered to modify the threshold. That is, when input I_1 is active, the thresholds effectively become $(\theta_E - I_1)$ and $(\theta_I - I_1)$.

This suggests defining a new variable,

$$Z_i(t+1) = \sum_j w_{ij} h\left(X_j(t)\right) = X_i(t+1) - I_i,$$

which is the sum of all the inputs from other units, not including the external input signal, then equation (4.9) solely in terms of Z_i becomes

$$\begin{aligned} Z_i(t+1) &= \sum_j w_{ij} h\left(Z_j(t) + I_j\right) \\ &= \sum_j w_{ij}\left(w_E g\left(Z_j(t) - (\theta_E - I_j)\right) - w_I g\left(Z_j(t) - (\theta_I - I_j)\right)\right) \end{aligned} \tag{4.11}$$

which is the same as the original system but with thresholds moved to the left by I_i for each unit i. In other words, the Z_i variables respond to a hump function shifted to the left by I_i.

In terms of this variable, our single unit with feedback on itself follows the mapping

$$Z_1(t+1) = w_{11} h\left(Z_1(t) + I_1\right). \tag{4.12}$$

It is easy to make h very similar to the quadratic map, by taking $s < \frac{1}{4}$, $\theta_E < \theta_I$, and $w_E = w_I$ for example (see Fig. 4.3).

If the hump has a steep enough downward face and is positioned correctly, the behaviour of this mapping will be chaotic. In particular, if there is an equilibrium point, $Z = w_{11}h(Z)$, such that $h'(Z) < -1$, then it is unstable and we expect at least periodic solutions. Like for the quadratic map, as $h'(Z)$ gets more negative, we expect a period doubling cascade and eventually chaotic dynamics.

Now the effect of adding an external input becomes clear. As the hump function is moved to the left by a positive input, the equilibrium point slides down into the right-hand tail of h, $|h'(Z)|$ becomes < 1 and therefore the equilibrium becomes stable. Chaotic activity is lost. This stable state corresponds to a situation where both the excitatory and inhibitory neurons are firing strongly. When a negative input is applied, the hump moves to the right and first the equilibrium point moves up on to the top of the hump, becoming stable (along with another stable equilibrium on the left-hand tail), and finally, with stronger negative input, only the stable equilibrium on the left-hand tail remains. This equilibrium corresponds to low activity in both the excitatory and inhibitory neurons. The intermediate state, where there is a stable equilibrium at the top of the hump corresponds to high activity in the excitatory neuron and low activity in the inhibitory neuron. (See Figs. 4.4a-c).

This makes it clear how inputs affect the dynamics. If the thresholds are such that the system is chaotic when no inputs are present, then addition of sufficiently strong inputs can stabilize the system in at least three different ways. Thus, our single unit model (4.12) has the same capabilities as the Lorenz system with input. We can now create more complex networks by interconnecting these units, and unlike the interconnected Lorenz systems, we can be sure that solutions are bounded.

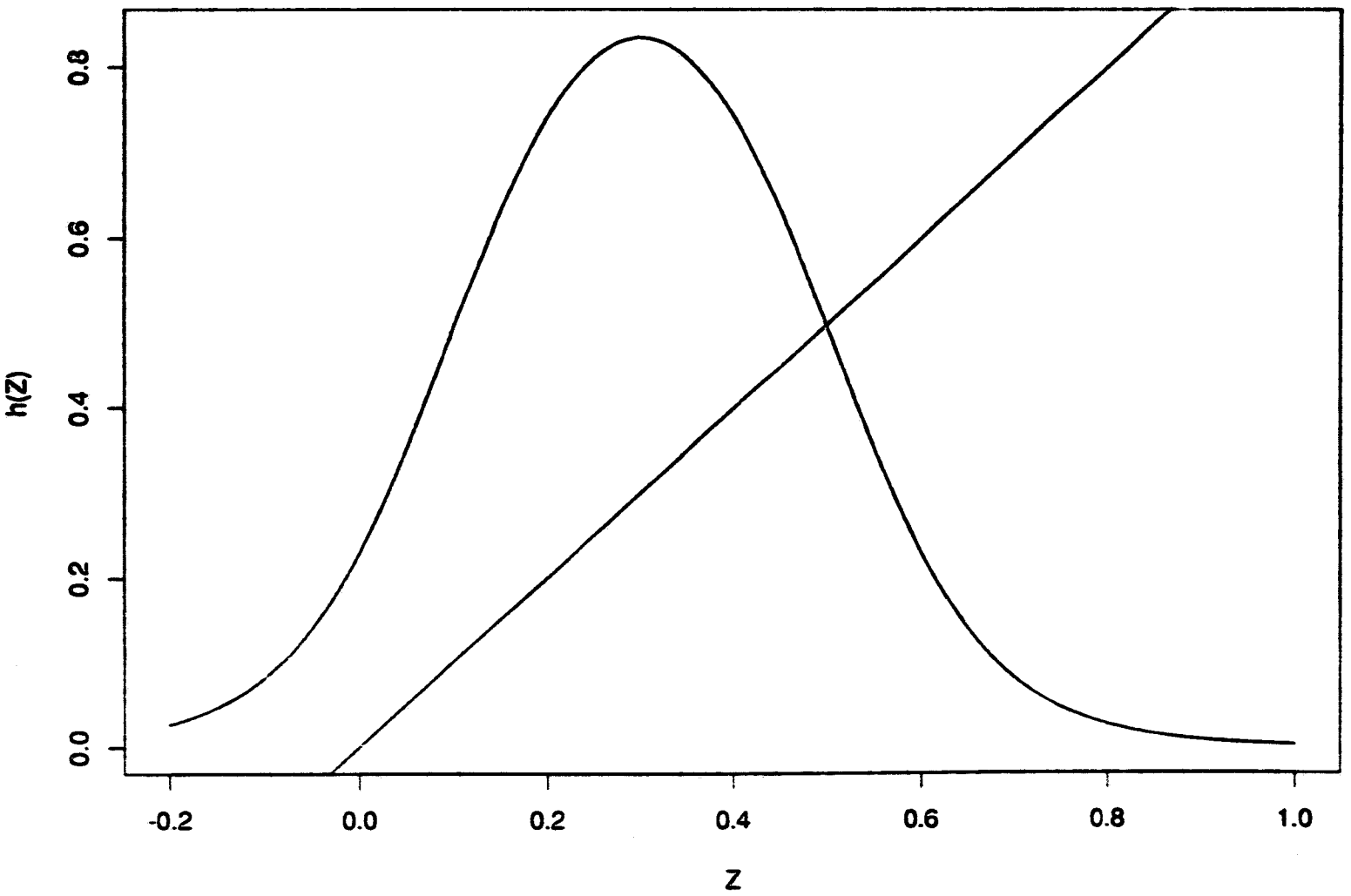

Figure 4.3 A 'hump' function: the difference of two logistic functions as in equation (4.12), with $w_E = w_I = 1$, $\theta_E = 0.1$, $\theta_I = 0.5$ and $s = \frac{1}{12}$. The straight line is $f(Z) = Z$.

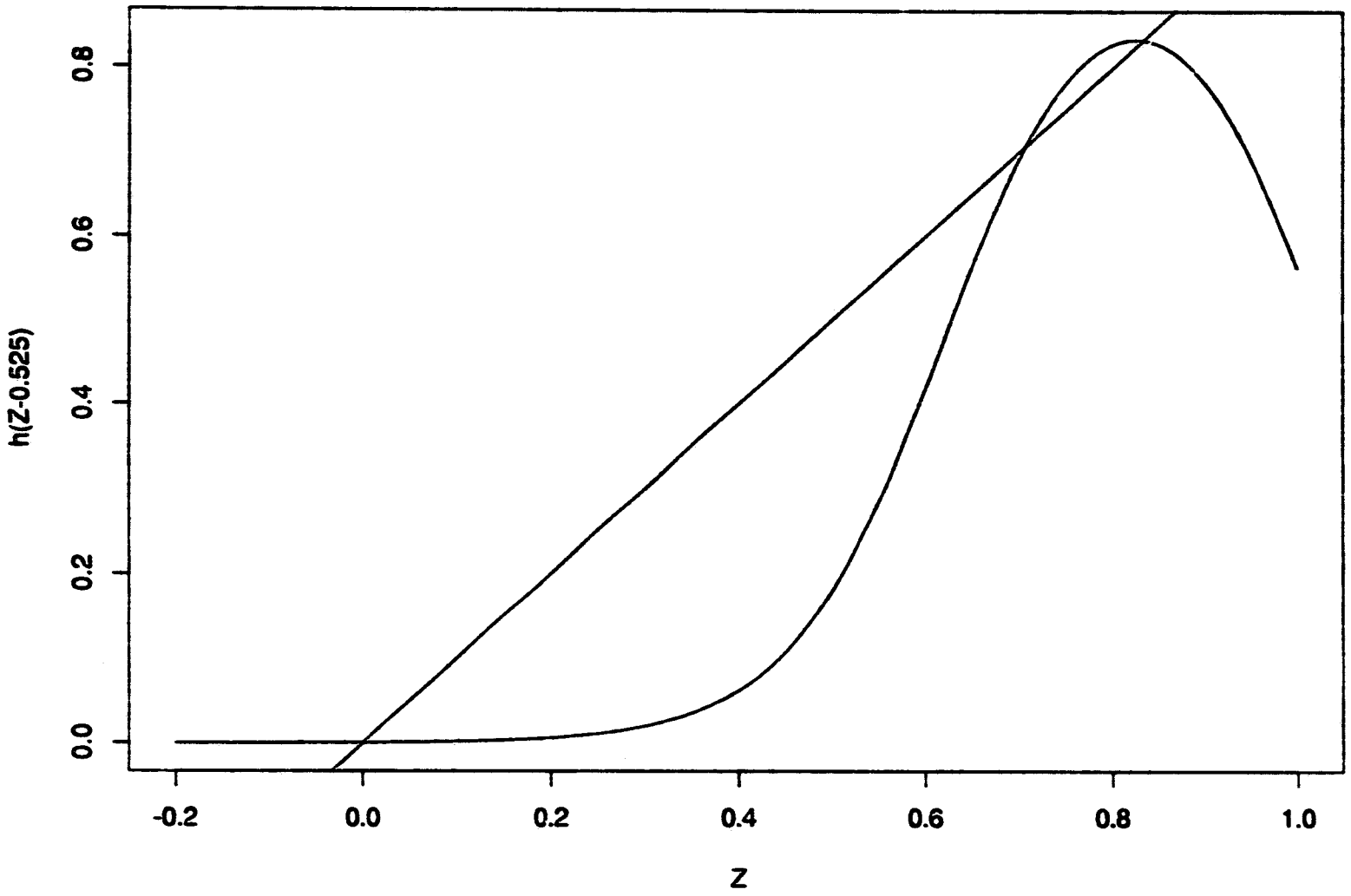

Figure 4.4a The input-shifted hump function of equation (4.12) with h as in Fig.4.3 and $I = -0.525$.

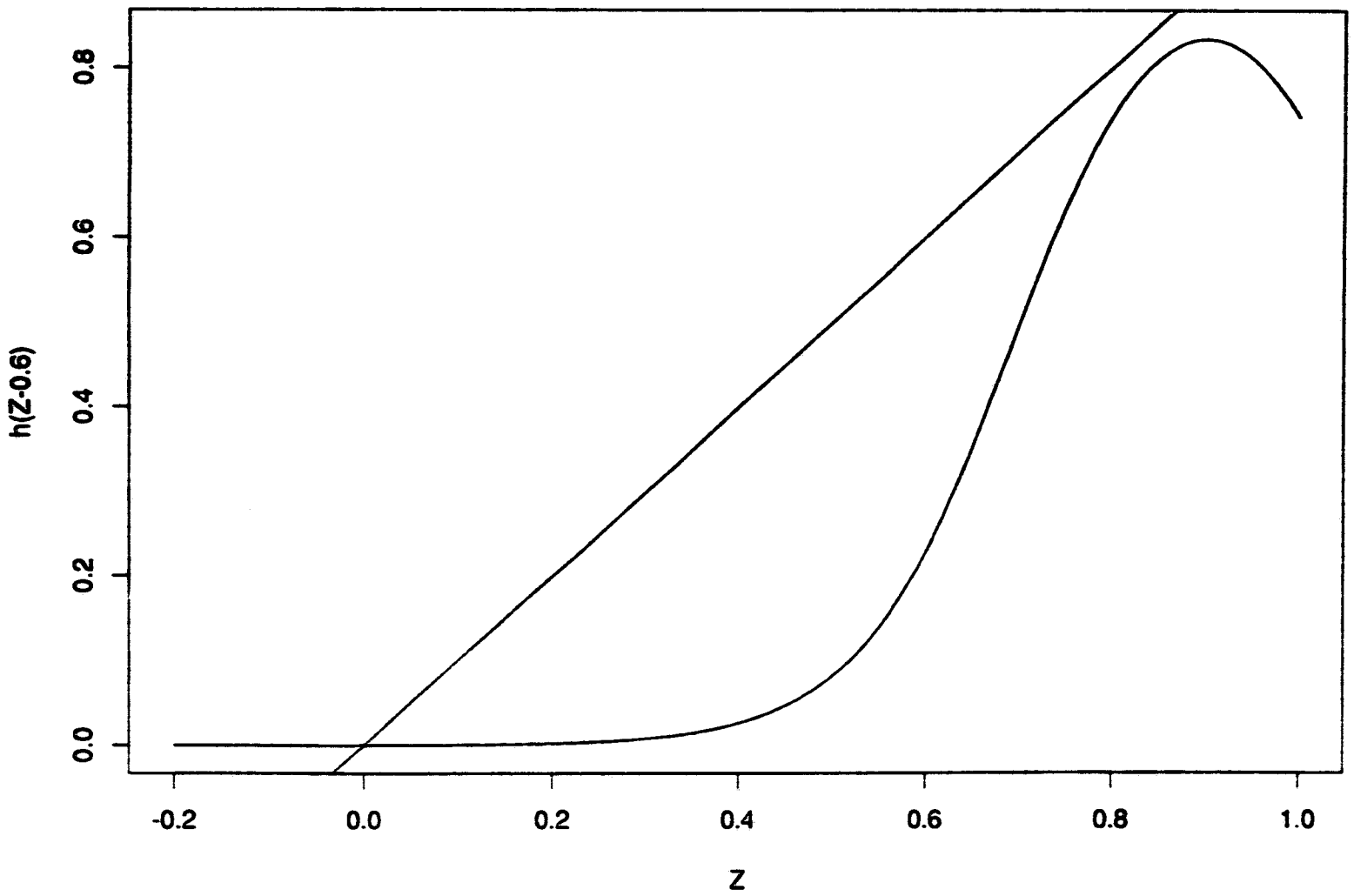

Figure 4.4b The input-shifted hump function of equation (4.12) with h as in Fig.4.3 and $I = -0.6$.

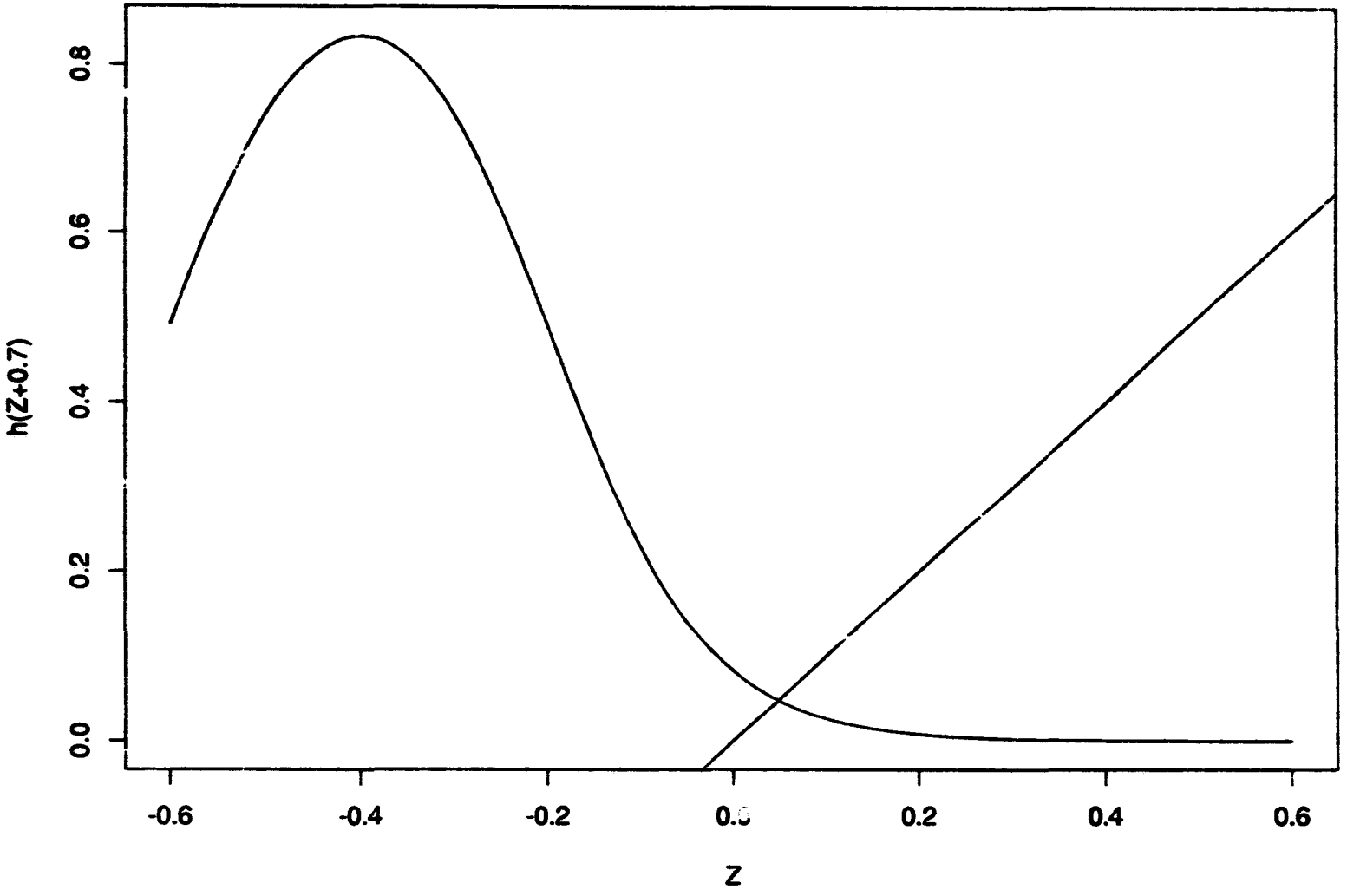

Figure 4.4c The input-shifted hump function of equation (4.12) with h as in Fig.4.3 and $I = 0.7$.

A profusion of potential fixed points is very likely for a large number of interconnected units and under conditions of weak or no inputs, many of these should be unstable. Input can clearly stabilize the dynamics, even for a large system. All units that receive a strong enough input will stabilize and other units that do not receive external inputs may still stabilize due to the collective signal they receive from the stabilized neurons.

It is not clear, however, that the dynamics of a large system in the absence of input must be chaotic. For instance, in the model described above, with $w_I = w_E$ and $\theta_E < \theta_I$, the output of each unit (the value of the hump function) is always positive. Thus, the collective input from other units will be positive, tending to shift everything to the stable right-hand tail. Another possibility is to set $w_I > w_E$. Then the right-hand tail of the hump function goes negative and the output of the unit can be negative as well as positive. In any case, the numerical simulations of Priesol, *et al.* [51] show that chaotic activity does occur in a 3-unit system for some choices of parameters.

Another possible modification would be to allow separate inputs to the excitatory and inhibitory neurons. In (4.7), J_k would be potentially a different number for each k, rather than being same for each pair of neurons. In (4.11), a different I_j would be subtracted from θ_E than from θ_I. The effect of allowing separate inputs is then to shift one side of the hump function independently of the other. For example, a negative input to the inhibitory neuron would cause the hump function to stretch out to the right (Fig. 4.5). A positive input to the inhibitory neuron would cause the hump to contract on the right, and if it was shifted enough, the inhibitory threshold would then be less than the excitatory threshold and the hump would become negative.

One example of a system of many units where we can get more information

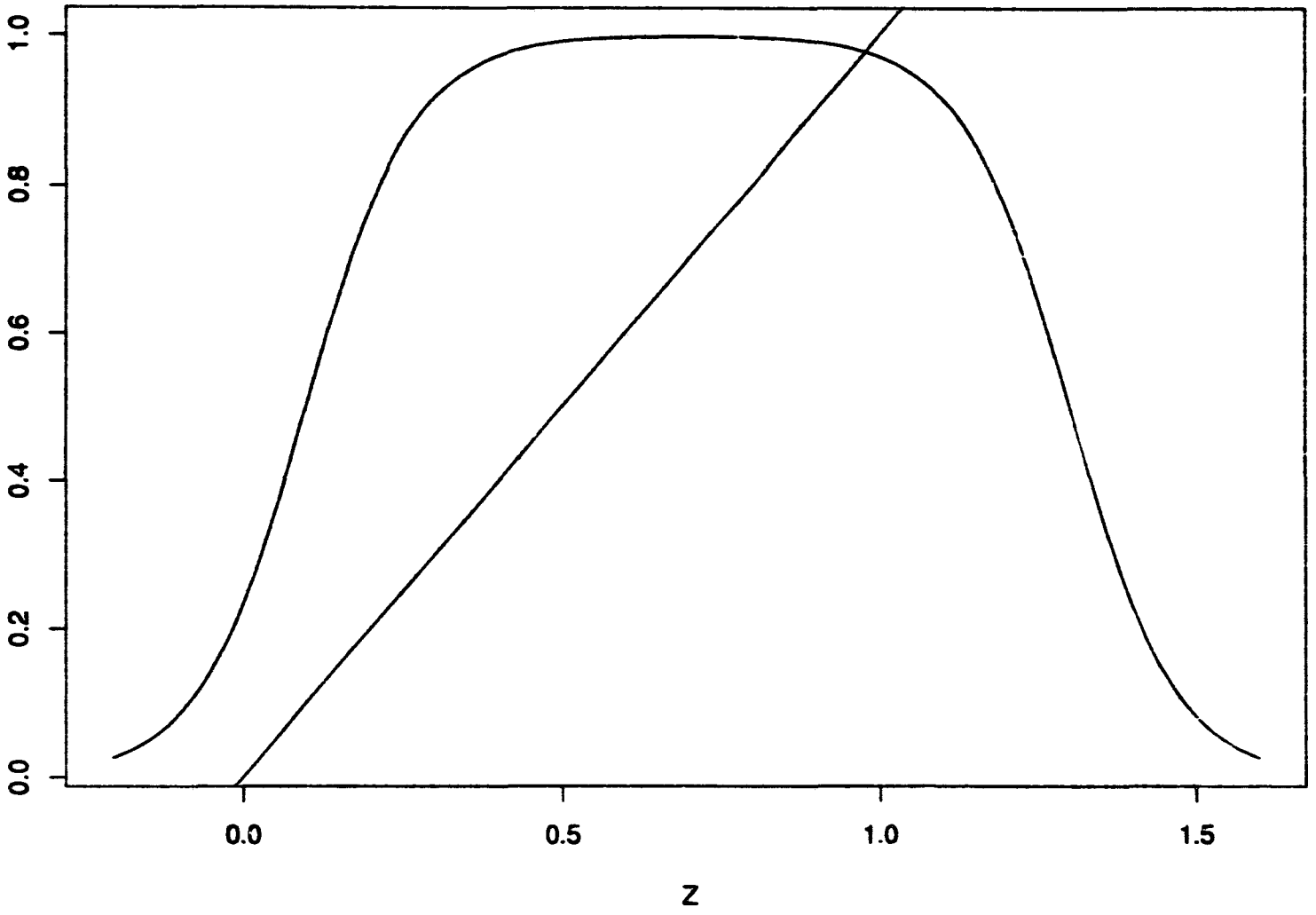

Figure 4.5 The hump function of Fig. 4.3 with input $I = -0.8$ applied only to the inhibitory neuron.

easily is the case where all rows of T are identical, i.e. w_{ij} is the same for all i. Then the only equilibria for equation (4.11) without external input are given by

$$Z_i = \sum_j w_{ij} h(Z_j) = k\,,$$

where k is constant. Stability of equilibria can be determined by the eigenvalues of $D(W\mathbf{h})$ at equilibrium, where W is the matrix of weights and $\mathbf{h}$ is the vector with $h_i = h(Z_i)$. But

$$D(W\mathbf{h}) = Wh'(k)$$

so the eigenvalues of $D(W\mathbf{h})$ at equilibrium are $h'(k)$ times the eigenvalues of W. Since W has identical rows, its eigenvalues are all 0 except one which is a row sum: $\sum_j w_{ij}$. Thus, if

$$h'(k) \sum_j w_{ij} < -1\,,$$

we have instability for the only possible equilibrium. Periodic behaviour is still possible, but we expect that if $h'(k)$ is large, behaviour will be chaotic. (This is a well-known result for discrete maps with 'hump' functions, see for example [20, p. 130ff]).

4.5 Tentative conclusions.

The simple models of this chapter exemplify some of the features discussed in Chapter 3, namely features 1–4 from the list given there: Asymmetry and chaotic background states, unsupervised activity and response to external input (in particular, bifurcation to stable states). They do not yet incorporate learning and it seems that to create a full-fledged neural network model that learns from experience requires some new insight as well as possibly more sophisticated mathematical tools from the theory of chaotic dynamical systems (symbolic dynamics might be useful,

for instance). However, as building blocks for such systems, they look promising. We emphasize that this has been achieved with Hopfield network equations having a particular structure to the connection matrix, with close interplay of positive and negative entries.

5. Integro-differential equations as neural fields

5.1 The integro-differential equation.

It has been observed in Chapter 3 that there is an argument that the appropriate level for neural modelling is that of averaged activity over an area consisting of many neurons. One way to implement this idea is to simply blur the distinction between individual neurons by starting with a continuous space model rather than a discrete space model. Instead of discrete neurons arranged with certain connections, we now imagine a 'neural field'. The system of ordinary differential equations describing the network then becomes a single integro-differential equation describing the field, with an integral replacing the sum.

An example of such an equation is

$$u_t(x,t) = -\alpha u(x,t) + \int_\Omega T(x,y) g(\lambda u(y,t))\, dy\,, \tag{5.1}$$

where $\Omega \subset \mathbf{R}^n$ is the domain over which the neural activity is defined (i.e. the spatial extent of the 'brain'). To be more general, we could allow T to be time-varying and include threshold and input terms. We might even let α be a function of the space variable, x, but for the time being we will stick with (5.1), and explore some of the possibilities. We remark that such equations have also been studied, but in a different light, by Amari [3,4] and by Coolen and Ruijgrok [13,57].

5.2 The energy functional.

When T is symmetric there is an energy functional for this equation analogous to the one for the Hopfield network (cf. equation 2.5), namely

$$E[u] = -\frac{1}{2}\int_\Omega \int_\Omega T(x,y) v(x,t) v(y,t)\, dy\, dx + \frac{\alpha}{\lambda}\int_\Omega \int_0^v G(V)\, dV\, dx\,,$$

where $v(x,t) = g(\lambda u(x,t))$ and $G = g^{-1}$, as before. If T is not symmetric, then this is not an energy functional, also as before. The demonstration of these facts is similar to the analogous demonstration for the Hopfield net in Section 2.2.

$$\begin{aligned}
\dot{E}[u] &= -\frac{1}{2}\int_\Omega\int_\Omega T(x,y)\left[v(x,t)v_t(y,t) + v_t(x,t)v(y,t)\right]dy\,dx \\
&\quad + \frac{\alpha}{\lambda}\int_\Omega G(v(x,t))v_t(x,t)\,dx \\
&= -\int_\Omega\int_\Omega\left[\frac{T(x,y)+T(y,x)}{2}\right]v(y,t)v_t(x,t)\,dy\,dx + \alpha\int_\Omega u(x,t)v_t(x,t)\,dx \\
&= -\int_\Omega\left[\int_\Omega T(x,y)g(\lambda u(y,t))\,dy - \alpha u(x,t)\right]v_t(x,t)\,dx \\
&= -\int_\Omega \lambda g'(\lambda(u(x,t))u_t^2(x,t)\,dx \le 0\,,
\end{aligned}$$

where we assume $Tvv_t \in L^1$, say, so that we can apply the Fubini-Tonelli theorem (see e.g. [25, p.65]).

5.3 Equilibria and gain.

Equilibria for equation (5.1) are found by setting

$$u(x) = \frac{1}{\alpha}\int_\Omega T(x,y)g(\lambda u(y))\,dy\,, \tag{5.2}$$

which is a kind of nonlinear Fredholm integral equation. We would like to say something about solutions to this equation, depending on the values of $T(x,y)$, α, and λ.

First, we can show that if $\frac{\lambda}{\alpha}\|T\|_{L^\infty} < \frac{1}{\mathrm{meas}(\Omega)}$, then the only solution is the trivial one, $u \equiv 0$. We do this for $x \in [a,b] \subset \mathbf{R}^1$. For simplicity, we will let $\alpha = 1$.

Lemma 5.1 *Let $T \in L^\infty([a,b]^2)$, $u(x) \in L^\infty([a,b])$ and $g \in L^\infty(\mathbf{R}^1)$. Let g be a monotone increasing sigmoid and $\|g'\|_{L^\infty} = g'(0) = 1$. Let $K = \|T\|_{L^\infty([a,b]^2)}$. Then, if $K\lambda < \frac{1}{(b-a)}$, the only solution to (5.2) is $u \equiv 0$.*

Proof Let

$$Au(x) = \int_a^b T(x,y)g(\lambda u(y))\,dy\,.$$

Then $Au(x) \in L^\infty[a,b]$. Note that $K < \infty$. Now,

$$\begin{aligned}\|Au - Av\|_{L^\infty} &= \left\| \int_a^b T(x,y)\left[g(\lambda u(y)) - g(\lambda v(y))\right]\,dy \right\|_{L^\infty} \\ &\leq \sup_{x\in[a,b]} \int_a^b |T(x,y)| \cdot |g(\lambda u) - g(\lambda v)|\,dy \\ &\leq K(b-a)\,\|g(\lambda u) - g(\lambda v)\|_{L^\infty}\,.\end{aligned}$$

Since g is monotone increasing and $\|g'\|_{L^\infty} = g'(0) = 1$, we have

$$|g(\lambda u) - g(\lambda v)| \leq \lambda\,|u - v|$$

and so

$$\|g(\lambda u) - g(\lambda v)\|_{L^\infty} \leq \lambda\,\|u - v\|_{L^\infty}\,.$$

So ther

$$\|Au - Av\|_{L^\infty} \leq K\lambda(b-a)\,\|u - v\|_{L^\infty}\,.$$

The contraction mapping theorem applies when $K\lambda(b-a) < 1$. Thus, if

$$K\lambda < \frac{1}{(b-a)}\,,$$

as assumed in the hypothesis, then there exists a unique $u(x) \in L^\infty([a,b|)$ satisfying

$$u(x) = \int_a^b T(x,y)g(\lambda u(y))\,dy\,.$$

Since $u(x) \equiv 0$ is a bounded solution to (5.2), there is no other by the contraction mapping theorem. □

We may normalize T by assuming that $K = 1$. Then, at least on bounded regions, the above lemma shows that the only equilibrium is the trivial one when the gain, λ, is low. The only chance for non-trivial equilibria is when the gain is sufficiently high.

5.4 Equilibria with a hard non-linearity.

One way to explore the possible behaviours and equilibrium patterns for the integro-differential equation is to go to the high gain limit and consider the relatively simple case of a hard nonlinearity in the transfer function, g, i.e. instead of a smooth sigmoid, such as tanh, we let

$$g(u) = \text{sgn}(u) = \begin{cases} -1 \text{ if } u < 0 \\ 0 \text{ if } u = 0 \\ 1 \text{ if } u > 0 \end{cases} .$$

Then (5.2) with $\alpha = 1$ becomes

$$u(x) = \int_\Omega T(x,y)\text{sgn}(u(y))\, dy . \tag{5.3}$$

Now we can examine what equilibria exist (equation 5.2) for various forms of the connection function, T. We expect that a continuous sigmoid with a high gain (large λ) will produce similar results but they are harder to characterize.

Consider first the following trivial case.

Example 5.1 If $T(x,y) = \chi_{[a,b]^2}(x,y)$, the characteristic function on all of $[a,b]^2$, then (5.3) becomes

$$u(x) = \int_a^b \text{sgn}u(y)\, dy$$

which is a constant for any $x \in [a,b]$. Either this constant is 0 or

$$u(x) > 0 \Rightarrow u(x) = \int_a^b dy = b - a \quad \text{or}$$

$$u(x) < 0 \Rightarrow u(x) = -\int_a^b dy = a - b .$$

Thus the only three equilibria are $u(x) = 0$ and $u(x) = \pm(b-a)\chi_{[a,b]}(x)$. Note that if we took instead $T(x,y) = -\chi_{[a,b]^2}(x,y)$, then the only equilibrium is the trivial one. □

We make a few comments about solutions to (5.3).

1. $u \equiv 0$ is always a solution.

2. If $u(x)$ is a solution, so is $-u(x)$.

3. If $T(x,y) \geq 0$ on Ω then there exists a unique positive solution $u(x) > 0$, namely

$$u(x) = \int_\Omega T(x,y)\,dy\,.$$

4. If T is a function of y only, then u is constant. Non-trivial solutions exist if $\int_\Omega T(y)\,dy > 0$. Then

$$u(x) = \pm \int_\Omega T(y)\,dy$$

are the two non-trivial solutions.

5. If T is of the form $T(x,y) = T_1(x)T_2(y)$, then $\frac{u(x)}{T_1(x)}$ must be constant and two non-trivial solutions exist if $\int_E T_2(y)\,dy > \int_F T_2(y)\,dy$, where $E = \{y : T_1(y) > 0\}$ and $F = \{y : T_1(y) < 0\}$. This can be seen by checking the two cases of a positive and a negative constant. If $K = \frac{u(x)}{T_1(x)} > 0$ then

$$\begin{aligned} K &= \int_\Omega T_2(y)\mathrm{sgn}(u(y))\,dy = \int_\Omega T_2(y)\mathrm{sgn}(KT_1(y))\,dy \\ &= \int_\Omega T_2(y)\mathrm{sgn}(T_1(y))\,dy = \int_E T_2(y)\,dy - \int_F T_2(y)\,dy \end{aligned}$$

and the last expression must be positive since K is positive. Similarly, if $K = \frac{u(x)}{T_1(x)} < 0$ then

$$K = -\int_\Omega T_2(y)\mathrm{sgn}(T_1(y))\,dy = \int_F T_2(y)\,dy - \int_E T_2(y)\,dy$$

and the last expression must be negative since K is negative.

5.5 The S – Σ exchange revisited.

Let us generalize from Section 2.3, the equivalence of two formulations of the problem given by the $S - \Sigma$ exchange.

Lemma 5.2 *Let $\Sigma : X \to Y$, $S : Y \to X$ be arbitrary mappings. Let $x_1, x_2 \in X$; $y_1, y_2 \in Y$. If $\Sigma x_1 = y_1$, $\Sigma x_2 = y_2$ and $x_2 = S\Sigma x_1$, then $y_2 = \Sigma S y_1$.*

Proof $x_2 = S\Sigma x_1 \Rightarrow x_2 = Sy_1 \Rightarrow \Sigma x_2 = \Sigma S y_1 \Rightarrow y_2 = \Sigma S y_1$. □

Corollary 5.3 *If $x \in X$, $y \in Y$, $\Sigma x = y$ and $x = S\Sigma x$, then $y = \Sigma S y$. Similarly, if $Sy = x$ and $y = \Sigma S y$, then $x = S\Sigma x$.*

Proof Direct from Lemma 5.2. □

This result is the simple idea behind the $S - \Sigma$ exchange. It was applied to equation (2.8) to get (2.10) and *vice versa* using (2.9) and (2.11). We can also apply it to equation (5.3) by defining

$$s(x) = \operatorname{sgn}(u(x)); \quad u(x) = \int_\Omega T(x,y)s(y)\,dy,$$

$$s = Su; \quad u = \Sigma s\,.$$

Then, from Corollary 5.3, $u = \Sigma S u \Leftrightarrow s = S\Sigma s$, so that solutions of (5.3) are equivalent to solutions of

$$s(x) = \operatorname{sgn} \int_\Omega T(x,y)s(y)\,dy\,, \tag{5.4}$$

where $s(x)$ takes values in $\{-1, 0, 1\}$.

5.6 Equilibrium solutions via S – Σ exchange.

Since $s(x)$ can only have the values -1, 0 or 1, all solutions to (5.4) are of the form $s(x) = \chi_E(x) - \chi_F(x)$. For example, if $\chi_{[a,b]}(x)$ is to be a solution for $[a,b] \subset \Omega$, then

$$\chi_{[a,b]}(x) = \text{sgn} \int_a^b T(x,y)\,dy\,.$$

This is true if $T(x,y) > 0$ and supported on $[a,b]^2$, for example. The same condition is required for $-\chi_{[a,b]}(x)$ to be a solution. To take this idea further, if T is supported (and positive) on n disjoint diagonal boxes, $[a_1,b_1]^2$, $[a_2,b_2]^2 \ldots [a_n,b_n]^2$, then there exist 3^n solutions, $s(x) = 0$ or ± 1 on each disjoint piece. Fig. 5.1 shows an example of a region of support for T with 3 disjoint boxes.

If T is simply a characteristic function on these n boxes, then the solutions to (5.3), $u(x)$, corresponding to these $s(x)$ are also constant on the intervals $[a_i,b_i]$ but the constants depend on the lengths of the intervals.

Overlaps in the diagonal boxes or off-diagonal regions, with T still a characteristic function on these regions are similar but some of the possible solutions may be lost, depending on the intervals. For example:

If T is taken to be negative then all the non-trivial equilibrium solutions are lost (and in the dynamic equations, oscillation occurs between positive and negative values).

If $T = \chi_{[a,b]}(x) \cdot \chi_{[c,d]}(y)$ with $[a,b]$, $[c,d]$ disjoint, then the trivial solution is the only one:

$$s(x) = \text{sgn} \int_c^d \chi_{[a,b]}(x) s(y)\,dy$$

and for $x \notin [a,b]$, $s(x) = 0$, and then for $x \in [a,b]$, $s(x) = \text{sgn} \int_c^d s(y)\,dy = 0$ since $s(y) = 0$ for $y \in [c,d]$.

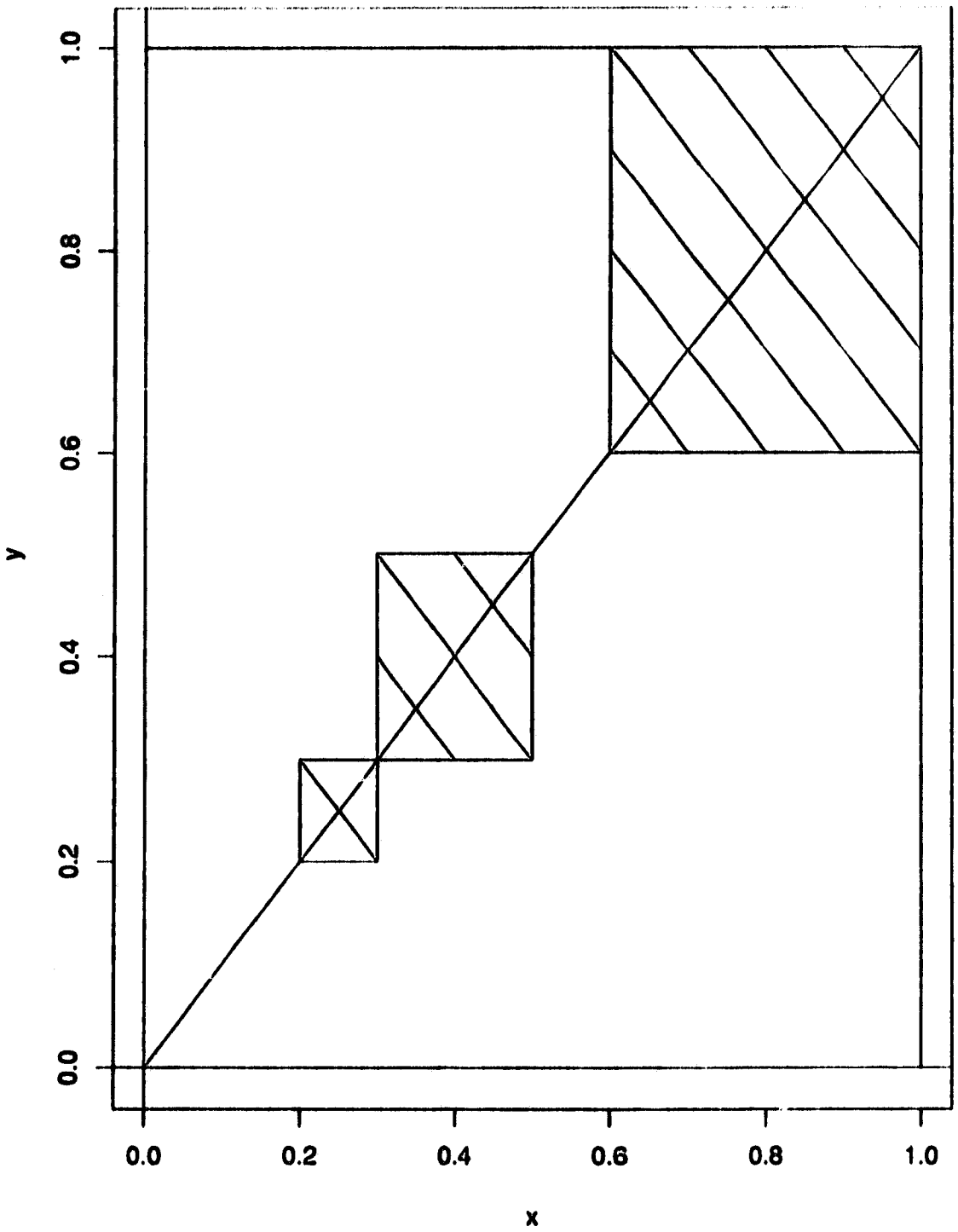

Figure 5.1 An example of a region of support for $T(x, y)$ with 3 disjoint boxes on the diagonal. Equation (5.4) with $\Omega = [0, 1]$ has 9 solutions in this case.

However, if with T as above, $[a,b]$ and $[c,d]$ have an overlap with positive measure (e.g. $a < c < b < d$), then $s(x) = \pm\chi_{[a,b]}(x)$ is a solution again. For $x \notin [a,b]$, $s(x) = 0$ and for $x \in [a,b]$, $s(x) = \operatorname{sgn} \int_c^d s(y)\,dy = \pm 1$, depending on the sign of s.

If T is taken to be something more complex than a characteristic function, then it is possible to get more equilibrium solutions. For example, let $E^+ \subset [a,b]$, $E^- \subset [a,b]$ be disjoint and

$$s(x) = \chi_{E_+}(x) - \chi_{E_-}(x)\,.$$

Then, from (5.4),

$$s(x) = \operatorname{sgn}\left[\int_{E_+} T - \int_{E_-} T\right]$$

so for $x \in E_-$,

$$\int_{E_+} T < \int_{E_-} T\,,$$

and for $x \in E_+$,

$$\int_{E_+} T > \int_{E_-} T\,,$$

so this s is a solution if T is bigger on $E_+ \times E_+$ and $E_- \times E_-$ than on $E_+ \times E_-$ and $E_- \times E_+$.

Example 5.2 Let

$$T(x,y) = 1 - (y-x)^2$$

on $[0,1]^2$. Then

$$s(x) = \chi_{(\frac{1}{2},1]}(x) - \chi_{[0,\frac{1}{2})}(x)$$

is a solution to (5.4) as is

$$s(x) = \chi_{[0,1]}(x)\,.$$

This is verified by noting that in the first case, equation (5.4) becomes

$$s(x) = \operatorname{sgn}\left[\int_{\frac{1}{2}}^{1} \left[1-(y-x)^2\right]\, dy - \int_{0}^{\frac{1}{2}} \left[1-(y-x)^2\right]\, dy\right]$$
$$= \operatorname{sgn}\left[\frac{x}{2} - \frac{1}{4}\right] = \chi_{(\frac{1}{2},1]}(x) - \chi_{[0,\frac{1}{2})}(x).$$

In the second case,

$$s(x) = \int_{0}^{1} \left[1-(y-x)^2\right]\, dy = \operatorname{sgn}\left[\frac{2}{3} + x - x^2\right] = \chi_{[0,1]}(x).$$

The functions $u(x)$ corresponding to these $s(x)$ are

$$u(x) = \pm\left(\frac{x}{2} - \frac{1}{4}\right),$$

$$u(x) = \pm\left(\frac{2}{3} + x - x^2\right),$$

and of course the trivial solution $u(x) \equiv 0$. □

Note that in this example we have five equilibrium solutions (at least) instead of the three for $T = \chi_{[0,1]^2}$. The equilibrium solutions in the u variable given above are not at all obvious from equation (5.3). The $S - \Sigma$ exchange allows a reformulation of the problem which makes solutions much easier to see.

5.7 Stability for discrete time dynamics.

So far, we have looked only at equilibria for the integro-differential equation model. The simplest way to re-introduce dynamics is to use discrete time steps, rather than going straightaway to the continuous time equation. Thus, we look at

$$s_{n+1}(x) = \operatorname{sgn} \int_{\Omega} T(x,y) s_n(y)\, dy$$

or the equivalent

$$u_{n+1}(x) = \int_{\Omega} T(x,y) \operatorname{sgn}(u_n(y))\, dy,$$

where the $S - \Sigma$ exchange consists of the mappings, $s_n(x) = \mathrm{sgn}(u_n(x))$ and $u_{n+1}(x) = \int_\Omega T(x,y)s_n(y)\,dy$.

Now, we can explore stability by considering slight perturbations of equilibria.

To get the feel of how this works, let us first look at our simple example (Example 5.1). Let $T(x,y) = \chi_{[a,b]^2}(x,y)$ so that $s(x) = \chi_{[a,b]}(x)$ is an equilibrium for the s equation, $u(x) = (b-a)\chi_{[a,b]}(x)$ is the corresponding equilibrium for the u equation. One type of perturbation is to add something small to the equilibrium of the u equation, say

$$u_0(x) = (b-a)\chi_{[a,b]}(x) + f(x),$$

where $\|f\|_{L^\infty} < b-a$. Then

$$\begin{aligned} u_1(x) &= \int_a^b \chi_{[a,b]}(x)\mathrm{sgn}(u_0(y))\,dy \\ &= \chi_{[a,b]}(x)\int_a^b dy = (b-a)\chi_{[a,b]}(x)\,. \end{aligned}$$

The equilibrium is restored in one time step. If, on the other hand, we perturb the zero equilibrium, by, say

$$u_0(x) = f(x),$$

with $f(x) > 0$ on E and $f(x) < 0$ on F, and $E, F \subset [a,b]$, then

$$\begin{aligned} u_1(x) &= \int_a^b \chi_{[a,b]}(x)\mathrm{sgn}(f(y))\,dy \\ &= \chi_{[a,b]}(x)\left(\int_E dy - \int_F dy\right) \\ &= (\mathrm{meas}(E) - \mathrm{meas}(F))\,\chi_{[a,b]}(x)\,, \end{aligned}$$

and

$$u_2(x) = \pm(b-a)\chi_{[a,b]}(x),$$

the sign depending on which has the greater measure, E or F. Thus, the zero equilibrium is unstable.

In fact, on whatever diagonal region, $[a,b]^2$, that T is supported, any positive or negative contribution in $s_0(x)$ will immediately be picked up by the entire region, $[a,b]$. In this sense, the region is 'recognized'. This idea can be used in simple forms of training for the model.

5.8 Training the discrete time equation.

As for the standard Hopfield network, training involves modifying the connections strengths, here represented by the connection function T. Most likely, this process will have to occur on a slower time scale, and the Hebb rule for long term potentiation is a reasonable mechanism to start with.

The simplest sort of learning rule, maintaining the simple structure of T as a characteristic function, would be

$$T_{n+1}(x,y) = \mathrm{sgn}\left[T_n(x,y) + s_{n+1}(x)s_{n+1}(y)\right], \tag{5.5}$$

with $T_0(x,y) = 0$. This rule simply superimposes squares representing each s. It is similar to the kind of learning rule already discussed for the discrete space Hopfield network, except that we have kept the size of T bounded by putting it through the nonlinearity at every step. We could leave this out and have a dynamic version of the 'learning' rule for the Hopfield network, (see Parisi [50] for a more fully worked out suggestion along these lines); however, we would like to continue to explore the type of connection functions suggested by our study of the integral equation.

We first examine alternate rules of the form

$$T_{n+1}(x,y) = T_n(x,y)s_{n+1}(x)s_{n+1}(y) \tag{5.6}$$

with $T_0(x,y) = 1$. By this rule any activity in s_1 will be registered in T_1, but then if s_2 is a characteristic function on a different region than s_1, the 'memory' of s_1 will be lost in T_2 (in fact, everything will be lost). We can do better by making

$$T_{n+1}(x,y) = T_n(x,y)s_{n+1}(x)s_{n+1}(y) + T_n(x,y)s^c_{n+1}(x)s^c_{n+1}(y), \qquad (5.7)$$

where $s^c = 1 - |s|$, i.e. s^c is a characteristic function on the complement of the set for which s is a characteristic function. Now, the activity s_1 is recorded in T_1 along with its complement (i.e. the inactive region). If s_2 is different, then it and its complement will also be registered in T_2. If the regions of activity of s_1 and s_2 are disjoint, then the equation will now be in a position to restore either s_1 or s_2 if only a portion of one of them is input to the s equation. It would also restore the complement of the union of the two regions if part of the complement is input. The process of modifying T by this rule suggests that we can train the equation to become sensitive to each activity state so far encountered and the differences between them and their complements.

Of course, this is just a way to get a feel for how the learning rule works. The equations for s and for T must evolve in conjunction and we must have a way to provide information to the system. This can be done by an external forcing term in the s equation,

$$s_{n+1}(x) = \mathrm{sgn}\left[\int_\Omega T_n(x,y)s_n(y)\,dy + I_n(x)\right]. \qquad (5.8)$$

If we stick to characteristic functions for the time being, an input at time step n will immediately become part of the activity at the next step. The learning equation for T might operate on a slower time scale so that only established activity states are registered in the connection function. For the time being, however, we try equations (5.7) and (5.8) together, and trace through a possible sequence of inputs, step by step.

Example 5.3

Time 0: $T_0 \equiv 1$, $s_0(x) \equiv 0$. Take $I_0(x) \equiv 0$.

Time 1: $s_1(x) = 0$ still and so $T_1 = 1$ still. Make $I_1(x) = \chi_{[a,b]}(x)$.

Time 2: Then $s_2(x) = \chi_{[a,b]}(x)$ and $T_2(x,y) = \chi_{[a,b]^2} + \chi_{([a,b]^c)^2}$ (see Fig. 5.2). Now if we let the input remain present for another time step, so $I_2 = I_1$, then the state of the equation also remains unchanged.

Time 3: $s_3 = s_2$ and $T_3 = T_2$. If we now turn off the input, $I_3 = 0$, the previous state has become fixed.

Time 4: $s_4 = s_2$ and $T_4 = T_2$, still. Now introduce a new input, $I_4(x) = \chi_{[c,d]}(x)$, where $a < c < b < d$, say.

Time 5: $s_5 = \chi_{[a,b]\cup[c,d]} = \chi_{[a,d]}$ and $T_5 = \chi_{[a,b]^2} + \chi_{[b,d]^2} + \chi_{([a,d]^c)^2}$. □

Thus, finer information is stored in T. This is on the right track but the problem is that once s has become active on a region, it can't be turned off there (except with a carefully tuned input) and so inputs cannot be introduced cleanly into the system.

In order to allow inputs to dominate the activity, they must be strong and consist of positive and negative regions; i.e. they should be binary with large positive and negative values Wherever the input is zero or small, the current state of s interferes. To exploit the above idea of progressive refinement in the connection function T, so that it can recognize previous inputs, and use the idea of the learning rule given by (5.6), we should keep T non-negative. An appropriate learning rule is

$$T_{n+1}(x,y) = \max\left(0, T_n(x,y)s_{n+1}(x)s_{n+1}(y)\right). \tag{5.9}$$

As before, whenever T_n gets set to zero in a region, it remains zero forever after. However we will now have $s(x) = \pm 1$, except on sets of measure 0, since our inputs

are to be essentially non-zero. Therefore the diagonal region of T will remain intact.

Let us follow the dynamics of the system of equations given by (5.8) and (5.9) for a few time steps.

Example 5.4 We initialize the system with no information.

Time 0: $s_0 \equiv -1$, $T_0 \equiv 1$. Now, if no input is initially present, $I_0 \equiv 0$, nothing changes.

Time 1: $s_1 \equiv -1$, $T_1 \equiv 1$. If we introduce an input $I_1 = K\left[\chi_{[a,b]}(x) - \chi_{[a,b]^c}(x)\right]$, with $K > \text{meas}(\Omega)$, the system activity immediately reflects the input and the connection function 'records' it.

Time 2: $s_2 = \chi_{[a,b]}(x) - \chi_{[a,b]^c}(x) = I_1/K$ and $T_2 = \chi_{[a,b]^2}(x,y) + \chi_{([a,b]^c)^2}(x,y)$ (Fig. 5.2). Now if the input remains active, nothing changes: $I_2 = I_1$.

Time 3: $s_3 = s_2$ and $T_3 = T_2$. If the input is now reset to zero, $I_3 = 0$, then the system still remains unchanged.

Time 4: $s_4 = s_2$, $T_4 = T_2$. Introducing a new input, $I_4 = K\left[\chi_{[c,d]}(x) - \chi_{[c,d]^c}(x)\right]$, with $a < c < b < d$ for example, we get

Time 5: $s_5 = \chi_{[c,d]}(x) - \chi_{[c,d]^c}(x) = I_4/K$ and $T_5 = \chi_{[a,c]^2}(x,y) + \chi_{[c,b]^2}(x,y) + \chi_{[b,d]^2}(x,y) + \chi_{([a,d]^c)^2}(x,y)$ (Fig. 5.3).

Now, if we could turn off the learning process, so that T remained fixed, then introducing an input close to either I_1 or I_4 would cause the system to immediately retrieve the corresponding previous input which is now 'stored' in the system's memory. For example, if we introduce $I_5 = K\left[\chi_{[a,b-\varepsilon]}(x) - \chi_{[a,b-\varepsilon]^c}(x)\right]$, with ε small, then

Time 6: $s_6 = I_1/K$. □

Many neural network models require learning and retrieval phases to be sepa-

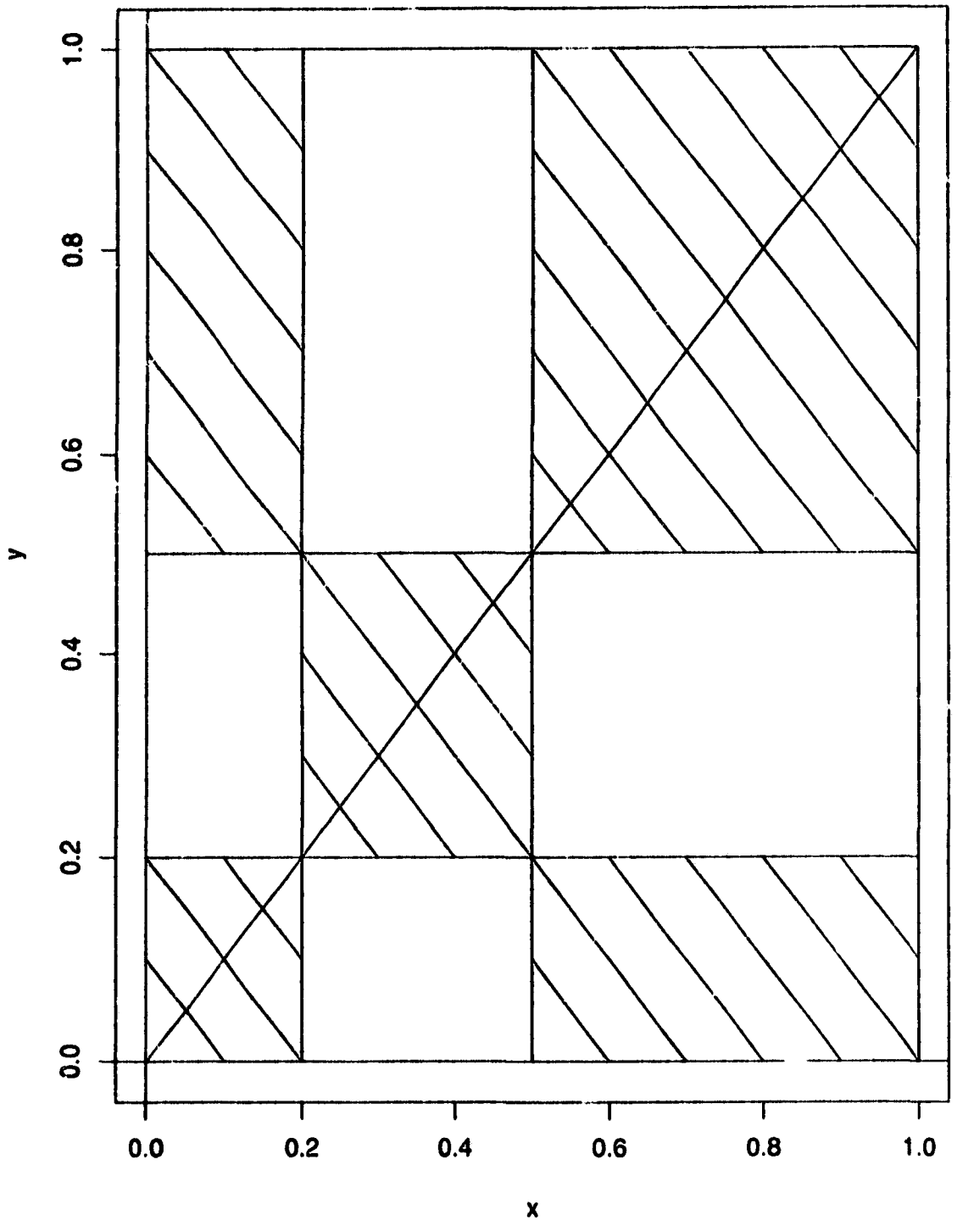

Figure 5.2 Support of $T(x,y)$ (shaded) after one input in Examples 5.3, 5.4 and 5.5. The region is $[a,b]^2 \cup ([a,b]^c)^2$ with $a = 0.2$, $b = 0.5$ and $\Omega = [0,1]$.

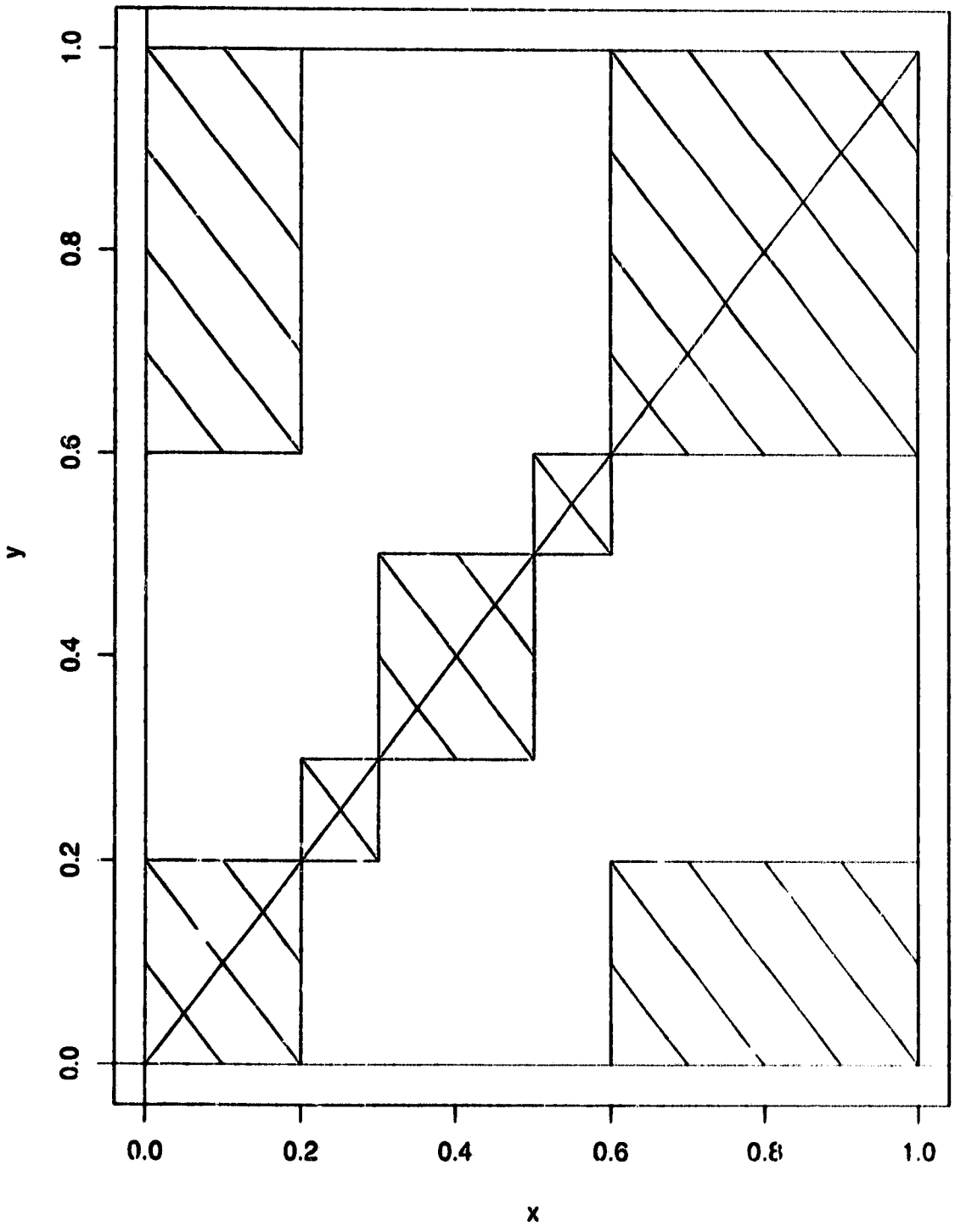

Figure 5.3 Support of $T(x, y)$ (shaded) after two inputs in Examples 5.4 and 5.5. The region is $[a, c]^2 \cup [c, b]^2 \cup [b, d]^2 \cup ([a, d]^c)^2$ with $a = 0.2$, $b = 0.5$, $c = 0.3$, $d = 0.6$ and $\Omega = [0, 1]$.

rated manually (i.e. an external operator is required to switch from one mode to another). From the point of view of biology, however, it is advantageous to have a dynamical system that is capable of handling both learning and retrieval in appropriate situations. For, example, the Carpenter-Grossberg classifier [45] takes an input as a new pattern to learn if it is not sufficiently similar to previously learned patterns. Otherwise it just retrieves the most similar previously learned pattern. Another way for a dynamical system to decide whether to learn a new pattern or to use it for retrieval is by the duration of the input. An input presented to the system for a short time might simply cause an attempt at retrieval of a similar pattern, and an input that remains present for a long period of time might be learned. This implies a difference in the time scales of neural activity and learning.

To introduce this idea into the above model, we could simply require that changes to the T function only occur if an input is present for two or more time steps (or, more precisely, if the system activity is present for two or more time steps).

This can be accomplished by changing the learning rule to

$$T_{n+1}(x,y) = \operatorname{sgn}\left[\sum_{k=0}^{m} \max\left(0, T_n(x,y)s_{n+1-k}(x)s_{n+1-k}(y)\right)\right], \tag{5.10}$$

where m determines the time scale of learning. Let $m = 1$, for example. Note that the evolution described by this equation consists of gradually setting parts of the T function to 0, from its initial state where it is 1 everywhere. It will be set to 0 only when a part of the space is consistently (over more than m time steps) pulled toward -1 by $s(x)s(y)$. Once $T(x,y)$ is set to 0 on a region, it must remain so, but note that the diagonal $T(x,x)$ will remain at 1 (as long as we ensure that activities $s(x)$ take only the values -1 and 1). Thus, the learning process is essentially the progressive refinement of the block structure on the diagonal.

With this rule and starting with the same time steps 0 and 1 as in the Example 5.4, we can follow the evolution from time step 2.

Example 5.5

Time 2: $s_2 = \chi_{[a,b]}(x) - \chi_{[a,b]^c}(x) = I_1/K$ and $T_2 = 1$. Now if the input is immediately turned off, $I_2 = 0$, then

Time 3: $s_3 = \text{sgn} \int_\Omega s_2(y)\,dy \equiv -1$, if $\text{meas}[a,b] < \text{meas}[a,b]^c$ (or $s_3 = 1$ if $\text{meas}[a,b]$ is larger). In this case, $T_3 \equiv 1$ still and the input is not learned. As yet, there are no stored patterns to retrieve. Now, introduce the same input for two time steps. So $I_3 = I_1$ and

Time 4: $s_4 = s_2$, $T_4 \equiv 1$. Also, $I_4 = I_1$ and

Time 5: $s_5 = s_2$ and $T_5 = \chi_{[a,b]^2}(x,y) + \chi_{([a,b]^c)^2}(x,y)$ (Fig. 5.2). Now the pattern is registered in memory. If the input remains active for further time steps, nothing changes. So consider $I_5 = 0$.

Time 6: Now $s_6 = s_2$ since this has become a fixed point of the dynamics. Also $T_6 = T_5$. Now introduce the second input, $I_6 = K\left[\chi_{[c,d]}(x) - \chi_{[c,d]^c}(x)\right]$, with $a < c < b < d$.

Time 7: $s_7 = I_6/K$, reflecting the input. $T_7 = T_5$ still. Again, let us first try turning off the input after one step, so $I_7 = 0$.

Time 8: For $x \in [a,b]^c$,

$$s_8(x) = \text{sgn} \int_{[a,b]^c} s_7(y)\,dy = \text{sgn}\left[\int_b^d dy - \int_{[a,d]^c} dy\right] = -1$$

if $\text{meas}[b,d] < \text{meas}[a,d]^c$. For $x \in [a,b]$,

$$s_8(x) = \text{sgn} \int_a^b s_7(y)\,dy = \text{sgn}\left[\int_c^b dy - \int_a^c dy\right] = \text{sgn}\{(b-c) - (c-a)\}.$$

Thus, if [c,d] is not too different from [a,b], i.e. a close to c and b close to d, then $b-c > c-a$ and $s_8(x) = \chi_{[a,b]}(x) - \chi_{[a,b]^c}(x)$. The stored pattern has been retrieved

from a similar input. If the input was closer to $\pm K\chi_\Omega$ or the negative of the stored pattern, then this other pattern would be retrieved. Of course, $T_8 = T_5$ still, since s_8 is essentially a previously known pattern. Now, introduce the same input for two time steps. $I_8 = I_6$ and

Time 9: $s_9 = I_6/K$, $T_9 = T_5$. Then $I_9 = I_6$ and

Time 10: $s_{10} = I_6/K$, $T_{10} = \chi_{[a,c]^2}(x,y) + \chi_{[c,b]^2}(x,y) + \chi_{[b,d]^2}(x,y) + \chi_{([a,d]^c)^2}(x,y)$ again (Fig. 5.3).

Now, as discussed above, if we introduce a pattern close to one of the known patterns, say $I_{10} = K\left[\chi_{[a,b-\varepsilon]} - \chi_{[a,b-\varepsilon]^c}\right]$, then the similar known pattern is retrieved in one step. As long as no input is active, the system can only remain at an existing fixed point pattern, and so no further changes occur to the connection function.

Of course, the possible inputs are not limited to characteristic functions on intervals, but can use much more complicated sets and could, for example, represent binary data, via characteristic functions of sets consisting of many small intervals.

Aside from the previous patterns and their negatives, there are now other fixed points of the dynamics, namely characteristic functions of set differences, intersections and unions of the sets used in the patterns stored. Thus, in the example used above, with $a < c < b < d$, if an input is constructed from the set $[a,c]$ or $[a,d]$, then $[a,c]$ or $[a,d]$ is retrieved. Moreover, if $[a, c-\varepsilon]$ is used for the input, then $[a,c]$ is retrieved but if $[a,c+\varepsilon]$ is used, then $[a,b]$ is retrieved. This suggests that too many inputs with large overlaps will prevent the effective retrieval of patterns, as there will be too many small 'spurious' fixed points. This also occurs in the conventional Hopfield net with Hebbian learning.

5.9 Tentative conclusions.

Although the inspiration for the model specified by equations (5.8) and (5.10) came from considering continuous space, it could be formulated in a discrete space setting again. The only operational difference is that the degree to which the connection function can be subdivided is limited in the discrete space setting, whereas with continuous space, there is no theoretical limit to the fineness of the divisions. In the discrete space setting, we would have the Hopfield network equations again but now with a different learning rule.

Our continuous space model, (5.8) and (5.10), has some of the features set out in Chapter 3. It exemplifies unsupervised activity with response to external input, unsupervised learning and the ability to store new memories or retrieve old ones appropriately as well as using a continuous space neural activity variable, i.e. features 3–7 from the list in Chapter 3. The development of a model with these features was facilitated by some of the simplifications such as the hard nonlinearity, the use of discrete time and the use of characteristic functions. However, we expect that it is possible to retain the desired features if we go back to a continuous sigmoid with high gain and continuous time, for example. The time scales of neural activity and learning differ according to the parameter m in equation (5.10). In a continuous time model, it would be necessary to replace the sum there by a time integral so that sustained activity would still be necessary to make significant changes to T. We do not pursue this further, however.

Note that the model is still symmetric and convergent, and therefore, while it may be an alternative to the conventional Hopfield model, it does not fulfil all our objectives in this regard. The positivity of the connection function T is necessary to retain fixed points but this is not biological: it excludes inhibition. As mentioned above, if T has negative values, fixed points are lost but oscillation becomes possible.

Also, we have so far no rigorous analysis of the implicit approximation of the discrete space neural network by a continuous space model when we take T to be a continuous function, say, rather than simply a characteristic function on a set. This could be pursued here, but we choose rather to explore another continuous space approximation in Chapter 6.

6. Approximation of neural network dynamics by a reaction-diffusion equation

6.1 Introduction.

The standard method by which neurobiologists take measurements of neural activity is either to place electrodes on the scalp or on the surface of the brain or in the extracellular space below the surface. They record the activity or potential of a pool of neurons in the general area around the electrode. (For a discussion of some of the complexities involved, see Traub and Miles [76, pp.191-193]). Although neurons are discrete entities, it appears that at least some (and perhaps all) of the interesting behaviour occurring in biological brains occurs at this level of averaged activity over an area. (See e.g. [60, pp.163,190; 26, pp.7-10]). This suggests that continuous space models might be capable of describing these behaviours.

Furthermore, it has been pointed out (e.g. by Cottet [15]) that it is appropriate to study the limiting system as the number of neurons in a network goes to infinity if one is interested in studying the behaviour of very large neural networks, particularly since the increase in size may cause changes in the type of dynamics and asymptotic behaviour.

The possibility of using a continuous space model with the same form as the discrete space model was explored in Chapter 5 (and in references given there). Neural activity is essentially modelled by a 'neural field' rather than a neural network, as in equation (5.1) for example. However, there needs to be some justification of the adoption of an 'analogous' continuous space model with continuous connection function. Is this really similar, or are there essential differences in taking this approach? In what sense is the continuous space model an approximation of the underlying discrete space network?

Cottet [14,15] (see also [16]) has taken the idea one step further in a way that helps to shed some light on these questions. He has shown at least formally that, in a restricted case of symmetric, translation-invariant connection functions T (*i.e.* $T(x,y) = T(y,x)$ and $T(x,y) = T(x+z, y+z)$), the integro-differential equation may be replaced by a certain partial differential equation (PDE). In fact, he claims (using techniques from particle methods in fluid dynamics, etc.) that solutions to Hopfield equations in this case are approximated arbitrarily closely by the solution to a PDE in a fixed time interval. Thus, he has attempted to give a rigorous analysis of the relationship between his continuous space and discrete space models, rather than simply presenting a new continuous space model. Although it appears that the convergence of the approximation cannot be rigorously carried through in general, it holds in some spatial regions (and a weaker form of convergence may still hold everywhere).

Cottet was particularly interested in applications of these equations to image processing problems (see also [15,16]) and for these problems, his restrictions on the connection function T are appropriate. In fact, he points out that Hopfield network equations with translation-invariant connection matrices are essentially applying a discrete convolution to an image. Another parallel can be drawn between connection matrices of this type and cellular automata (translation-invariant connections restricted to a window of nearby neurons correspond to the rules for updating cells in a cellular automaton). This has apparently not been explored before.

However, it is possible to apply the mathematical theory used by Cottet to more general T and obtain PDEs with solutions approximating those of the corresponding Hopfield nets over fixed time intervals. In particular, it is not necessary to insist that T be symmetric or translation-invariant. There are still, of course, restrictions on the form of T (smoothness, moment conditions, concentration around the diagonal),

but there is a class of such functions (and corresponding matrices) for which the Hopfield nets are approximated by the behaviour of PDEs. The particular PDEs that are obtained are of a type for which some theory exists so we have an idea as to how they behave. In any case, this opens up the possibility of bringing the analytic theory of PDEs to bear on the behaviour of neural network dynamics.

It is also of interest to observe for what types of connection function (or matrix) T the PDE approximation theorems break down. PDEs may not be capable of approximating all network dynamics, particularly complex, chaotic dynamics. The results of this chapter, for example, lead us to look at connection matrices T with wildly varying entries (in other words, interspersed inhibitory and excitatory connections) to produce the kind of complex behaviour not possible for the PDEs.

Our main object, however, is to prove a theorem estabishing the approximation of Hopfield net equations with a more general form of the function (or matrix) T than covered by Cottet's result. A discussion of the implications and limitations follows.

6.2 Approximation of an integral operator by a differential operator.

The theorem of this section is a generalization of that presented in Degond and Mas-Gallic [17] (see their Remark 2, p.491) and uses similar notation.

We let $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ (similarly β, γ, ...) denote a multi-index and we use the usual notational conventions:

$$\partial^\alpha f = \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}} \text{ and } x^\alpha = \prod_{i=1}^{n} x_i^{\alpha_i},$$

with

$$|\alpha| = \sum_{i=1}^{n} \alpha_i\,,\ \alpha! = \prod_{i=1}^{n} \alpha_i!,\ \text{and } \alpha + \beta = (\alpha_i + \beta_i)_{1 \le i \le n}\,.$$

We use the classical Sobolev spaces, $W^{m,p}(\Omega)$, of functions with m weak derivatives in L^p, with seminorm and norm given by

$$|f|_{m,p} = \left(\sum_{|\alpha|=m} \|\partial^\alpha f\|_{L^p}^p \right)^{\frac{1}{p}} \text{ and } \|f\|_{m,p} = \left(\sum_{0 \le |\alpha| \le m} \|\partial^\alpha f\|_{L^p}^p \right)^{\frac{1}{p}},$$

for $1 \le p < \infty$ and

$$|f|_{m,\infty} = \operatorname{ess} \sup_{|\alpha|=m, x \in \Omega} |\partial^\alpha f(x)| \text{ and } \|f\|_{m,\infty} = \sup_{0 \le k \le m} |f|_{k,\infty}\,.$$

When we deal with functions that depend on both space and time, we somewhat loosely use the above norms when the time dependence is not under consideration. Finally, let e_i denote the standard basis vector in $\mathbf{R}^n$ with 1 in the i^{th} component.

For $x \in \mathbf{R}^n$, we will be concerned with a function $\eta(x)$ with the following properties (moment conditions):

$$\int_{\mathbf{R}^n} x^\alpha \eta(x)\,dx = \begin{cases} \tau_0 & \text{if } \alpha = 0 \\ \tau_i & \text{if } \alpha = 2e_i \\ 0 & \text{if } \alpha \ne 2e_i,\ 1 \le |\alpha| \le r+1 \end{cases} \tag{6.1a}$$

and

$$\int_{\mathbf{R}^n} |x|^{r+2} |\eta(x)|\,dx \equiv K_{r+2} < \infty \tag{6.1b}$$

for some integer $r \ge 1$. For example, if η is even ($\eta(x) = \eta(-x)$), compactly supported and $\eta \in L^1(\mathbf{R}^n)$, then the moment conditions are satisfied with $r = 2$ (see Lemma 6.1 below). For $\varepsilon > 0$, we define the cutoff function

$$\eta_\varepsilon(x) = \frac{1}{\varepsilon^n} \eta\left(\frac{x}{\varepsilon}\right). \tag{6.1c}$$

We consider a continuous space analogue of the connection matrix of the Hopfield net of the following form. For $x, y \in \mathbf{R}^n$, define

$$T^\varepsilon(x,y) = \frac{\mu(x,y)\eta_\varepsilon(x-y)}{\varepsilon^2}. \tag{6.2}$$

We define the integral operator I^ε, which corresponds to the summed inputs from other neurons, by

$$I^\varepsilon f(x) = \int_{\mathbf{R}^n} T^\varepsilon(x,y) f(y)\, dy. \tag{6.3}$$

We will also use the shorthand (Grossberg and Mas-Gallic [17]):

$$I^\varepsilon f(x) = \int_{\mathbf{R}^n} T^\varepsilon(x,y) f(y)\, dy.$$

$$\begin{aligned} Z^\varepsilon_\alpha(x) &= \int_{\mathbf{R}^n} (y-x)^\alpha \eta_\varepsilon(x-y)\, dy\,, \\ \bar{Z}^\varepsilon_\alpha(x) &= \int_{\mathbf{R}^n} |(y-x)^\alpha|\, |\eta_\varepsilon(x-y)|\, dy\,. \end{aligned} \tag{6.4}$$

Notice first that in order to make the moments τ_0 and τ_i positive, as we will later on, it is not necessary that $\eta(x) \geq 0$.

Although it is not required for the main theorem, the following lemma gives an example of a type of function η to which the theorem will apply. Specifically even functions will satisfy the moment conditions (6.1a). First, observe that for any (integrable) function ϕ on $\mathbf{R}^1$,

$$\int_{-\infty}^{\infty} \phi(-v)\, dv = \int_{\infty}^{-\infty} \phi(w)(-dw) = -\int_{-\infty}^{\infty} \phi(w)(-dw) = \int_{-\infty}^{\infty} \phi(w)\, dw$$

with $w = -v$ and similarly in $\mathbf{R}^n$,

$$\int_{\mathbf{R}^n} \phi(-v)\, dv = (-1)^n \int_{\mathbf{R}^n} \phi(w)(-1)^n dw = \int_{\mathbf{R}^n} \phi(w)\, dw.$$

This will be used below.

Lemma 6.1 *If $\eta \in L^1(\mathbf{R}^n)$ and $\eta(-x) = \eta(x)$ then η satisfies conditions* (6.1a), *with $r = 2$.*

Proof Since $\eta \in L^1(\mathbf{R}^n)$, the $\alpha = 0$ condition is satisfied. Now consider the case where $|\alpha| = 1$. For $i = 1, 2, \ldots n$, letting $x = (x_i, \hat{x})$,

$$\begin{aligned}
\int_{\mathbf{R}^n} x_i \eta(x)\, dx &= \int_{-\infty}^{\infty} x_i \int_{\mathbf{R}^{n-1}} \eta(x_i, \hat{x})\, d\hat{x} dx_i \\
&= \int_0^{\infty} x_i \int_{\mathbf{R}^{n-1}} \eta(x_i, \hat{x})\, d\hat{x}\, dx_i + \int_{-\infty}^{0} x_i \int_{\mathbf{R}^{n-1}} \eta(x_i, \hat{x})\, d\hat{x}\, dx_i \\
&= \int_0^{\infty} x_i \int_{\mathbf{R}^{n-1}} \eta(x_i, \hat{x})\, d\hat{x}\, dx_i + \int_{-\infty}^{0} x_i \int_{\mathbf{R}^{n-1}} \eta(x_i, -\hat{x})\, d\hat{x}\, dx_i \\
&= \int_0^{\infty} x_i \int_{\mathbf{R}^{n-1}} \eta(x_i, \hat{x})\, d\hat{x}\, dx_i + \int_0^{\infty} (-x_i) \int_{\mathbf{R}^{n-1}} \eta(-x_i, -\hat{x})\, d\hat{x}\, dx_i \\
&= \int_0^{\infty} x_i \int_{\mathbf{R}^{n-1}} [\eta(x_i, \hat{x}) - \eta(-x_i, -\hat{x})]\, d\hat{x}\, dx_i = 0 .
\end{aligned}$$

The same reasoning works in the above for any α_i odd. When α_i is odd for some i, and $\alpha_j \neq 0$ for some other j, letting $\alpha = (\alpha_i, \hat{\alpha})$ so that $x^\alpha = x_i^{\alpha_i} \hat{x}^{\hat{\alpha}}$, we can apply the above reasoning with $\eta(x)$ replaced by $\hat{x}^{\hat{\alpha}} \eta(x)$ to get

$$\int_{\mathbf{R}^n} x^\alpha \eta(x)\, dx = \int_0^{\infty} x_i^{\alpha_i} \left(\int_{\mathbf{R}^{n-1}} \left[\hat{x}^{\hat{\alpha}} \eta(x_i, \hat{x}) - (-\hat{x})^{\hat{\alpha}} \eta(-x_i, -\hat{x}) \right] d\hat{x} \right) dx_i$$

which is 0 if the inner integral is 0. If all components of $\hat{\alpha}$ are even, then $\hat{x}^{\hat{\alpha}} = (-\hat{x})^{\hat{\alpha}}$ and we are done. If $\hat{\alpha}$ has an odd component, then we can show that both inner integrals are 0 by applying the same argument to $\int_{\mathbf{R}^{n-1}} \left[\hat{x}^{\hat{\alpha}} \eta(x_i, \hat{x}) \right] d\hat{x}$, and repeat until all odd components of α are removed. Thus, for any α with an odd component, the α^{th} moment of η is 0. So, in particular, the only α with $|\alpha| < 4$ which can give non-zero moments of η are $\alpha = 2e_i$. □

An example of a function $\eta(x)$ satisfying the moment conditions is given in Fig. 6.1. (This is the η used later in the example of Fig. 6.5).

Lemma 6.2 *Let $x, y \in \mathbf{R}^n$. Suppose η satisfies conditions* (6.1a) *and* (6.1b). *Then*

$$Z_\alpha^\varepsilon(x) = \begin{cases} \tau_0 & \text{if } \alpha = 0 \\ \varepsilon^2 \tau_i & \text{if } \alpha = 2e_i \\ 0 & \text{if } \alpha \neq 2e_i,\ 1 \leq |\alpha| \leq r+1 \end{cases}$$

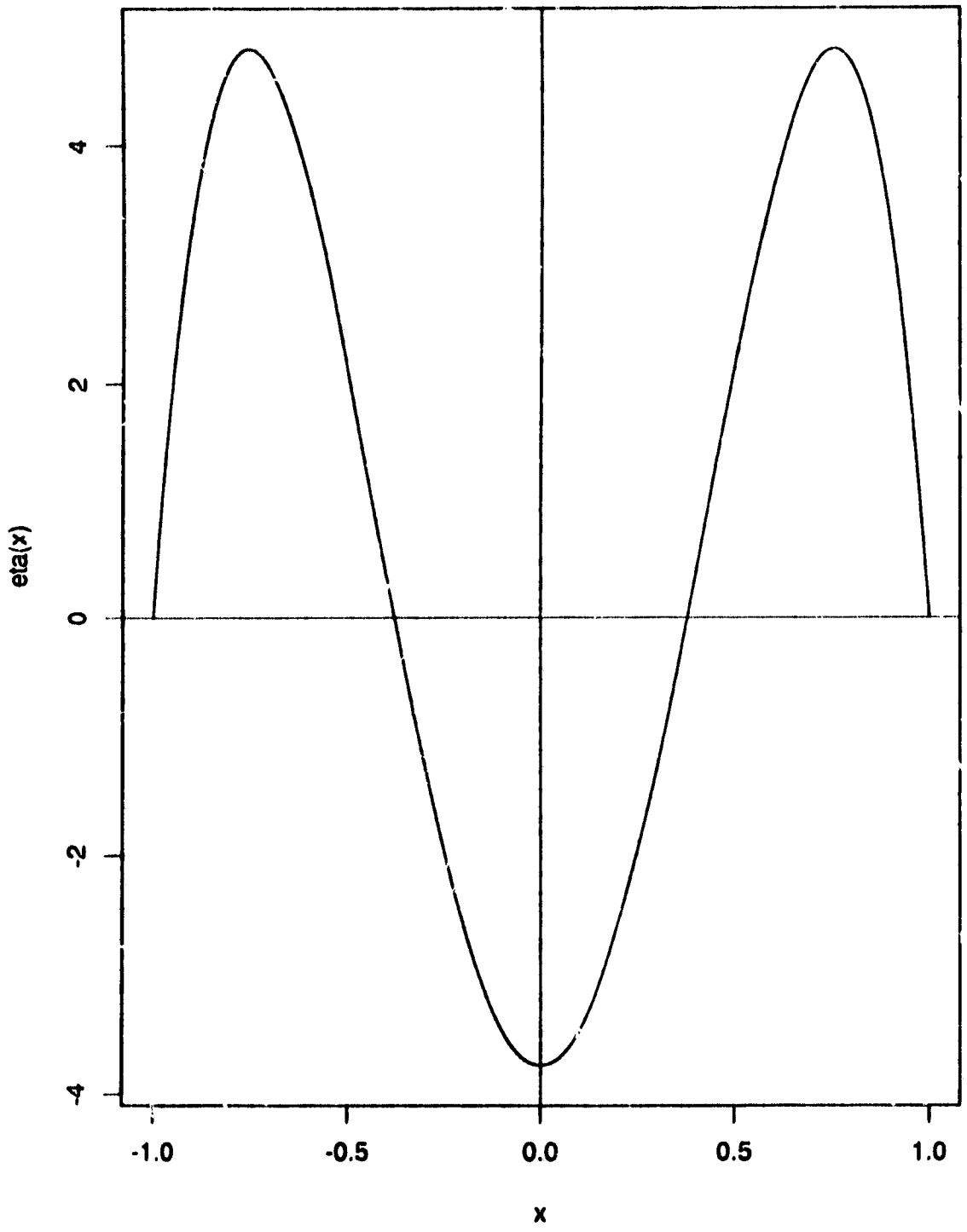

Figure 6.1 An example of an even $\eta(x)$ satisfying moment conditions (6.1a, 6.1b): $\eta(x) = \frac{15}{4}(x^2 - 1)(1 - 7x^2)$ on $[-1, 1]$. For this example, $\tau_1 = \tau_0 = \frac{2}{3}$.

and when $|\alpha| = r+2$,

$$|Z_\alpha^\varepsilon(x)| \le \bar{Z}_\alpha^\varepsilon(x) \le \varepsilon^{r+2} K_{r+2},$$

where Z_α^ε *and* $\bar{Z}_\alpha^\varepsilon$ *are defined by* (6.4).

Proof Let $v = \frac{y-x}{\varepsilon}$. Then $\varepsilon^n dv = dy$ and

$$Z_\alpha^\varepsilon(x) = \int_{\mathbf{R}^n} (\varepsilon v)^\alpha \left[\frac{1}{\varepsilon^n}\eta(-v)\right](\varepsilon^n dv) = \varepsilon^{|\alpha|}(-1)^{|\alpha|}\int_{\mathbf{R}^n} v^\alpha \eta(v)\, dv$$

and the result follows directly from (6.1a) and (6.1b). □

We now state and prove the theorem generalizing that of Degond and Mas-Gallic. The integral operator I^ε given by (6.3) will be approximated by the differential operator D_2 defined by

$$D_2 f(x) = \frac{1}{2}\mathrm{div}_y\left(\tau \nabla_y \left[\mu(x,y)f(y)\right]\right)\Big|_{(x,x)} + \frac{\tau_0}{\varepsilon^2}\mu(x,x)f(x), \tag{6.5}$$

where $\tau = \mathrm{diag}(\tau_1, \tau_2, \ldots, \tau_n)$, τ_i from equation (6.1*a*).

Theorem 6.3 *Let* $x, y \in \mathbf{R}^n$. *Suppose that* T^ε *is of the form given by* (6.2) *and* $I^\varepsilon f(x)$ *by* (6.3) *with*

$$\mu \in L^\infty\left(\mathbf{R}_x^n, W^{r+2,\infty}\left(\mathbf{R}_y^n\right)\right)$$

and also that η *satisfies conditions* (6.1a) *and* (6.1b). *Suppose also that* D_2 *is as given in* (6.5). *Then there exists a constant* $C > 0$, *depending on* K_{r+2} *(from* (6.1b)*) and* $\|\mu\|$, *the norm of* μ *in the above space, such that*

$$\|D_2 f - I^\varepsilon f\|_{0,\infty} \le C\varepsilon^r \|f\|_{r+2,\infty}$$

for any $f \in W^{r+2,\infty}(\mathbf{R}^n)$.

Proof In (6.3), we consider x to be fixed for the time being and expand $f(y)$ as a Taylor polynomial about $f(x)$ with integral remainder:

$$f(y) = \sum_{|\alpha|=0}^{r+1} \frac{1}{\alpha!}(y-x)^\alpha \partial^\alpha f(x) + R_f(x,y)$$

where

$$R_f(x,y) = (r+2)\sum_{|\alpha|=r+2}\frac{1}{\alpha!}(y-x)^\alpha\int_0^1(1-\theta)^{r+1}\partial^\alpha f(x+\theta(y-x))\,d\theta\,. \quad (6.6)$$

This gives the following representation for $I^\varepsilon f(x)$:

$$I^\varepsilon f(x) = I_T^\varepsilon f(x) + R_f^\varepsilon(x)\,, \quad (6.7)$$

where

$$I_T^\varepsilon f(x) = \sum_{|\alpha|=0}^{r+1}\frac{1}{\alpha!}\partial^\alpha f(x)\int_{\mathbf{R}^n}(y-x)^\alpha T^\varepsilon(x,y)\,dy\,,$$

$$R_f^\varepsilon(x) = \int_{\mathbf{R}^n} T^\varepsilon(x,y)R_f(x,y)\,dy\,. \quad (6.8)$$

Now we approximate $I_T^\varepsilon f(x)$ by expressing T^ε in terms of μ and η and expanding $\mu(x,y)$ about (x,x) for a particular α as follows:

$$\mu(x,y) = \sum_{|\beta|=0}^{r+1-|\alpha|}\frac{1}{\beta!}(y-x)^\beta\partial_y^\beta\mu(x,x) + R_{\mu,\alpha}(x,y)$$

where

$$R_{\mu,\alpha}(x,y) = (r+2-|\alpha|)$$
$$\times\sum_{|\beta|=r+2-|\alpha|}\frac{1}{\beta!}(y-x)^\beta\int_0^1(1-\tau)^{r+1-|\alpha|}\partial_y^\beta\mu(x,x+\tau(y-x))\,d\tau\,. \quad (6.9)$$

Then $I_T^\varepsilon f(x)$ can be expressed as

$$I_T^\varepsilon f(x) = \frac{1}{\varepsilon^2}\sum_{|\alpha|+|\beta|=0}^{r+1}\frac{1}{\alpha!}\frac{1}{\beta!}\partial^\alpha f(x)\partial_y^\beta\mu(x,x)Z_{\alpha+\beta}^\varepsilon(x) + R_{\mu,f}^\varepsilon(x)\,,$$

where

$$R_{\mu,f}^\varepsilon(x) = \frac{1}{\varepsilon^2}\sum_{|\alpha|=0}^{r+1}\frac{1}{\alpha!}\partial^\alpha f(x)\int_{\mathbf{R}^n}(y-x)^\alpha\eta_\varepsilon(x-y)R_{\mu,\alpha}(x,y)\,dy\,. \quad (6.10)$$

By Lemma 6.2, the only non-zero terms in the sum in $I_T^\varepsilon(x)$ are the term with $\alpha = \beta = 0$ and those with $\alpha + \beta = 2e_i$, $i = 1, \ldots, n$. Thus,

$$I_T^\varepsilon f(x) = \frac{\tau_0}{\varepsilon^2} f(x)\mu(\quad) + I_{2,0}(x) + I_{2,1}(x) + I_{2,2}(x) + R^\varepsilon_{\mu,f}(x) \tag{6.11a}$$

where

$$\begin{aligned} I_{2,0}(x) &= \frac{1}{2} f(x) \sum_{i=1}^n \tau_i \frac{\partial^2 \mu}{\partial y_i^2}\bigg|_{(x,x)} = \frac{1}{2} f(x) \left[\operatorname{div}_y (\tau \nabla_y \mu)\right]\Big|_{(x,x)}, \\ I_{2,1}(x) &= \sum_{i=1}^n \tau_i \frac{\partial f(x)}{\partial x_i} \frac{\partial \mu}{\partial y_i}\bigg|_{(x,x)} = (\tau \nabla f(x)) \cdot (\nabla_y \mu)\Big|_{(x,x)}, \\ I_{2,2}(x) &= \frac{1}{2}\mu(x,x) \sum_{i=1}^n \tau_i \frac{\partial^2 f(x)}{\partial x_i^2} = \frac{1}{2}\mu(x,x) \left[\operatorname{div} (\tau \nabla f(x))\right]. \end{aligned} \tag{6.11b}$$

Putting (6.7) and (6.11a) together, we have

$$I^\varepsilon f(x) = \frac{\tau_0}{\varepsilon^2} f(x)\mu(x,x) + I_{2,0}(x) + I_{2,1}(x) + I_{2,2}(x) + R^\varepsilon(x), \tag{6.12a}$$

with

$$R^\varepsilon(x) = R^\varepsilon_{\mu,f}(x) + R^\varepsilon_f(x). \tag{6.12b}$$

Note that the expressions in (6.11b) can be rewritten together as

$$I_{2,0}(x) + I_{2,1}(x) + I_{2,2}(x) = \frac{1}{2} \operatorname{div}_y (\tau \nabla_y [\mu(x,y) f(y)])\Big|_{(x,x)},$$

so by (6.5) and (6.12a),

$$I^\varepsilon f(x) = D_2 f(x) + R^\varepsilon(x).$$

Thus, the theorem is proved by obtaining a bound on R^ε.

To obtain this bound first note that the condition on μ in the statement of the theorem says that there is a single (essential) bound on the magnitude of μ and its

first $(r+2)$ derivatives in y everywhere. We will denote this bound, which is a norm on its space, by $\|\mu\|$, i.e.

$$\|\mu\| = \operatorname*{ess\,sup}_x \left(\|\mu\|_{W^{r+2,\infty}(\mathbf{R}^n_y)} \right) ,$$

$$\|\mu\|_{W^{r+2,\infty}(\mathbf{R}^n_y)} = \sup_{0 \le |\alpha| \le r+2} \|\partial_y^\alpha \mu\|_{L^\infty} .$$

Now we estimate each component of the error. From (6.6) and (6.8), and using (6.2), (6.4) and the hypotheses on μ and f,

$$|R_f^\varepsilon(x)| \le (r+2)|f|_{r+2,\infty} \sum_{|\alpha|=r+2} \frac{1}{\alpha!} \int_{\mathbf{R}^n} |T^\varepsilon(x,y)| \, |(y-x)^\alpha| \int_0^1 \left|(1-\theta)^{r+1}\right| \, d\theta \, dy ,$$

where $|\cdot|_{r+2,\infty}$ denotes the usual seminorm in $W^{r+2,\infty}$. But $\int_0^1 \left|(1-\theta)^{r+1}\right| \, d\theta = \frac{1}{r+2}$, so

$$\begin{aligned} |R_f^\varepsilon(x)| &\le |f|_{r+2,\infty} \sum_{|\alpha|=r+2} \frac{1}{\alpha!} \int_{\mathbf{R}^n} |T^\varepsilon(x,y)| \, |(y-x)^\alpha| \, dy \\ &\le \frac{1}{\varepsilon^2} |f|_{r+2,\infty} \|\mu\| \sum_{|\alpha|=r+2} \frac{1}{\alpha!} \bar{Z}_\varepsilon^\alpha(x) \\ &\le \varepsilon^r S_{r+2,n} K_{r+2} |f|_{r+2,\infty} \|\mu\| , \end{aligned}$$

by Lemma 6.2, where $S_{r+2,n} = \sum_{|\alpha|=r+2} \frac{1}{\alpha!}$. Similarly, from (6.9) and (6.10),

$$\begin{aligned} |R_{\mu,f}^\varepsilon(x)| &\le \frac{1}{\varepsilon^2} \|f\|_{r+1,\infty} \|\mu\| \sum_{|\alpha|=0}^{r+1} (r+2-|\alpha|) \sum_{|\beta|=r+2-|\alpha|} \frac{1}{\alpha!} \frac{1}{\beta!} \\ &\qquad \times \int_{\mathbf{R}^n} \left|(y-x)^{\alpha+\beta}\right| \, |\eta_\varepsilon(x-y)| \, dy \int_0^1 \left|(1-\tau)^{r+1-|\alpha|}\right| \, d\tau \\ &\le \frac{1}{\varepsilon^2} \|f\|_{r+1,\infty} \|\mu\| \sum_{|\alpha|=0}^{r+1} \sum_{|\beta|=r+2-|\alpha|} \frac{1}{\alpha!} \frac{1}{\beta!} \bar{Z}_\varepsilon^{\alpha+\beta}(x) \\ &\le \varepsilon^r C_{r+2,n} K_{r+2} \|f\|_{r+1,\infty} \|\mu\| , \end{aligned}$$

where $C_{r+2,n} = \sum_{|\alpha|=0}^{r+1} \sum_{|\beta|=r+2-|\alpha|} \frac{1}{\alpha!} \frac{1}{\beta!}$. Thus, (6.12b) gives

$$|R^\varepsilon(x)| \le \varepsilon^r \left(S_{r+2,n} + C_{r+2,n}\right) K_{r+2} \|f\|_{r+2,\infty} \|\mu\| .$$

This is true for all x, so

$$\|D_2 f - I^\varepsilon f\|_{0,\infty} = \|R^\varepsilon\|_{0,\infty} \leq C\varepsilon^r \|f\|_{r+2,\infty},$$

for some $C > 0$, depending on $\|\mu\|$ and $K_{r+2} = \int_{\mathbf{R}^n} |x|^{r+2} |\eta(x)|\, dx$. ||

Remark 1 If η is even and $\eta \in L^1$ then the above theorem is true with $r = 2$ (as long as μ also satisfies the conditions of the theorem).

Remark 2 The above theorem is also true if $\mathbf{R}^n$ is replaced by a subset $\Omega \subset \mathbf{R}^n$, as long as Ω is also used in the conditions of the theorem (moment conditions, etc.). In the case of Lemma 6.1, an even function η will satisfy the moment conditions with $r = 2$ as long as Ω is symmetric about the origin (so that if $x \in \Omega$, so is $-x$).

Remark 3 Simpler forms of T^ε (6.2) lead to simpler forms of the differential operator D_2 (6.5). For example:

(i) If $\tau_i = \hat{\tau}$, a constant, then

$$D_2 f(x) = \frac{\hat{\tau}}{2} \Delta_y \left[\mu(x,y) f(y)\right]\Big|_{(x,x)} + \frac{\tau_0}{\varepsilon^2} \mu(x,x) f(x).$$

(ii) If $\mu \equiv 1$ then

$$D_2 f(x) = \frac{1}{2} \mathrm{div}\left(\tau \nabla f(x)\right) + \frac{\tau_0}{\varepsilon^2} f(x).$$

(iii) If $\mu \equiv 1$ and $\tau_i = \hat{\tau}$ then

$$D_2 f(x) = \frac{\hat{\tau}}{2} \Delta f(x) + \frac{\tau_0}{\varepsilon^2} f(x). \tag{6.13}$$

To aid in an intuitive understanding of this idea of approximating a differential operator by an integral operator, consider standard finite difference approximations. For second order derivatives, a central difference scheme amounts to an approximation by a weighted sum of neighbouring points: $u_{xx} \sim -2u_i + u_{i-1} + u_{i+1}$. Other

techniques use more neighbouring points: $u_{xx} \sim -au_i + \sum_{j\sim i} a_{ij}u_j$. It is not so different, then, to use an integral in place of the sum over neighbouring points.

6.3 Quadrature for the integral.

We have shown that the differential operator D_2 is approximated by the integral operator I^ε. We now wish to approximate I^ε by a sum over points on a grid with grid point spacing h. This can be done as in the literature on particle methods. Particle methods allow us to obtain approximations of arbitrarily high order, given enough smoothness in η and μ. (Of course, high order approximations may involve very large constants). The idea behind particle methods is to approximate a continuous flow (e.g. of a fluid) by motions of a collection of delta functions or small but finite-sized 'particles'. Although *motions* of the particles are not involved in our present problem (our 'particles' are neurons), we can use the same techniques. For more background on particle methods, see [17,46,52] for example.

Let $x_i = ih$, $i \in \mathbf{Z}^n$ and similarly for x_j. Our discrete analogue of the integral operator (6.3) will be:

$$I_h^\varepsilon f(x_i) = \frac{h^n}{\varepsilon^{2+n}} \sum_{j\in\mathbf{Z}^n} T_{ij} f(x_j)$$

where

$$T_{ij} = \varepsilon^{2+n} T^\varepsilon(x_i, x_j) = \mu(x_i, x_j)\, \eta\left(\frac{x_i - x_j}{\varepsilon}\right).$$

Theorem 6.4 *If $\eta \in W^{m,1}(\mathbf{R}^n)$ and $\mu \in W^{m,\infty}(\mathbf{R}^n \times \mathbf{R}^n)$ for $m > n$, then there exists a constant C, depending on m, n, $\|\mu\|_{m,\infty}$ and $\|\eta\|_{m,1}$ (but not h or ε) such that*

$$\sup_i |I^\varepsilon f(x_i) - I_h^\varepsilon f(x_i)| \le C\left(\frac{h^m}{\varepsilon^{m+2}}\right) \|f\|_{m,\infty},$$

for all $f \in W^{m,\infty}(\mathbf{R}^n)$.

Proof Define a cell, B_j, around each grid point by

$$B_j = \left\{ y \in \mathbf{R}^n, \left(j_k - \frac{1}{2}\right) h \le y_k \le \left(j_k + \frac{1}{2}\right) h, 1 \le k \le n \right\},$$

where the subscript k here refers to a component of an n-vector. Thus, the grid point $x_j = jh$ is the centre of B_j. Let

$$f_h(y) = h^n \sum_{j \in \mathbf{Z}^n} f(x_j) \delta(y - x_j). \tag{6.14}$$

This is the particle approximation. Now let

$$E_j(g) = \int_{B_j} g(y)\, dy - h^n g(x_j).$$

We let $g = fT^\varepsilon(x, \cdot)$ so that our quadrature error is,

$$\begin{aligned} \langle f - f_h, T^\varepsilon(x, \cdot) \rangle &= \int_{\mathbf{R}^n} T^\varepsilon(x, y) f(y)\, dy - h^n \sum_{j \in \mathbf{Z}^n} T^\varepsilon(x, x_j) f(x_j) \\ &= \sum_{j \in \mathbf{Z}^n} E_j \left(f T^\varepsilon(x, \cdot) \right), \end{aligned}$$

where $\langle \cdot, \cdot \rangle$ refers to the usual inner product in L^2. Raviart's [52] Theorem 3.1 with $p = 1$ says that there exists a $C > 0$ so

$$\left| \sum_{j \in \mathbf{Z}^n} E_j(g) \right| \le C h^m \sum_{j \in \mathbf{Z}^n} |g|_{m,1,B_j},$$

as long as $m > n$, and $g \in W^{m,1}(\mathbf{R}^n)$. With $g(y) = f(y) T^\varepsilon(x, y)$, this gives a bound on the quadrature error of

$$\left| \sum_{j \in \mathbf{Z}^n} E_j \left(f T^\varepsilon(x, \cdot) \right) \right| \le C h^m \sum_{j \in \mathbf{Z}^n} |f T^\varepsilon(x, \cdot)|_{m,1,B_j}.$$

Now, we take L^∞ norms on both sides of this inequality to obtain

$$\begin{aligned} \left\| \sum_{j \in \mathbf{Z}^n} E_j \left(f T^\varepsilon(x, \cdot) \right) \right\|_{L^\infty} &\le C h^m \left\| \sum_{j \in \mathbf{Z}^n} \sum_{|\alpha| = m} \int_{B_j} \left| \partial^\alpha \left(f(y) T^\varepsilon(x, y) \right) \right| dy \right\|_{L^\infty} \\ &\le C h^m \sum_{|\alpha| = m} \sum_{\beta + \gamma = \alpha} \left\| \int_{\mathbf{R}^n} |\partial^\beta f| \cdot |\partial_y^\gamma T^\varepsilon(x, y)|\, dy \right\|_{L^\infty} \\ &\le C h^m \sum_{|\beta| + |\gamma| = m} \left\| \int_{\mathbf{R}^n} |\partial^\beta f| \cdot |\partial_y^\gamma T^\varepsilon(x, y)|\, dy \right\|_{L^\infty}. \end{aligned}$$

But from (6.2)

$$\begin{aligned}|\partial_y^\gamma T^\varepsilon(x,y)| &= \frac{1}{\varepsilon^2}\left|\partial_y^\gamma\left(\mu(x,y)\eta_\varepsilon(x-y)\right)\right| \\ &\le \frac{C}{\varepsilon^2}\sum_{\zeta+\xi=\gamma}\left|\partial_y^\zeta\mu(x,y)\right|\left|\partial^\xi\eta_\varepsilon(x-y)\right| .\end{aligned}$$

So,

$$\begin{aligned}&\left\|\sum_{j\in\mathbf{Z}^n} E_j\left(fT^\varepsilon(x,\cdot)\right)\right\|_{L^\infty} \\ &\quad\le C\frac{h^m}{\varepsilon^2}\sum_{|\beta|+|\zeta|+|\xi|=m}\left\|\int_{\mathbf{R}^n}\left|\partial^\beta f\right|\left|\partial_y^\zeta\mu(x,y)\right|\left|\partial^\xi\eta_\varepsilon(x-y)\right|dy\right\|_{L^\infty} \\ &\quad\le C\frac{h^m}{\varepsilon^2}\sum_{|\beta|+|\zeta|+|\xi|=m}\|\mu\|_{\zeta,\infty}\|\partial^\beta f\|_{L^\infty}\|\partial^\xi\eta_\varepsilon\|_{L^1} ,\end{aligned}$$

by Young's inequality. Finally, noting that $\|\partial^\xi\eta_\varepsilon\|_{L^1} = \frac{1}{\varepsilon^{|\xi|}}\|\partial^\xi\eta\|_{L^1}$ and taking $\varepsilon < 1$, we obtain the error bound

$$\left\|\sum_{j\in\mathbf{Z}^n} E_j\left(fT^\varepsilon(x,\cdot)\right)\right\|_{L^\infty} \le C\frac{h^m}{\varepsilon^{m+2}}\|\mu\|_{m,\infty}\|\eta\|_{m,1}\|f\|_{m,\infty} .$$

The quadrature error is a function of x, so the maximum error at a grid point, x_i, is bounded by the above:

$$\begin{aligned}\sup_i |I^\varepsilon f(x_i) - I_h^\varepsilon f(x_i)| &= \sup_i\left|\int_{\mathbf{R}^n} T^\varepsilon(x_i,y)f(y)\,dy - h^n\sum_{j\in\mathbf{Z}^n} T^\varepsilon(x_i,x_j)f(x_j)\right| \\ &\le \left\|\int_{\mathbf{R}^n} T^\varepsilon(x,y)f(y)\,dy - h^n\sum_{j\in\mathbf{Z}^n} T^\varepsilon(x,x_j)f(x_j)\right\|_{L^\infty} \\ &= \left\|\sum_{j\in\mathbf{Z}^n} E_j\left(fT^\varepsilon(x,\cdot)\right)\right\|_{L^\infty} \\ &\le C\frac{h^m}{\varepsilon^{m+2}}\|\mu\|_{m,\infty}\|\eta\|_{m,1}\|f\|_{m,\infty} .\end{aligned}$$

Although T^ε and f are only defined almost everywhere, $\mu \in W^{m,\infty}(\mathbf{R}^n\times\mathbf{R}^n)$, $\eta \in W^{m,1}(\mathbf{R}^n)$ and $f \in W^{m,\infty}(\mathbf{R}^n)$ guarantee that fT^ε is in $W^{m,1}(\mathbf{R}^n)$ and

since by the Sobolev embedding theorems [1] $W^{m,1}(\Omega) \hookrightarrow C^0(\Omega)$, we may assume that we have a representative of the equivalence class of f^ε that is continuous on bounded sets. This establishes the result. □

Remark 1 The result is also true if we restrict the spatial domain to $\Omega \subset \mathbf{R}^n$.

Remark 2 It is possible to carry this proof through almost unchanged with the L^p norm of the error if $f \in W^{m,p}(\mathbf{R}^n)$, at least if we are on a compact domain or if η has compact support (recall that $I_h^\varepsilon f(x)$ is defined in terms of delta functions (6.14)). This gives

$$\|I^\varepsilon f - I_h^\varepsilon f\|_{L^p} \leq C \frac{h^m}{\varepsilon^{m+2}} \|\mu\|_{m,\infty} \|\eta\|_{m,1} \|f\|_{m,p}.$$

6.4 Generalization of Cottet's result.

We now use the above results to generalize the theorem of Cottet [14]. This amounts to combining the two approximations above to obtain an integral operator close to the differential operator and then a sum close to the integral. These approximations are in terms of the solution to the PDE, so we must also have an *a priori* bound on the solution in terms of the initial data. These are then used to show that solutions to an equation involving the sum (the Hopfield network) are approximated by solutions to a partial differential equation involving the differential operator D_2.

Consider the Hopfield network equations expressed in terms of firing rates as in (2.7) but without threshold or input terms for the time being, i.e.

$$\dot{v}_i = \frac{1}{G'(v_i)} \left[\lambda \sum_j T_{ij} v_j - \alpha G(v_i) \right]. \tag{6.15}$$

We will suppose that g is a strictly increasing odd function taking values in (-1,1) and with a maximum slope at 0 as in conditions (2.4). Thus, $G = g^{-1}$ is well-defined

and strictly increasing on (-1,1). Also, $G'(v) \geq G'(0) = \beta$ for all v, since G' has a minimum at 0. (We could set $\beta = 1$ without loss of generality, since we already have the parameter λ to modify the gain or slope of g and hence of G).

Consider also the partial differential equation,

$$\begin{aligned} v_t &= \frac{1}{G'(v)} \left[\gamma \varepsilon^2 D_2 v - \alpha G(v)\right] \\ &= \frac{1}{G'(v)} \left[\frac{\gamma \varepsilon^2}{2} \mathrm{div}_y \left(\tau \nabla_y \left[\mu(x,y) v(y)\right]\right)\Big|_{(x,x)} + \gamma \tau_0 \mu(x,x) v(x) - \alpha G(v)\right], \end{aligned} \tag{6.16}$$

where $D_2 v$ is as in (6.5) above, G is as described above for equation (6.15) and

$$\gamma = \lambda \left(\frac{\varepsilon}{h}\right)^n. \tag{6.17}$$

Here, ε and h are the parameters involved in the two approximations, ε for replacing the differential operator by the integral and h for quadrature of the integral. Existence and uniqueness are also not difficult to establish for this equation under the conditions

$$\tau_i \mu(x,x) > 0, \forall x, \forall i \tag{6.18}$$

$$v_0(x) \in (-1,1). \tag{6.19}$$

In fact, there is a maximum principle guaranteeing boundedness and global existence of solutions as long as initially $v \in (-1,1)$. We could also recast the equation in terms of u, the membrane potential, giving another quasi-linear parabolic equation, and then global existence for this equation guarantees that $v = g(\lambda u)$ remains in $(-1,1)$. Proofs of global existence and uniqueness for such quasi-linear equations can be found, for example, in [44].

However, we also need control over derivatives of solutions to equation (6.16) since the approximations in Theorem 6.3 and Theorem 6.4 involve Sobolev norms of

the solution. It is implied in Cottet's Theorem 1 that there exist bounds, independent of ε and γ, on derivatives of the solution to his equation, in terms of similarly bounded initial data and a fixed time interval. It appears that such a condition is unlikely to hold globally (see discussion in Section 6.5) but we may expect it to hold on some regions.

For the time being we will assume for the sake of simplicity that we have an *a priori* bound:

$$\|v\|_{m,\infty} \leq C\|v_0\|_{m,\infty}, \tag{6.20}$$

for $x \in \mathbf{R}^n$, $v_0 \in W^{m,\infty}$, where C is independent of ε and γ.

Now we give our main result.

Theorem 6.5 *Let τ, τ_0 and $T^\varepsilon(x,y)$ be defined above in* (6.1a), (6.2) *and Theorem 6.3, and let μ and η satisfy the conditions of that theorem with $\eta \in W^{m,1}(\mathbf{R}^n)$ and $\mu \in W^{m,\infty}(\mathbf{R}^n \times \mathbf{R}^n)$ for $m > n$, $m \geq r+2$ (r from the moment conditions on η,* (6.1)). *Let $v_i(t)$ be the solution to* (6.15) *with initial conditions*

$$v_i(0) = v_{i,0},\ i \in \mathbf{Z}^n.$$

Let $w(x,t)$ be the solution to (6.16) *with initial conditions*

$$w(x,0) = v_0(x),$$

such that $v_0 \in W^{m,\infty}(\mathbf{R}^n)$ and

$$v_0(x_i) = v_{0,i}$$

(i.e. initial conditions coincide). Assume also that conditions (6.18), (6.19) *and* (6.20) *hold. Then, for $t \in [0,M]$ these solutions satisfy*

$$\sup_i \|v_i(t) - w(x_i,t)\|_{L^\infty} \leq C\gamma\varepsilon^2\left(\varepsilon^r + \left(\frac{h^m}{\varepsilon^{m+2}}\right)\right)$$

for some constant $C > 0$, depending on the parameters α, λ, β, and T_{ij} as well as

$$\|\mu\|_{m,\infty}, \|\eta\|_{m,1}, K_{r+2}, \|v_0\|_{m,\infty}, M.$$

Proof By Theorem 6.3,

$$\begin{aligned} w_t &= \frac{1}{G'(w)}\left[\gamma\varepsilon^2 D_2 w - \alpha G(w)\right] \\ &= \frac{1}{G'(w)}\left[\gamma\varepsilon^2 I^\varepsilon w - \alpha G(w)\right] + \frac{\gamma\varepsilon^2}{G'(w)} R^\varepsilon(x), \end{aligned} \tag{6.21}$$

where $|R^\varepsilon(x)| \le C\varepsilon^r \|v\|_{m,\infty}$. Note that $W^{m,\infty}(\mathbf{R}^n \times \mathbf{R}^n) \subset L^\infty(\mathbf{R}^n_x, W^{m,\infty}(\mathbf{R}^n_y))$ so that any μ satisfying the conditions of this theorem must also satisfy those of Theorem 6.3, as long as $m \ge r+2$, which we have assumed. Then, by Theorem 6.4,

$$w_t = \frac{1}{G'(w)}\left[\gamma\varepsilon^2 I^\varepsilon_h w - \alpha G(w)\right] + \frac{\gamma\varepsilon^2}{G'(w)} R(x),$$

where $|R(x)| \le C\left(\varepsilon^r + \left(\frac{h^m}{\varepsilon^{m+2}}\right)\right)\|v(t)\|_{m,\infty}$ at any time $t \in [0, M]$, and so by condition (6.20), $|R(x)| \le C\left(\varepsilon^r + \left(\frac{h^m}{\varepsilon^{m+2}}\right)\right)$. In other words,

$$G'(w_i)w_t(x_i,t) = \lambda\sum_j T_{ij}w_j - \alpha G(w_i) + \gamma\varepsilon^2 R(x_i)$$

at grid points. So the solution w_i of the PDE at grid points and the solution v_i of the system of ODE's at grid points satisfy

$$G'(v_i)\dot{v}_i - G'(w_i)\dot{w}_i = \lambda\sum_j T_{ij}(v_j - w_j) + \alpha(G(v_i) - G(w_i)) - \gamma\varepsilon^2 R.$$

Now integrate and take absolute values, noting that the initial conditions for v and w are equal.

$$\begin{aligned} &|G(v_i) - G(w_i)| \\ &\le \int_0^t \left|\lambda\sum_j T_{ij}(v_j(s) - w_j(s)) + \alpha(G(v_i(s)) - G(w_i(s))) - \gamma\varepsilon^2 R(s)\right| ds \\ &\le \int_0^t \left[\lambda\sum_j |T_{ij}|\,|v_j(s) - w_j(s)| + \alpha\,|G(v_i(s)) - G(w_i(s))| + \gamma\varepsilon^2\,|R(s)|\right] ds. \end{aligned}$$

Taking the supremum over all i and writing

$$T = \sup_i \sum_j |T_{ij}|,$$

we have

$$\|G(v(t)) - G(w(t))\|_\infty$$

$$\leq \int_0^t \left[\lambda T \|v(s) - w(s)\|_\infty + \alpha \|G(v(s)) - G(w(s))\|_\infty + \gamma \varepsilon^2 \|R(s)\|_\infty\right] ds.$$

Note that the supremum above exists since $\mu \in W^{m,\infty}(\mathbf{R}^n)$ and $\eta \in W^{m,1}(\mathbf{R}^n)$, and since $W^{m,1}(\Omega) \hookrightarrow C^0(\Omega)$ for bounded Ω by a Sobolev embedding theorem we may take η to be continuous.

Now observe that $|v_i - w_i| \leq \frac{1}{\beta} |G(v_i) - G(w_i)|$ where $\beta = G'(0)$ since by the mean value theorem:

$$\frac{G(v_i) - G(w_i)}{v_i - w_i} = G'(c) \geq \beta.$$

Thus,

$$\|v - w\|_\infty \leq \frac{1}{\beta} \|G(v) - G(w)\|_\infty \tag{6.22}$$

and

$$\|G(v(t)) - G(w(t))\|_\infty \leq \left(\frac{\lambda T}{\beta} + \alpha\right) \int_0^t \|G(v(s)) - G(w(s))\|_\infty ds$$

$$+ \gamma \varepsilon^2 \int_0^t \|R(s)\|_\infty ds.$$

Now apply Gronwall's inequality to get

$$\|G(v(t)) - G(w(t))\|_\infty \leq \exp\left[\left(\frac{\lambda T}{\beta} + \alpha\right) t\right] \cdot \gamma \varepsilon^2 \int_0^t \|R(s)\|_\infty ds.$$

Now, for $t \in [0, M]$, letting $\|R\|$ denote $\max_{t \in [0,M]} \|R(t)\|_\infty$,

$$\|G(v(t)) - G(w(t))\|_\infty \leq \gamma \varepsilon^2 M \|R\| \exp\left[\left(\frac{\lambda T}{\beta} + \alpha\right) M\right]$$

and using (6.22) again,

$$\begin{aligned}\|v(t) - w(t)\|_\infty &\le \frac{1}{\beta}\|G(v(t)) - G(w(t))\|_\infty \\ &\le \frac{M}{\beta}\exp\left[\left(\frac{\lambda T}{\beta} + \alpha\right)M\right]\gamma\varepsilon^2\left(C_1\varepsilon^r + C_2\left(\frac{h^m}{\varepsilon^{m+2}}\right)\right) \\ &\le C\gamma\varepsilon^2\left(\varepsilon^r + \left(\frac{h^m}{\varepsilon^{m+2}}\right)\right),\end{aligned}$$

where C depends on M and the constants α, λ, β, and T as well as (through C_1 and C_2) $\|\mu\|_{m,\infty}$, $\|\eta\|_{m,1}$, K_{r+2} and $\|v_0\|_{m,\infty}$. □

Remark 1. η, μ, v_0, are fixed functions, so making $\varepsilon \to 0$ and $\frac{h^m}{\varepsilon^{m+2}} \to 0$ (which implies $\gamma \to \infty$) gives convergence of the approximation. This amounts to letting the size of the 'visible window' shrink to zero while at the same time letting the number of neurons in the 'visible window' increase to infinity. (See Section 6.6).

Remark 2. Special cases (see Remark 3 following Theorem 6.3):

(i) If $\tau_i = \hat{\tau}$, a constant, then

$$v_t = \frac{1}{G'(v)}\left[\gamma\varepsilon^2\frac{\hat{\tau}}{2}\Delta_y\left[\mu(x,y)v(y)\right]\Big|_{(x,x)} + \gamma\tau_0\mu(x,x)v(x) - \alpha G(v)\right] \tag{6.23}$$

(ii) If $\mu \equiv 1$, then

$$v_t = \frac{1}{G'(v)}\left[\frac{1}{2}\gamma\varepsilon^2\mathrm{div}\left(\tau\nabla v\right) + \left(\gamma\tau_0 v - \alpha G(v)\right)\right].$$

(iii) If $\tau_i = \hat{\tau}$, and $\mu \equiv 1$, then

$$v_t = \frac{1}{G'(v)}\left[\gamma\varepsilon^2\frac{\hat{\tau}}{2}\Delta v + \left(\gamma\tau_0 v - \alpha G(v)\right)\right], \tag{6.24}$$

which is the equation of Cottet.

Remark 3. The initial value problem for equation (6.16) is well-posed if $\tau_i\mu(x,x) > 0$, for each i and x (recall (6.5)). Note that it is not well-posed if $\tau_i\mu(x,x) < 0$ for any i. For example, in equation (6.24) if $\hat{\tau} < 0$, we have backwards diffusion.

Remark 4. Thresholds and external inputs in the Hopfield model may be incorporated easily into the above analysis with no significant change in the proofs. If $c(x)$ is an external forcing function such that $c(x_i) = c_i$ is the external input to the i^{th} neuron in the Hopfield network equations and similarly, $\theta(x)$ is a threshold function such that $\theta(x_i) = \theta_i$, then the Hopfield network may be formulated as in (2.7) and the PDE becomes

$$v_t = \frac{1}{G'(v)} \left[\gamma \varepsilon^2 D_2 v - \alpha G(v) + \lambda \left(c(x) - \alpha \theta(x) \right) \right] . \tag{6.25}$$

Remark 5. A more biologically realistic sigmoid with range $(0, 1)$ and maximum slope not at zero can be handled by recasting the model in terms of a sigmoid with range $(-1, 1)$ and maximum slope at zero as described in Section 2.4. Thus, we can still carry out the approximation of this theorem with an extra threshold term.

Remark 6. $\mathbf{R}^n$ may be replaced by Ω throughout if appropriate boundary conditions are imposed. Then if $\|v\|_{m,\infty} \leq C \|v_0\|_{m,\infty}$ only on a bounded set, the result still holds there.

6.5 Convergence of the approximation.

Theorem 6.5 applies only where we have a bound on the solution independent of ε and γ (condition 6.20). Under what circumstances can we expect this condition to hold? We attempt to answer this question by looking at a simple one-dimensional example.

One standard way to obtain the desired bounds is by energy estimates (as described, for example, in [41]). We show what happens when we attempt to apply this technique to a simple form of equation (6.16). The basic idea is that, if the small ε is neglected, derivatives of the solution to the resulting equation can grow

exponentially. For a fixed time interval, they could be bounded except that the factor γ occurs in the growth rate. There are regions, however, where the growth is stopped or reversed.

We take $x \in \Omega \subset \mathbf{R}^1$, an interval, and D_2 given by (6.13) with $\hat{\tau} = 2$, $\tau_0 = 1$, and for simplicity, assume periodic boundary conditions. Thus, from (6.24)

$$v_t = \gamma \varepsilon^2 a(v) v_{xx} + \gamma b(v) \tag{6.26}$$

where

$$a(v) = \frac{1}{G'(v)} > 0$$

and

$$b(v) = \frac{\gamma v - \alpha G(v)}{\gamma G'(v)} .$$

For initial conditions we take

$$v(x,0) = v_0(x) \in [-v_s, v_s] ,$$

where v_s is the positive solution to $\gamma v = \alpha G(v)$ (see Fig. 6.2), and $\alpha\beta < \gamma$. (Note that if $\alpha\beta \geq \gamma$, then there is no $v_s > 0$, so all solutions collapse to 0. This is another occurrence of our low gain result from Section 2.4. Under these circumstances, condition (6.20) holds globally, but for the neural network application, we are only interested in the other case).

First recall from Section 6.4 that v itself is already bounded for $v_0(x) \in (-1,1)$. Also, $v \in [-v_s, v_s]$ if $v_0 \in [-v_s, v_s]$ by the maximum principle for this equation. Differentiating (6.26) with respect to x gives

$$w_t = \varepsilon^2 \gamma \left[a(v) w_x \right]_x + \gamma b'(v) w \tag{6.27}$$

where $w = v_x$. Now, working in L^2 (integrating over one period) with the usual inner product, denoted by $< \cdot, \cdot >$, and using integration by parts with the periodic

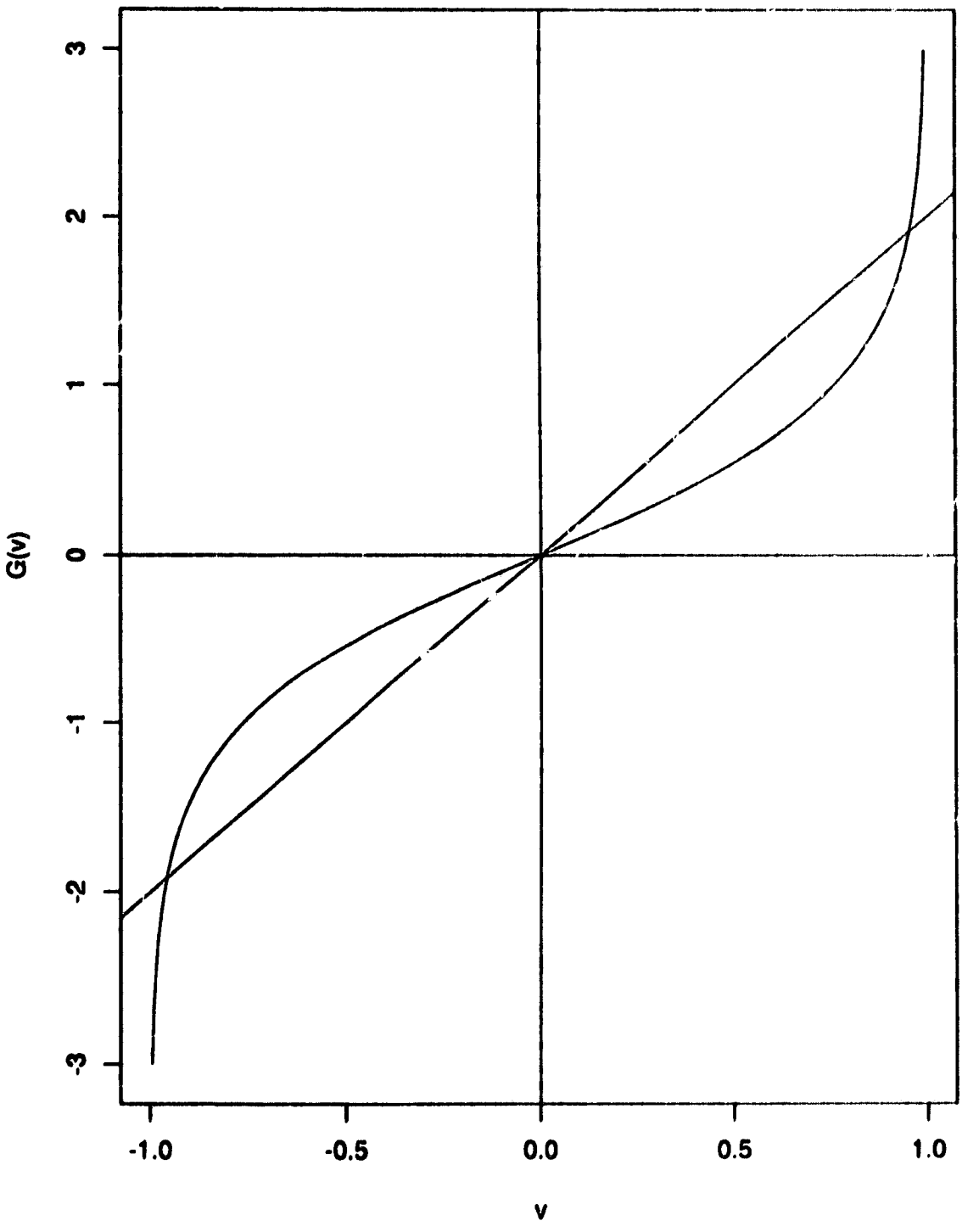

Figure 6.2 An example of the function $G(v)$ from Cottet's equation (6.26). Here, we show $G(v) = \tanh^{-1}(v)$ and $f(v) = \gamma v$, with $\gamma = 2$, showing the intersections at $\pm v_s$.

boundary conditions, we can approximate the L^2 norm of w as follows.

$$\begin{aligned}\frac{d}{dt}\|w\|_2^2 &= \frac{d}{dt}\langle w, w\rangle = 2\langle w, w_t\rangle = 2\int w\left[\varepsilon^2\gamma\left[a(v)w_x\right]_x + \gamma b'(v)w\right]dx \\ &= -2\varepsilon^2\gamma\int a(v)w_x^2\,dx + 2\gamma\int b'(v)w^2\,dx \\ &\le -2\varepsilon^2\gamma\int a(v)w_x^2\,dx + 2\gamma\max_v|b'(v)|\,\|w\|_2^2 .\end{aligned}$$

Let $B = \max_v |b'(v)|$, which exists for $v \in [-v_s, v_s]$. It is also independent of γ for γ large, since $b'(v)$ is even and decreasing on $(0, v_s]$ (see Fig. 6.3a,b) and

$$0 < b'(0) = \frac{1}{\beta} - \frac{\alpha}{\gamma} < \frac{1}{\beta};$$

$$0 > b'(v_s) = \frac{1}{G'(v_s)} - \frac{\alpha}{\gamma} > -\frac{\alpha}{\gamma} \to 0 \text{ as } \gamma \to \infty .$$

Now applying variation of constants to the above, we obtain

$$\|w\|_2^2 \le \exp(2B\gamma t)\left[\int_0^t(-2\varepsilon^2\gamma)\exp(-2B\gamma\tau)\left(\int a(v)w_x^2\,dx\right)d\tau + \|w(\cdot,0)\|_2^2\right], \tag{6.28}$$

and again throwing out the first term, which is negative, we get

$$\|w\|_2^2 \le \exp(2B\gamma t)\,\|w(\cdot,0)\|_2^2. \tag{6.29}$$

Letting $z = w_x$ and differentiating (6.27) again, we can use (6.28) to estimate $\|z\|_2^2$ in a similar way to show that

$$\|z\|_2^2 \le \exp(2B\gamma t)\left[\frac{1}{2}\max_v\left(\frac{a'(v)}{a(v)}\right)^2|w(\cdot,0)|_\infty^2\,\|w(\cdot,0)\|_2^2 + \|z(\cdot,0)\|_2^2\right]. \tag{6.30}$$

As there are some technical points involved and as it is not crucial to the discussion, we give the derivation of this inequality in Appendix 6.A at the end of this chapter. The same procedure works for still higher derivatives.

These estimates give us an upper bound on L^2 norms of v_x, v_{xx}, etc. for $t \in [0, M]$ in terms of initial data, but this bound depends also on γ. For convergence

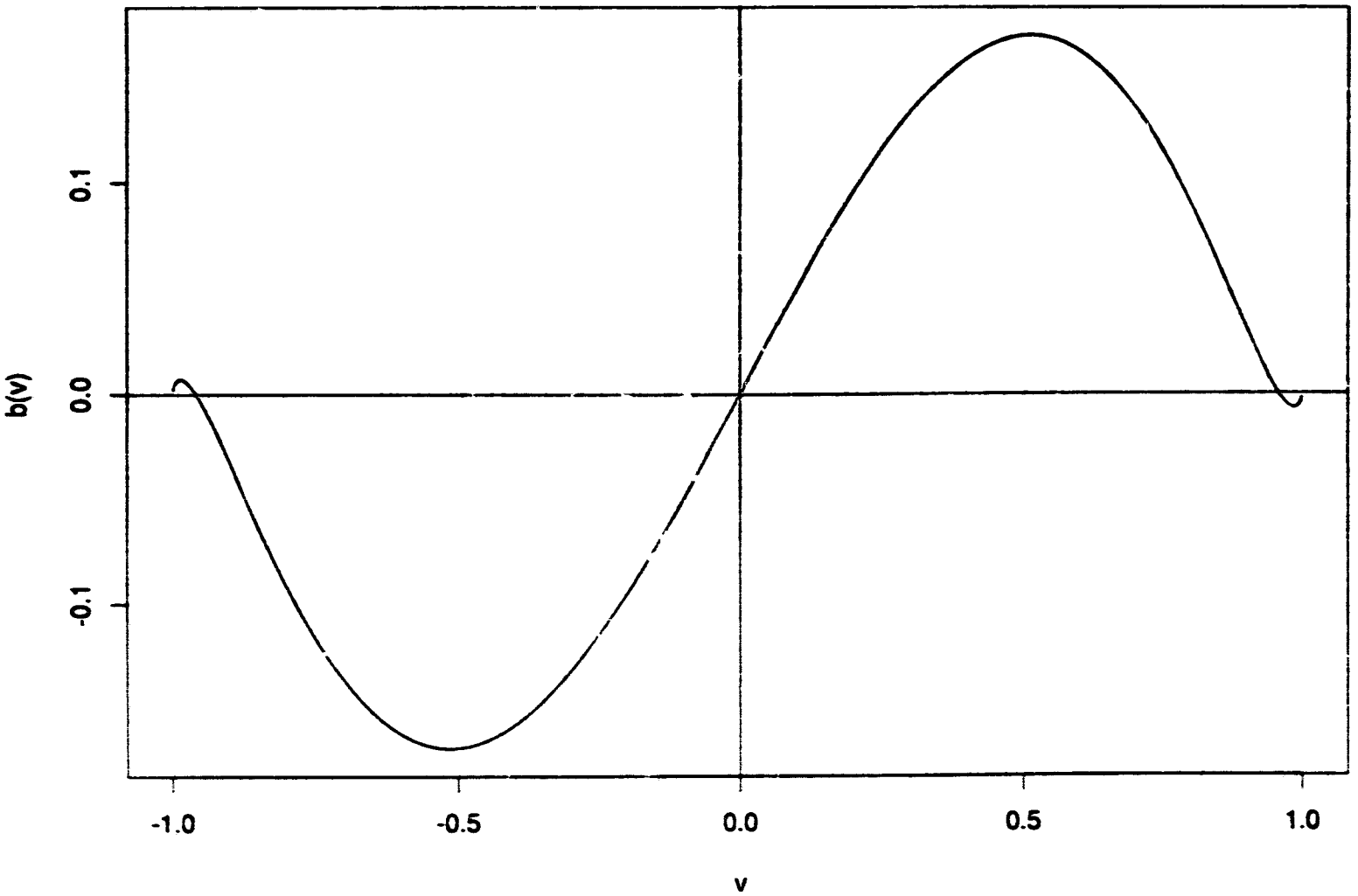

Figure 6.3a An example of the reaction term, $b(v)$, in Cottet's equation (6.26) with $\gamma = 2$, $a = 1$, and $G = \tanh^{-1}$, showing the roots at $\pm v_s$.

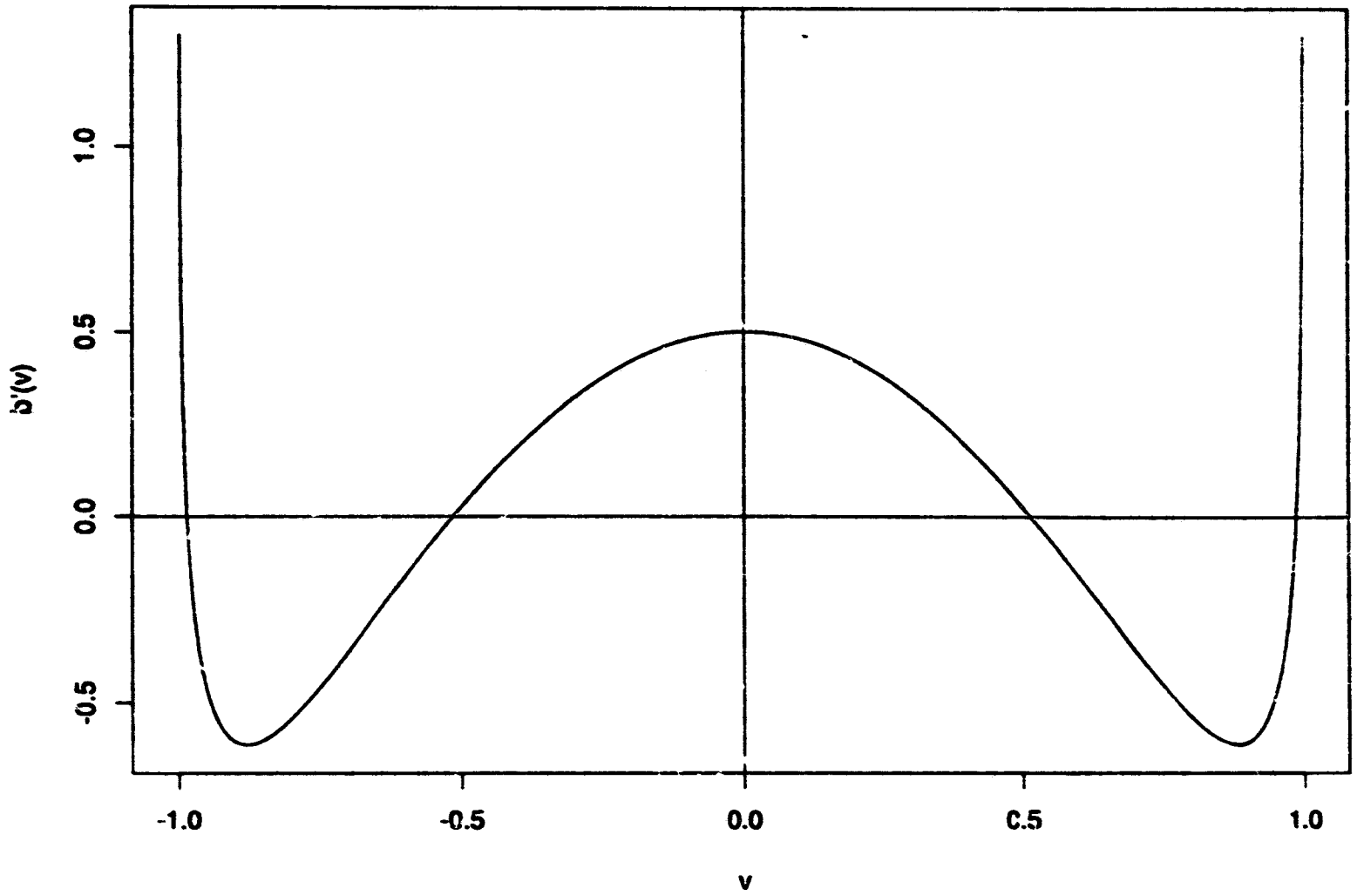

Figure 6.3b Derivative of the reaction term in Cottet's equation (6.26), i.e., $b'(v)$ for $b(v)$ as in Figure 6.3a. Note that $b'(v_s) < 0$.

of the approximation we require $\gamma \to \infty$, so this technique does not give us an appropriate upper bound on derivatives of the solution everywhere. If a derivative is unbounded in L^2 on a compact set, then of course it is also unbounded in L^∞.

However, looking again at (6.27) and neglecting the small ε term, we see that the growth rate depends on both γ and on $b'(v)$. The latter quantity has a maximum of $\frac{1}{\beta} - \frac{\alpha}{\gamma}$ at $v = 0$ and becomes negative for $|v|$ near v_s (Fig. 6.3b). Thus, for larger values of v, w actually decays. Consider, for example, an initial condition, $v_0(x)$, that is an odd function and monotone increasing on $\mathbf{R}$. Then $v(0,t) = 0$ for all t and

$$w_t(0) \simeq \left(\frac{\gamma}{\beta} - \alpha\right) w(0),$$

so $w(0)$ grows exponentially with rate depending on γ. However, for x away from 0, $|v|$ also increases since

$$v_t \simeq \gamma b(v)$$

and $b(v)$ has the sign of v for $v \in (-v_s, v_s)$ (Fig. 6.3a). Then as $|v|$ exceeds the value where $b'(v) = 0$, w_t becomes negative and v flattens out.

In fact, solutions typically look like the sketch in Fig. 6.4. Sharp transition layers form quickly and persist for a long time. The above analysis suggests that in the region of transition layers we cannot expect to get a bound on v_x or higher derivatives independent of γ. However, away from transition layers, v flattens out and approaches $\pm v_s$, so derivatives of v will be bounded.

Another approach is to attempt to obtain a time-independent bound. In the transition layers, the reaction term in (6.26) drives the magnitude of v and its derivatives up, but eventually, the small diffusion term dampens this growth. So there must be an upper bound, even in the transition layers, which holds for all time. However, such bounds must depend on ε. For example, let us look at equation (6.26)

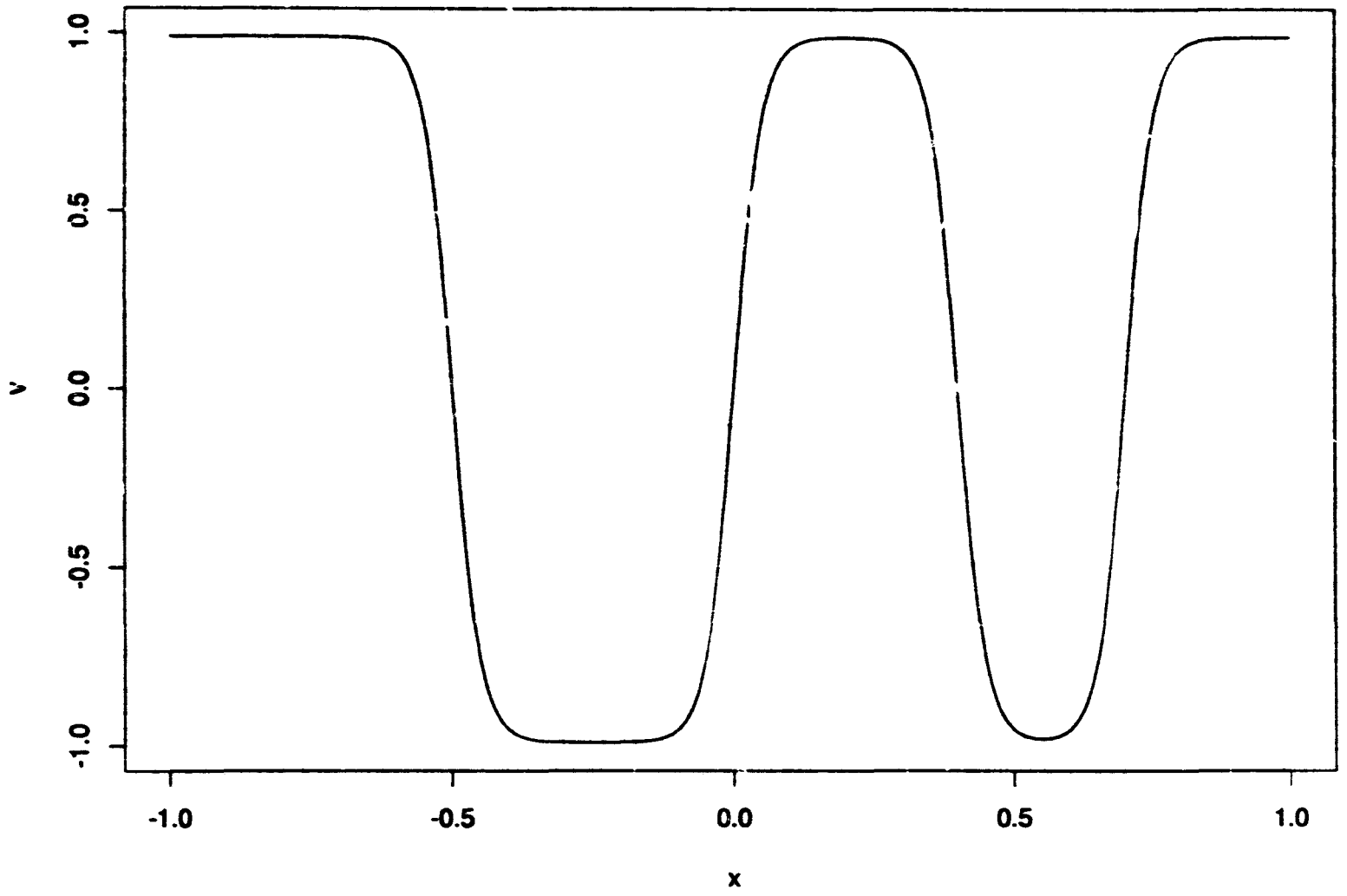

Figure 6.4 A sketch of a solution to Cottet's equation (6.26) with periodic boundary conditions, showing transition layers. This type of pattern develops quickly and then persists for a long time.

with $G(v) = \tanh^{-1}(v)$:

$$v_t = (1 - v^2)\left[\varepsilon^2 \gamma v_{xx} + \gamma v - \tanh^{-1}(v)\right]. \tag{6.31}$$

This is similar to

$$v_t = \varepsilon^2 \gamma v_{xx} + \gamma v - \tanh^{-1}(v) \tag{6.32}$$

which is an example of what is variously known as the Ginzburg-Landau equation or the Allen-Cahn equation. This type of equation has been studied by Carr and Pego [8,9,10], as well as Reyna and Ward [55] and Ward [67], for example, and it is clear that (6.31) and (6.32) have the same energy functional and the same equilibria. It is well-known that equations of this type develop transition layers between positive and negative regions that are metastable in general (see Section 7.1). The transition layers are approximated by the equilibrium solution that is monotonically increasing (or decreasing) in x (namely the heteroclinic orbit between the positive and negative solutions to $\gamma v = \tanh^{-1} v$, $-v_s$ and v_s). The norms of derivatives of solutions will be dominated by these transition layers, since elsewhere the solutions become quite flat. Thus, we expect that the norms of the transition layer in the equilibrium state will control the norms of all solutions to (6.31) and (6.32).

Consider this monotonic equilibrium, $v(x)$, solving

$$\varepsilon^2 \gamma v_{xx} + \gamma v - \tanh^{-1}(v) = 0. \tag{6.33}$$

Clearly v itself is bounded and $\|v\|_\infty = v_s$. Also, v takes on all values in $(-v_s, v_s)$ for $x \in \mathbf{R}^1$ so

$$\|v_{xx}\|_\infty = \frac{1}{\varepsilon^2} \max_v \left| \frac{1}{\gamma} \tanh^{-1}(v) - v \right| = \frac{1}{\varepsilon^2} \left| \frac{1}{\gamma} \tanh^{-1}\left(\sqrt{1 - \frac{1}{\gamma}}\right) - \sqrt{1 - \frac{1}{\gamma}} \right|.$$

We are interested in what happens as $\varepsilon \to 0$ and $\gamma \to \infty$ (the order in which these limits are taken does not matter here). It is not difficult to show that

$$\lim_{\substack{\varepsilon \to 0 \\ \gamma \to \infty}} \|v_{xx}\|_\infty = \infty$$

and that the rate of growth is $O(\varepsilon^{-2})$. In the limit as $\varepsilon \to 0$, this equilibrium approaches the discontinuous function,

$$v(x) = \begin{cases} -1 \text{ if } x < 0 \\ 0 \text{ if } x = 0 \\ 1 \text{ if } x > 0 \end{cases} .$$

Thus, $\|v_x\|_\infty$ must also grow without bound as $\varepsilon \to 0$.

This example shows that L^∞ norms of derivatives of solutions cannot in general be bounded independently of ε and t. The possibility remains that bounds exist in L^p for $p < \infty$. Then we might hope for Theorem 6.5 to hold globally in space if L^∞ norms were replaced by L^p norms. This appears not to be the case, as we see below.

We examine the L^1 norms of derivatives of the solution to (6.32) (the transition layer) which is increasing and centred at 0, on an interval $(-a, a)$. Note that if an L^1 norm is not bounded, neither is any L^p norm for $1 < p < \infty$. First,

$$\|v_x\|_1 = \int_{-a}^{a} v_x dx = v(a) - v(-a) = 2v(a) ,$$

which is constant, independent of ε and γ. However,

$$\|v_{xx}\|_1 = \int_{-a}^{a} |v_{xx}| \, dx = -2 \int_0^a v_{xx} dx = 2 \left[v_x(0) - v_x(a) \right] .$$

As $\varepsilon \to 0$, $v_x(a) \to 0$ and $v_x(0) \to \infty$ as shown above.

The order of growth of $v_x(0)$ is $O(\varepsilon^{-1})$, shown as follows. In (6.33) let

$$f(v) = \frac{1}{\gamma} \left(\tanh^{-1} v \right) - v \tag{6.34}$$

so that the equation becomes

$$\varepsilon^2 v_{xx}(x) = f\left(v(x)\right) . \tag{6.35}$$

Let $x = \varepsilon y$ and define

$$w(y) = v(\varepsilon y) = v(x)\,. \tag{6.36}$$

Then

$$w_y(y) = \varepsilon v_x(\varepsilon y)\,,$$

$$w_{yy}(y) = \varepsilon^2 v_{xx}(\varepsilon y)$$

and so

$$w_{yy}(y) = f\left(v(\varepsilon y)\right) = f\left(w(y)\right)\,. \tag{6.37}$$

This implies that w is independent of ε (γ is not critical here, see Appendix 6.B at the end of the chapter).

Letting $x = \varepsilon y$,

$$\begin{aligned} v_x(s) &= \int_{-\infty}^{s} v_{xx}(x)\,dx \\ &= \varepsilon \int_{-\infty}^{\frac{s}{\varepsilon}} v_{xx}(\varepsilon y)\,dy \\ &= \frac{1}{\varepsilon}\int_{-\infty}^{\frac{s}{\varepsilon}} w_{yy}(y)\,dy \\ &= \frac{1}{\varepsilon}\int_{-\infty}^{\frac{s}{\varepsilon}} f\left(w(y)\right)\,dy\,. \end{aligned}$$

Since we know that v_x is positive and takes its supremum at 0,

$$\|v_x\|_\infty = v_x(0) = \frac{1}{\varepsilon}\int_{-\infty}^{0} f\left(w(y)\right)\,dy = \frac{c}{\varepsilon}\,,$$

for some $c > 0$, since w is independent of ε.

We have already shown that $\|v_{xx}\|_1 = 2\left[v_x(0) - v_x(a)\right]$ and $v_x(a) \to 0$ as $\varepsilon \to 0$, so $\|v_{xx}\|_1 = O\left(\frac{1}{\varepsilon}\right)$ also. Similarly, denoting the value of x where v_{xx} or $f\left(v(x)\right)$

takes its minimum as b,

$$\begin{aligned}\|v_{xxx}\|_1 &= 2\int_0^a |v_{xxx}|\,dx = 2\int_b^a v_{xxx}dx - 2\int_0^b v_{xxx}dx \\ &= 2\left[v_{xx}(a) - v_{xx}(b)\right] - 2\left[v_{xx}(b) - v_{xx}(0)\right] \\ &= 2v_{xx}(a) - 4v_{xx}(b) \\ &= 2v_{xx}(a) - 4\|v_{xx}\|_\infty.\end{aligned}$$

Since $\|v_{xx}\|_\infty = O\left(\frac{1}{\varepsilon^2}\right)$, so is $\|v_{xxx}\|_1$.

The approximation error for the diffusion operator in Cottet's equation is $O\left(\varepsilon^r + \varepsilon^{-2}\gamma^{-\frac{m}{n}}\right)$ if the solution is bounded in $W^{m,\infty}$ by initial conditions. If not and we try to use the above time-independent bounds on the solution, we lose three orders of ε by demanding that three derivatives of v be in L^∞ (and in general n orders of ε for $v \in W^{m,\infty}$). Recall that Theorem 6.5 requires $m \geq r+2$ and $r \geq 1$, so r is at most $m-2$, but even then the approximation becomes $\varepsilon^{-m} \cdot O\left(\varepsilon^{m-2} + \varepsilon^{-2}\gamma^{-\frac{m}{n}}\right) = O\left(\varepsilon^{-2} + \varepsilon^{-2-m}\gamma^{-\frac{m}{n}}\right)$ and convergence is lost. Even if we could do the approximations in $W^{m,1}$, we only save one order of ε and convergence is still lost.

Thus, in the transition layers, time-independent bounds depend too strongly on ε, and energy estimates only give us a bound if γt is bounded, whereas for convergence we require $\varepsilon \to 0$ and $\gamma \to \infty$. Nevertheless, away from transition layers, we expect condition (6.20) to hold and therefore Theorem 6.5 as well. Moreover, numerical simulations appear to confirm that the approximation does hold reasonably well.

We speculate that the approximation is accurate enough for the application to neural networks. The approximation holds in regions where the solution's derivatives do not grow, i.e. away from transition layers. Typically, we know that steep transition layers form quickly and then move only extremely slightly for a long time.

If the transition layers are slightly offset by the approximation (a shift of order ε, say), it is not of critical importance for the neural network dynamics; positive and negative regions are still virtually the same. The transition layers that form in the evolution of the reaction-diffusion equation should thus closely model the corresponding neural network dynamics despite the lack of rigorous convergence in the entire domain. A rigorous analysis of this question would require properly accounting for boundary conditions. After a long period of time, when the layers start to collapse, we might no longer expect to have a close approximation, of course.

Since the transition layers become sharper as $\varepsilon \to 0$ and $\gamma \to \infty$, it is possible that the approximation converges in a weaker sense such as, for example,

$$\text{meas}\{x : \exists t, \alpha, |\alpha| \leq m \text{ such that } |D_x^\alpha v(x,t)| \geq c\} \to 0 \text{ as } \varepsilon \to 0.$$

We can at least show that transition layers will exist on domains of size $O(\varepsilon)$ (see Appendix 6.B at the end of the chapter).

If we consider equation (6.16) in its general form, the above conclusions should be substantially the same. The presence of the function μ in (6.16) will modify the solution, but away from transition layers solutions should still be controlled by the norm of μ in $W^{m,\infty}(\mathbf{R}^n \times \mathbf{R}^n)$.

To summarize, we expect that the reaction-diffusion equation will be good at defining regions of high and low activity in the network and at approximating the activity in these regions, but may not accurately model the structure of the transition layers themselves.

Finally, we comment that the difficulty in convergence is in the error term from the quadrature. Therefore, if we are content to approximate an integro-differential equation model by a reaction-diffusion equation, the convergence can be carried through without any problem. In the proof of Theorem 6.5, if we stop after the first

approximation we get (see (6.21))

$$v_t = \frac{1}{G'(v)}\left[\gamma\varepsilon^2 I^\varepsilon v - \alpha G(v)\right] + \frac{\gamma\varepsilon^2}{G'(v)}O(\varepsilon^r),$$

which is

$$v_t = \frac{1}{G'(v)}\left[\gamma\varepsilon^2 \int T^\varepsilon(x,y)v(y)\,dy - \alpha G(v)\right] + \frac{\gamma\varepsilon^2}{G'(v)}O(\varepsilon^r).$$

Now write

$$T(x,y) = \varepsilon^{2+n}T^\varepsilon(x,y) = \mu(x,y)\eta\left(\frac{x-y}{\varepsilon}\right)$$

so that $T(x_i,x_j) = T_{ij}$ corresponds to the connection matrix. If we replace T^ε by T in the equation, the coefficient becomes

$$\frac{\gamma\varepsilon^2}{\varepsilon^{2+n}} = \gamma\varepsilon^{-n} = \lambda\left(\frac{\varepsilon}{h}\right)^n\varepsilon^{-n} = \lambda h^{-n},$$

(recall the definition of γ from (6.17)) and the equation is

$$v_t = \frac{1}{G'(v)}\left[\lambda h^{-n}\int T(x,y)v(y)\,dy - \alpha G(v)\right] + \frac{\gamma\varepsilon^2}{G'(v)}O(\varepsilon^r).$$

The h^{-n} is there to balance the h^n term coming from quadrature of the integral, $\int Tv \sim h^n\sum T_{ij}v_j$. If we wish to approximate only the integro-differential equation model and not do the quadrature, then we don't want the h^{-n} term in front of the integral. Thus, the corresponding reaction-diffusion equation will be as before but with its diffusion coefficient multiplied by h^n. In other words, the reaction diffusion equation

$$v_t = \frac{1}{G'(v)}\left[\lambda\varepsilon^{2+n}D_2 v - \alpha G(v)\right]$$

is approximated as

$$v_t = \frac{1}{G'(v)}\left[\lambda\int T(x,y)v(y)\,dy - \alpha G(v)\right] + \frac{\lambda\varepsilon^{2+n}}{G'(v)}O(\varepsilon^r).$$

Now Theorem 6.5 can be carried through essentially as before, resulting in an error term of $\lambda\varepsilon^{2+n}O(\varepsilon^r)$. As $\varepsilon \to 0$, the approximation converges.

6.6 Discussion.

Approximating a system of many ordinary differential equations (enormous systems for real biological brains) by a single partial differential equation has the advantages of allowing us to study a simpler model (even though we have gone from a finite dimensional to an infinite dimensional system) and allowing us to bring a different body of theory to bear on the problem. It is similar in spirit to the approximation of gas dynamics by statistical equations and thermodynamics where it is too difficult to keep track of vast numbers of individual particles. We have mentioned in Chapters 3 and 5 the arguments that the level of averaged neural activity is more significant for behaviour than the level of the individual neurons. It is possible that there is something crucial about the behaviour of neural nets at the level of individual neurons, at least for some purposes, but this itself may become clear from a study of their collective behaviour with an understanding of the relationship between the collective and individual levels.

What do the restrictions on the function T^ε in Theorem 6.5 mean? Also, what exactly does the approximation mean when, as we let ε get smaller and γ get larger (i.e. $\frac{h}{\varepsilon} \to 0$ since $\gamma = \lambda \frac{\varepsilon^n}{h^n}$), the parameters of the PDE change and also the connection matrix of the Hopfield net changes? In fact, we cannot decide for a Hopfield net with a specific connection matrix whether it is a good approximation to a PDE. However, Theorem 6.5 does give a general idea of what the matrices for nets approximating PDEs will be like.

We must have ε small, h small and $\frac{\varepsilon^{m+2}}{h^m}$ large. We also want $\eta \in L^1$ so we could take, for example,

$$\text{supp}(\eta) \subset [-1,1]^n,$$

so that η_ε, the cutoff function has

$$\text{supp}(\eta_\varepsilon) \subset [-\varepsilon, \varepsilon]^n.$$

Thus, $T^\varepsilon(x, y)$ and

$$T_{ij} = \mu(x_i, x_j)\eta\left(\frac{x_i - x_j}{\varepsilon}\right)$$

effectively sample μ near its 'diagonal', i.e. near (x, x). We can think of this as a 'visible window' of size 2ε in each direction around each neuron such that other neurons within its window are connected to it but neurons outside its window are not. The spacing between the neurons is determined by the step size used in the quadrature, h, so the number of neurons in a visible window will be

$$N = \left(\frac{2\varepsilon}{h}\right)^n.$$

In the case of one space dimension ($n = 1$), the connection matrix T will be large (h small), will have a band about the diagonal (visible window) which is wide ($\frac{\varepsilon}{h}$ large) but narrow in relation to all of T (ε small). Also, the entries of the matrix will not fluctuate wildly; i.e. an entry will not differ too radically from neighbouring entries (as a result of the smoothness requirements on η and μ). Finally, in order to obtain a well-posed PDE, we require T^ε to be predominantly positive ($\mu\tau_i > 0$). Note that specific connections, T_{ij}, may still be negative.

For purposes of illustration, let us look at an example of T^ε that satisfies the conditions for Theorem 6.5, and one that does not.

Example 6.1 Let

$$\eta(x) = \begin{cases} \frac{15}{4}(x^2 - 1)(1 - 7x^2) \text{ if } x \in [-1, 1] \\ 0 \text{ otherwise,} \end{cases}$$

so that

$$\tau_0 = \int_{-1}^{1} \eta(x)\, dx = \frac{15}{4}\int_{-1}^{1} (-7x^4 + 8x^2 - 1)\, dx = \frac{15}{2}\left[-\frac{7}{5} + \frac{8}{3} - 1\right] = 2,$$

and

$$\hat{\tau} = \int_{-1}^{1} x^2 \eta(x)\, dx = \frac{15}{4} \int_{-1}^{1} (-7x^6 + 8x^4 - x^2)\, dx = \frac{15}{2}\left[-1 + \frac{8}{5} - \frac{1}{3}\right] = -2.$$

Now let $\mu(x,y) = \frac{x^2}{2} + y^2$ on $[-1,1]$ so that

$$T^\varepsilon(x,y) = \begin{cases} \dfrac{15}{8\varepsilon^3}(x^2+2y^2)\left(\left(\dfrac{x-y}{\varepsilon}\right)^2 - 1\right)\left(1 - 7\left(\dfrac{x-y}{\varepsilon}\right)^2\right) & \text{if } |x-y| < \varepsilon \\ 0 \text{ otherwise.} \end{cases}$$

Note that in this case, T^ε is not symmetric or translation-invariant and that it is negative for $|x-y| < \frac{\varepsilon}{\sqrt{7}}$. Now $T_{ij} = \varepsilon^3 T^\varepsilon(x_i, x_j)$, where $x_i = ih$ and $x_j = jh$. If $\varepsilon = 0.5$ and $h = 0.12$, for example, then the matrix T is as in Fig. 6.5. Of course, for a good approximation, we will most likely need ε and h much smaller, but this illustrates the general appearance of the matrix. □

Example 6.2 If, on the other hand, $\mu(x,y) = y \sin\left(\frac{1}{y}\right)$, then its derivatives are not bounded and no matter how fine the discretization gets (i.e. no matter how small h gets) there will always be large oscillations in the matrix entries near the diagonal. □

Our approximation theorem, then, gives a degree of classification of Hopfield nets, some having PDE-like behaviour, others not. The restriction to banded matrices with wide bands is reasonable from the biological point of view as, typically, a neuron will be connected to many other neurons but mainly those in its vicinity.

Note that in the PDE obtained from the network, equation (6.16), the connection matrix has been 'squeezed' into the diagonal of μ, namely $\mu(x,x)$ itself and the first and second $y-$derivatives of $\mu(x,y)$ evaluated at (x,x), and into the parameters τ_0 and τ_i. In this sense, at least, the PDE-like nets are simpler than general ones. In the case studied by Cottet ($\mu \equiv 1$, $\tau_i = \hat{\tau}$), all the information of the connection matrix is squeezed into the two parameters, τ_0 and $\hat{\tau}$. In one space dimension, the matrix T has constant diagonals in Cottet's case (this is the translation invariance).

$$T = \begin{bmatrix}
-5.18 & -2.46 & 1.73 & 3.90 & 1.11 & & & & & & & & & & & \\
-2.69 & -3.97 & -1.84 & 1.26 & 2.77 & 0.77 & & & & & & & & & & \\
2.09 & -2.03 & -2.92 & -1.31 & 0.87 & 1.85 & 0.51 & & & & & & & & & \\
5.23 & 1.57 & -1.47 & -2.02 & -0.87 & 0.55 & 1.13 & 0.31 & & & & & & & & \\
1.66 & 3.90 & 1.12 & -1.00 & -1.30 & -0.52 & 0.31 & 0.62 & 0.18 & & & & & & & \\
 & 1.23 & 2.77 & 0.75 & -0.62 & -0.73 & -0.26 & 0.14 & 0.31 & 0.13 & & & & & & \\
 & & 0.88 & 1.85 & 0.46 & -0.33 & -0.32 & -0.09 & 0.05 & 0.21 & 0 14 & & & & & \\
 & & & 0.59 & 1.13 & 0.24 & -0.14 & -0.08 & -0.02 & 0.04 & 0.31 & 0.22 & & & & \\
 & & & & 0.37 & 0.62 & 0.10 & -0.03 & 0.00 & -0.03 & 0.10 & 0 62 & 0.37 & & & \\
 & & & & & 0.22 & 0.31 & 0.04 & -0.02 & -0.08 & -0.14 & 0.24 & 1.13 & 0.59 & & \\
 & & & & & & 0.14 & 0.21 & 0.05 & -0.09 & -0.32 & -0.33 & 0.46 & 1.85 & 0.88 & \\
 & & & & & & & 0.13 & 0.31 & 0.14 & -0.26 & -0.73 & -0.62 & 0.75 & 2.77 & 1.23 \\
 & & & & & & & & 0.18 & 0.62 & 0.31 & -0.52 & -1.30 & -1.00 & 1.12 & 3.90 \\
 & & & & & & & & & 0.31 & 1.13 & 0.55 & -0.87 & -2.02 & -1.47 & 1.57 \\
 & & & & & & & & & & 0.51 & 1.85 & 0.87 & -1.31 & -2.92 & -2.03 \\
 & & & & & & & & & & & 0.77 & 2.77 & 1.26 & -1.84 & -3.97 \\
 & & & & & & & & & & & & 1.11 & 3.90 & 1.73 & -2.46
\end{bmatrix}$$

Figure 6.5 An example of a connection matrix of the type covered by Theorem 6.5. Here, $\eta(x) = \frac{15}{4}(x^2 - 1)(1 - 7x^2)$ on $[-1, 1]$, $\mu(x, y) = \frac{x^2}{2} + y^2$ on $[-1, 1]^2$, $\varepsilon = 0.5$ and $h = 0.12$. (Blank entries are 0).

Another level of classification is obtained by this method. Any Hopfield net with solutions closely approximating those of a PDE also closely approximate solutions to standard discretizations of the PDE, as long as appropriate discretizations are used. These standard discretizations are, of course, also Hopfield nets of a very simple type, so that all Hopfield nets that have the same PDE approximation form a class with similar behaviour and there is at least one very simple representative of each class. (These are studied further in Chapter 8).

Learning in the conventional Hopfield net consists in setting or modifying the connection matrix, T. There is still some room for making analogous modifications directly in the PDE formulation, by altering $\mu(x,x)$ and the y-derivatives of μ at (x,x) and the parameters τ_i and τ_0, though the range of possibilities for modification may be smaller. In Cottet's simpler case only the two parameters τ_0 and $\hat{\tau}$ are available for modification.

6.A Appendix on bounds for the second derivative of the solution.

We wish to prove inequality (6.30). First, we need a maximum principle.

Lemma 6.6 *For equation* (6.26) *with* $w = v_x$,

$$|w|_\infty^2 \leq \exp(-2B\gamma t)|w(\cdot,0)|_\infty^2 ,$$

where $B = \max_v |b'(v)|$.

Proof Let $\tilde{w} = e^{-kt}w$. So

$$\tilde{w}_t = \varepsilon^2\gamma[a(v)]_x\tilde{w}_x + \varepsilon^2\gamma a(v)\tilde{w}_{xx} + \gamma b'(v)\tilde{w} - k\tilde{w} .$$

At a maximum of $\tilde{w}(x,t)$ for $x \in \mathbf{R}$ in a time interval, $(0,M]$, with $\tilde{w} > 0$ we have $\tilde{w}_t \geq 0$ ($\tilde{w}_t > 0$ can only occur at $t = M$), $\tilde{w}_x = 0$ and $\tilde{w}_{xx} \leq 0$, which will give us

a contradiction if $\gamma b'(v) - k < 0$ for all v, i.e. $k > \gamma B$. We are at liberty to pick such a k, so $\tilde{w}$ must achieve its maximum at $t = 0$, and

$$|w(\cdot,t)|_\infty \leq \exp(-B\gamma t)|w(\cdot,0)|\,,$$

giving our result. □

Proposition 6.7 *For equation* (6.26) *with* $w = v_x$, $z = w_x$, *inequality* (6.30) *holds.*

Proof We will need another estimate obtained from (6.28),

$$2\varepsilon^2\gamma \int_0^t \exp\left(-2B\gamma\tau\right)\left(\int a(v)w_x^2\,dx\right)\,d\tau \leq \|w(\cdot,0)\|_2^2 - \exp\left(-2B\gamma t\right)\|w\|_2^2\,,$$

and so

$$\int_0^t \exp\left(-2B\gamma\tau\right)\left(\int a(v)w_x^2\,dx\right)\,d\tau \leq \frac{1}{2\varepsilon^2\gamma}\|w(\cdot,0)\|_2^2\,. \tag{6.38}$$

Differentiating (6.27) with respect to x gives

$$z_t = \varepsilon^2\gamma\left[a(v)z\right]_{xx} + \gamma\left[b'(v)w\right]_x\,.$$

Thus,

$$\begin{aligned}\frac{d}{dt}\|z\|^2 &= 2\int z\left(\varepsilon^2\gamma\left[a(v)z\right]_{xx} + \gamma\left[b'(v)w\right]_x\right)\,dx\\ &= -2\int z_x\varepsilon^2\gamma\left[a(v)z\right]_x\,dx + 2\gamma\int\left[b'(v)w\right]_x z\,dx\\ &= -2\varepsilon^2\gamma\int z_x\left[(a(v))_x z + a(v)z_x\right]\,dx + 2\gamma\int\left(b''(v)w^2 z + b'(v)z^2\right)\,dx\,.\end{aligned}$$

Now, we complete the square in the first integral.

$$2[a(v)]_x z z_x + a(v)z_x^2 = \left[\sqrt{a(v)}z_x + \frac{[a(v)]_x}{\sqrt{a(v)}}z\right]^2 - \frac{[a(v)]_x^2}{a(v)}z^2\,.$$

Since we have the negative of the expression on the left here, we can simply throw away the term in brackets on the right to get

$$\begin{aligned}\frac{d}{dt}\|z\|^2 &\le \varepsilon^2\gamma\int\frac{[a(v)]_x^2}{a(v)}z^2\,dx - \varepsilon^2\gamma\int a(v)z_x^2\,dx + 2\gamma\int\left(b''(v)w^2z + b'(v)z^2\right)\,dx\\ &\le \varepsilon^2\gamma\max_v\left(\frac{a'(v)}{a(v)}\right)^2\int|w|_\infty^2 a(v)z^2\,dx + 2\gamma\max_v(b''(v))\int w^2z\,dx\\ &\quad + 2B\gamma\int z^2\,dx\,.\end{aligned}$$

Now, we apply the lemma in the first integral and note that

$$\int w^2w_x\,dx = -\int(2ww_x)w\,dx = -2\int w^2w_x\,dx$$

implies that the second integral is 0. Thus,

$$\frac{d}{dt}\|z\|^2 \le \varepsilon^2\gamma\max_v\left(\frac{a'(v)}{a(v)}\right)^2\exp(-2B\gamma t)|w(\cdot,0)|_\infty^2\int a(v)z^2\,dx + 2B\gamma\|z\|^2\,.$$

Now we are in a position to use variation of constants to get

$$\begin{aligned}\|z\|^2 &\le \exp(2B\gamma t)\\ &\quad\times\left[\int_0^t\varepsilon^2\gamma\max_v\left(\frac{a'(v)}{a(v)}\right)^2|w(\cdot,0)|_\infty^2\exp(-4B\gamma s)\int a(v)z^2\,dx\,ds + \|z(\cdot,0)\|^2\right]\end{aligned}$$

and from (6.38), this gives

$$\begin{aligned}\|z\|^2 &\le \exp(2B\gamma t)\left[\varepsilon^2\gamma\max_v\left(\frac{a'(v)}{a(v)}\right)^2|w(\cdot,0)|_\infty^2\frac{\|w(\cdot,0)\|^2}{2\varepsilon^2\gamma} + \|z(\cdot,0)\|^2\right]\\ &\le \exp(2B\gamma t)\left[\max_v\left(\frac{a'(v)}{a(v)}\right)^2|w(\cdot,0)|_\infty^2\frac{\|w(\cdot,0)\|^2}{2} + \|z(\cdot,0)\|^2\right]\end{aligned}$$

as desired. □

6.B Appendix on width of transition layers.

We wish to show that the transition layer of the monotone increasing equilibrium (with $v(0) = 0$) is of width $O(\varepsilon)$. We will measure the width of the transition layer by the location of the point where the equilibrium solution v to equation (6.33) or (6.35) has maximum concavity (the 'sharp corners' of the layer). Let this point be denoted by $x^* > 0$, so that $|v_{xx}|$ has a maximum at x^*. We show here that $x^* = O(\varepsilon)$ as $\varepsilon \to 0$ and $\gamma \to \infty$. First, we show that the corresponding point y^* for equation (6.37) is $O(1)$ as $\gamma \to \infty$.

Lemma 6.8 *Let w satisfy equation (6.37) where $f(v) = \frac{1}{\gamma}\tanh^{-1} v$ as in (6.34). Let y^* be the positive solution to $w_{yyy}(y) = 0$, i.e. to $f'(w(y)) = 0$. Then $y^* = O(1)$ as $\gamma \to \infty$. Also, $w_y(0) = O(1)$.*

Proof First, note that $|f|$ has maximum value

$$f_{max} = \sqrt{1 - \frac{1}{\gamma}} - \frac{1}{\gamma}\tanh^{-1}\left(\sqrt{1 - \frac{1}{\gamma}}\right) \tag{6.39}$$

because $f'(w) = 0$ implies $\frac{1}{\gamma(1-w^2)} - 1 = 0$ so that $w = \pm\sqrt{1 - \frac{1}{\gamma}}$. Thus, $w(y^*) = \sqrt{1 - \frac{1}{\gamma}}$, and $f_{max} = f\left(-\sqrt{1 - \frac{1}{\gamma}}\right)$. This also shows that $f_{max} = O(1)$ as $\gamma \to \infty$ since

$$\begin{aligned}
\lim_{\gamma\to\infty} f_{max} &= \lim_{\gamma\to\infty}\sqrt{1 - \frac{1}{\gamma}} - \lim_{\gamma\to\infty}\frac{\tanh^{-1}\left(\sqrt{1 - \frac{1}{\gamma}}\right)}{\gamma} \\
&= 1 - \lim_{\gamma\to\infty}\frac{1}{1 - \left(1 - \frac{1}{\gamma}\right)} \cdot \frac{1}{2}\left(1 - \frac{1}{\gamma}\right)^{-\frac{1}{2}}\gamma^{-2} \\
&= 1 - \lim_{\gamma\to\infty}\frac{1}{2\gamma\sqrt{1 - \frac{1}{\gamma}}} = 1.
\end{aligned}$$

Now, for $y > 0$, we have $w_{yy} > -f_{max} > -1$, so for any $s > 0$,

$$w_y(s) - w_y(0) = \int_0^s w_{yy}(y)\,dy > \int_0^s (-1)\,dy = -s,$$

and

$$w_y(s) > w_y(0) - s.$$

Furthermore, for any $r > 0$,

$$w(r) = \int_0^r w_y(s)\,ds > \int_0^r [w_y(0) - s]\,ds = w_y(0)r - \frac{r^2}{2}.$$

Since $|w(y)| < 1$, we have

$$1 > w(r) > w_y(0)r - \frac{r^2}{2} \text{ for all } r > 0.$$

In particular,

$$\sup_{r>0}\left[w_y(0)r - \frac{r^2}{2}\right] < 1.$$

Since the quantity in brackets above has a maximum when $r = w_y(0)$, we have

$$w_y^2(0) - \frac{1}{2}w_y^2(0) = \frac{1}{2}w_y^2(0) < 1,$$

and so

$$w_y^2(0) < 2\,. \tag{6.40}$$

This upper bound on $w_y(0)$ also gives us a lower bound on y^*, since

$$w_y(0) > \frac{w(y^*)}{y^*} = \frac{\sqrt{1 - \frac{1}{\gamma}}}{y^*},$$

giving

$$y^* > \frac{\sqrt{1 - \frac{1}{\gamma}}}{w_y(0)} > \frac{\sqrt{1 - \frac{1}{\gamma}}}{\sqrt{2}}\,. \tag{6.41}$$

However, we get a better lower bound as well as an upper bound below. Looking at Fig. 6.6, it is evident that the straight line from the origin to the point $(y^*, f(w(y^*)))$

$$\ell(y) = \frac{-f_{max}}{y^*}y\,, \tag{6.42}$$

lies above the function $f(w(y))$ on the interval $(0, y^*)$. In fact, for y in this interval, we have $w \in \left(0, \sqrt{1 - 1/\gamma}\right)$ since w is increasing, and also $f''(w) > 0$, $f'(w) < 0$ and finally $w_{yy}(y) = f(w(y)) < 0$, so $\ell(y) > f(w(y))$ for $0 < y < y^*$.

Therefore,

$$w_y(s) - w_y(0) = \int_0^s f(w(y))\,dy < \int_0^s \frac{-f_{max}}{y^*}\,y\,dy = \frac{-f_{max}}{y^*}\,\frac{s^2}{2},$$

for $0 < s < y^*$. I.e.

$$w_y(s) < w_y(0) - \frac{1}{2}\frac{f_{max}}{y^*}s^2.$$

Now, $w(r) = \int_0^r w_y(s)\,ds$, so

$$w(y^*) = \int_0^{y^*} \left[w_y(0) - \frac{1}{2}\frac{f_{max}}{y^*}s^2\right] ds = w_y(0)y^* - \frac{1}{2}\frac{f_{max}}{y^*}\frac{(y^*)^3}{3},$$

$$\sqrt{1 - \frac{1}{\gamma}} < w_y(0)y^* - \frac{f_{max}}{6}(y^*)^2$$

or

$$\frac{f_{max}}{6}(y^*)^2 - w_y(0)y^* + \sqrt{1 - \frac{1}{\gamma}} < 0. \tag{6.43}$$

Since $f_{max} > 0$, this quadratic is positive when $|y^*|$ is large. In order to make it negative, it must have real roots, i.e. values of y such that

$$y = \frac{3}{f_{max}}\left[w_y(0) \pm \sqrt{w_y^2(0) - \frac{2}{3}f_{max}\sqrt{1 - \frac{1}{\gamma}}}\right], \tag{6.44}$$

and these only exist when

$$w_y^2(0) > \frac{2}{3}f_{max}\sqrt{1 - \frac{1}{\gamma}}. \tag{6.45}$$

This condition must be satisfied or the quadratic inequality (6.43) is never true, contradicting the existence of y^*. Combining (6.40) and (6.45), we get

$$\frac{2}{3}f_{max}\sqrt{1 - \frac{1}{\gamma}} < w_y^2(0) < 2. \tag{6.46}$$

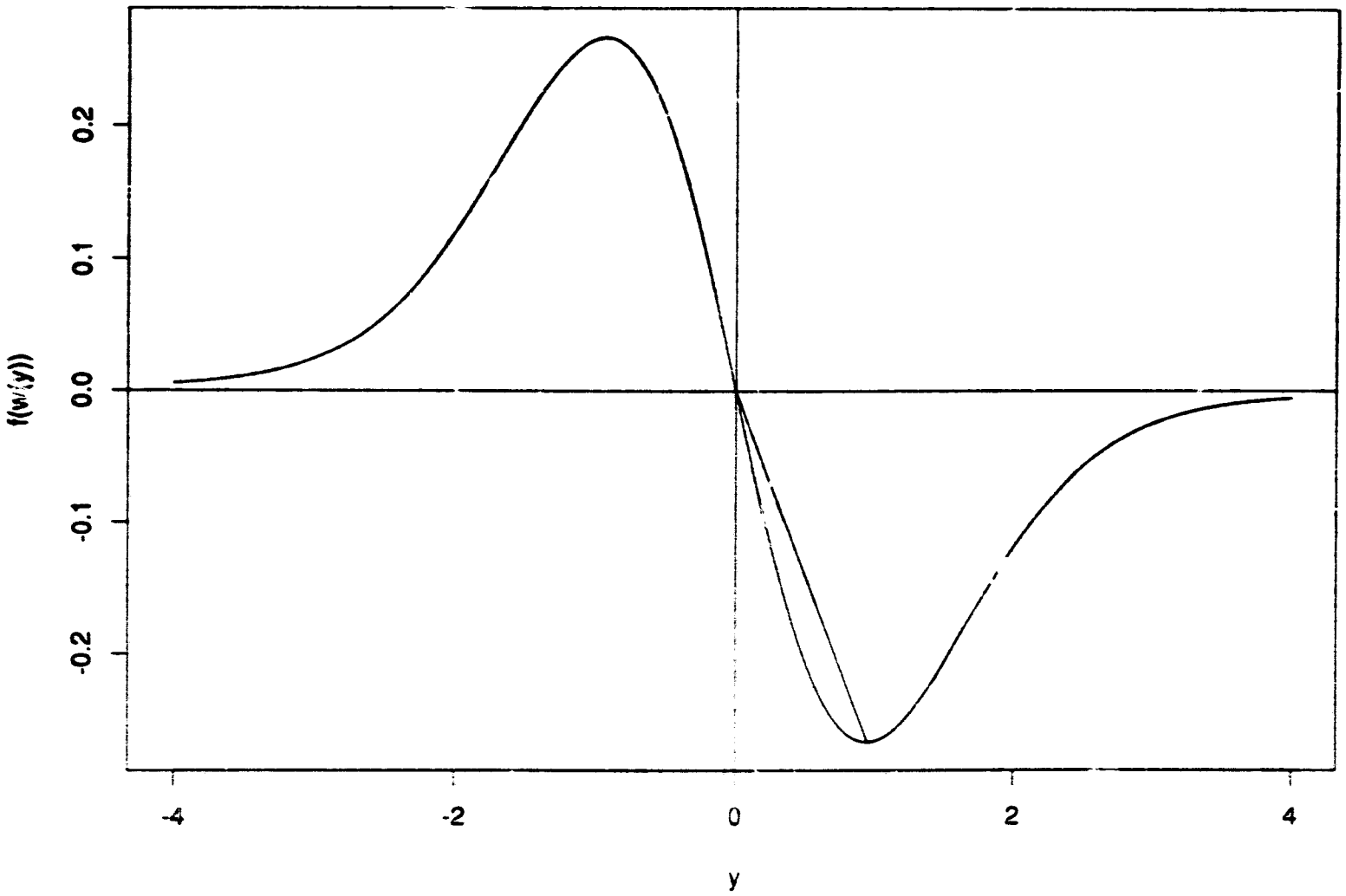

Figure 6.6 A sketch of $f(w, y))$, from equation (6.37), for use in energy estimates. The line from the origin to $(y^*, f(w(y^*)))$ is also shown.

As $\gamma \to \infty$, the left hand bound $\to \frac{2}{3}$ so clearly $w_y(0) = O(1)$. We also have from (6.44)

$$\frac{f_{max}}{3} y^* \in \left(A - \sqrt{A^2 - B}, A + \sqrt{A^2 - B} \right), \tag{6.47}$$

where $A = w_y(0)$ and $B = \frac{2}{3} f_{max} \sqrt{1 - \frac{1}{\gamma}}$. The size of this range depends on $w_y(0)$, i.e. on A in (6.47). Without determining the value of $w_y(0)$ exactly, we can find a range larger than that in (6.47) by noting that as A increases, the range widens, as follows.

$$\frac{d}{dA}\left[A - \sqrt{A^2 - B}\right] = 1 - \frac{1}{2}\left(A^2 - B\right)(2A) = 1 - \frac{A}{\sqrt{A^2 - B}} < 0,$$

since $B > 0 \Rightarrow \frac{A}{\sqrt{A^2 - B}} > 1$. Also

$$\frac{d}{dA}\left[A + \sqrt{A^2 - B}\right] = 1 + \frac{A}{\sqrt{A^2 - B}} > 0.$$

We have (6.40) that $w_y(0) < \sqrt{2}$, so we can be sure that

$$\frac{f_{max}}{3} y^* \in \left(\sqrt{2} - \sqrt{2 - \frac{2}{3} f_{max} \sqrt{1 - \frac{1}{\gamma}}}, \sqrt{2} + \sqrt{2 - \frac{2}{3} f_{max} \sqrt{1 - \frac{1}{\gamma}}} \right).$$

As $\gamma \to \infty$, this range $\to \left(\sqrt{2} - \frac{2}{\sqrt{3}}, \sqrt{2} + \frac{2}{\sqrt{3}} \right)$, and $f_{max} \to 1$, so clearly $y^* = O(1)$. □

Corollary 6.9 *Let v satisfy equation (6.35) where f is given by (6.34). Let x^* be the positive solution to $v_{xxx}(0) = 0$, i.e. to $f'(v(x)) = 0$. Then $x^* = O(\varepsilon)$ as $\varepsilon \to 0$ and $\gamma \to \infty$. Also, $v_x(0) = O(\varepsilon^{-1})$.*

Proof Lemma 6.8 establishes that $y^* = O(1)$ for equation (6.37) but since $x = \varepsilon y$ (6.36), we have $x^* = \varepsilon y^*$ so $x^* = O(\varepsilon)$. Similarly, since $w_y(0) = O(1)$, we have $v_x(0) = O(\varepsilon^{-1})$. □

Note that although we have used a specific function f here, the lemma will hold with any f of the same form, i.e. of the form

$$f(v) = \frac{1}{\gamma} G(v) - v,$$

where $G = g^{-1}$ is as specified in (2.4).

7. Behaviour of the reaction-diffusion equations

7.1 Cottet's equation.

In the case studied by Cottet ($\mu \equiv 1$), there is an energy functional corresponding to that for the Hopfield net. Assuming appropriate boundary conditions, we express equation (6.24) as

$$v_t = \frac{\gamma}{G'(v)} \left[\varepsilon^2 \frac{\hat{\tau}}{2} \Delta v - f(v) \right], \tag{7.1}$$

where $f(v) = \alpha \frac{G(v)}{\gamma} - \tau_0 v$, and we require $\tau_0 > \alpha \frac{G'(0)}{\gamma} = \frac{\alpha\beta}{\gamma}$. In one space dimension, for example, we have the energy functional

$$E[v] = \int \left[\varepsilon^2 \frac{\hat{\tau}}{4} v_x^2 + F(v) \right] dx, \tag{7.2}$$

where $F'(v) = f(v)$. It is easy to show that

$$\dot{E} = -\int \frac{G'(v)}{\gamma} v_t^2 \, dx \leq 0.$$

This energy functional can be modified in the obvious way if a threshold function and an external input function are included. The existence of this energy functional is not surprising since Cottet's connection matrices are symmetric. However, it can be generalized for equation (6.16), still in one spatial dimension, as long as $\mu_{yy}(x,x) = \mu_{xx}(x,x)$ and $\mu_y(x,x) = \mu_x(x,x)$ as proven in the following proposition.

Proposition 7.1 *If $\mu_{yy}(x,x) = \mu_{xx}(x,x)$ and $\mu_y(x,x) = \mu_x(x,x)$, then equation (6.16), in one space dimension with periodic boundary conditions, has an energy functional given by*

$$E[v] = -\frac{1}{2} \int \varepsilon^2 \gamma (D_2 v) v \, dx + \alpha \int \left(\int_0^v G(V) \, dV \right) dx.$$

Proof We check that energy decreases with time along solution trajectories:

$$\dot{E} = -\frac{1}{2}\int \varepsilon^2\gamma\left[(D_2v)v_t + v(D_2v)_t\right]\,dx + \alpha\int G(v)v_t\,dx\,.$$

Now

$$\int \left[(D_2v)v_t + v(D_2v)_t\right]\,dx = \int\left(\frac{\tau}{2}\left(\mu_{yy}v + 2\mu_y v_x + \mu v_{xx}\right) + \frac{\tau_0}{\varepsilon^2}\mu v\right)v_t\,dx$$
$$+ \int\left(\frac{\tau}{2}\left(\mu_{yy}v_t v + 2\mu_y v_{xt}v + \mu v_{xxt}v\right) + \frac{\tau_0}{\varepsilon^2}\mu v_t v\right)\,dx\,,$$

where μ and its derivatives are understood to be evaluated at (x,x). The terms with v_{xt} and v_{xxt} we integrate by parts,

$$\int \mu_y(x,x)v_{xt}v\,dx = -\int\left(\mu_{xy}(x,x)v + \mu_{yy}(x,x)v + \mu_y(x,x)v_x\right)v_t\,dx\,,$$

$$\int \frac{1}{2}\mu(x,x)v_{xxt}v\,dx = -\int\frac{1}{2}\left(\mu_x(x,x)v + \mu_y(x,x)v + \mu(x,x)v_x\right)v_{xt}\,dx$$
$$= \int\frac{1}{2}\left(\mu_{xx}v + 2\mu_{xy}v + \mu_{yy}v + 2\mu_x v_x + 2\mu_y v_x + \mu v_{xx}\right)v_t\,dx\,.$$

Thus, collecting all like terms together,

$$\int \left[(D_2v)v_t + v(D_2v)_t\right]\,dx$$
$$= \int\left(\tau\left(\mu v_{xx} + (\mu_x+\mu_y)v_x + \frac{\mu_{xx}+\mu_{yy}}{2}v\right) + \frac{2\tau_0}{\varepsilon^2}\mu v\right)v_t\,dx\,.$$

Under the symmetry conditions on the derivatives of μ, this becomes

$$\int \left[(D_2v)v_t + v(D_2v)_t\right]\,dx = \int\left(\tau\left(\mu v_{xx} + 2\mu_y v_x + \mu_{yy}v\right) + \frac{2\tau_0}{\varepsilon^2}\mu v\right)v_t\,dx$$
$$= \int 2(D_2v)v_t\,dx\,,$$

so finally,

$$\dot{E} = -\int\left(\varepsilon^2\gamma D_2 v - \alpha G(v)\right)v_t\,dx = -\int G'(v)v_t^2\,dx \le 0\,.$$

□

Interestingly, since this is a weaker condition than symmetry of μ ($\mu(x,y) = \mu(y,x)$), and since also η need not be even as long as it satisfies the moment conditions, we have an energy functional that applies to a wider class of connection functions T^ε than the symmetric ones.

Note that if $\tau_0 < \frac{\alpha\beta}{\gamma}$, $f(v)$ does not have non-trivial zeros, so the gain in equation (7.1) is too low and all solutions collapse immediately to zero. (This is not true, of course, when threshold and input terms are present). Also, if $\tilde{\tau} < 0$, the diffusion operator is negative and the reaction-diffusion equation is ill-posed as mentioned in the general case above. In Cottet's case, T is essentially just η, so this says that the neural connections should be predominantly excitatory for the reaction-diffusion equation to be well-posed.

Equation (7.1) is similar to the Ginzburg-Landau equation or Allen-Cahn equation [2] (see also [7], which deals with a different discrete form of this equation). The factor $\frac{\gamma}{G'(v)}$ does not alter the dynamics significantly, as the energy functional is the same (recall that $G'(v) \geq \beta$, for all $v \in (-1,1)$). In the one-dimensional case, we mentioned in section 6.5 that an equation of this type has been studied in depth [8,9,10,55,67]:

$$v_t = \varepsilon^2 v_{xx} - f(v),$$

where f has the same form as above. They show that solutions typically develop transition layers that are metastable, eventually collapsing to a constant state. It is known that no stable patterns exist for this equation with Neumann or periodic boundary conditions [31,47]. Numerical evidence suggests that Cottet's equation behaves similarly. The image processing application suggested by Cottet for this type of equation depends on this development of transition layers: diffusion smooths out noise and the reaction term enhances contrast [14,15].

It is possible that some of the metastable solutions of the reaction-diffusion

equation correspond to stable states of corresponding Hopfield nets (recall that the approximation is only good for a fixed time interval). In fact there are PDEs of essentially the same type (therefore having no stable patterns) whose discretizations (by standard finite difference methods) do have stable patterns (see [22] and Chapter 8).

An external forcing function of sufficient strength will dominate the dynamics, as for the Hopfield network and the integro-differential equation model of Chapter 5. A constant, positive input over a domain, for example, would make the reaction term positive, driving the solution up towards v_s. For example, in one dimension, Cottet's equation with an input term is (from equation 6.25)

$$v_t = \frac{1}{G'(v)}\left[\varepsilon^2\gamma\frac{\hat{\tau}}{2}v_{xx} + \gamma\tau_0 v - \alpha G(v) + \lambda c(x)\right].$$

Since $\gamma\tau_0 v > \alpha G(v)$ for $v \in (0, v_s)$, an input $c(x) > \frac{\gamma\tau_0}{\lambda}v_s$ would be more than enough to guarantee that v_t was driven upwards. In general, an input of large magnitude forces v to duplicate its sign pattern, except for smooth transition layers between positive and negative regions. The effects of inputs of smaller magnitude are harder to characterize.

Note also that if a strong input is active only over some region (or even only at the boundary of the region) it will determine the behaviour of the solution nearby, but the effect of this input will propagate in space only very slowly. The small spatial diffusion means that the transition layers, once developed can move only extremely slowly, so the solution in an area not immediately affected by an external force will be determined by initial data rather than effects propagating in space from the area where the input is active.

7.2 The general equation.

The generalized equation (6.16) has apparently not been studied. It should retain features of the specific equation in which $\mu \equiv 1$, but the variation in μ must have a modulating effect. Analysis and numerical experimentation are required. Preliminary numerical simulations of this equation discretized in space (see Chapter 8) suggest that in areas where $\mu(x,x)$ is small the solution will decay to zero, and in areas where $\mu(x,x)$ is sufficiently large it will behave as it would without the μ term (see, e.g. Fig. 8.4). If we allow μ to take on negative values (with positive $\hat{\tau}$), then oscillations can apparently occur. Once again, this is an indication of the value of inhibition in generating complex behaviour.

7.3 Conclusions.

Reaction-diffusion equations of the type obtained in Chapter 6 give us approximations to the behaviour of Hopfield net equations with connection matrices of a certain form. Study of the reaction-diffusion equations can shed some light on the dynamics of such networks. Also, the conditions of the theorem giving the approximation give a rough classification of connection matrices: those that approximate the PDEs and those that do not. Those that are approximations to PDEs are also approximations to the discrete systems obtained by standard discretizations of the PDEs.

The theorems also provide a semi-rigorous analysis of the relationship between continuous and discrete space models and suggest what form of connection matrix is more likely to produce complex behaviour. Specifically, if we want a Hopfield net with more complex behaviour than the PDE-like ones, we should look at connection matrices that either vary wildly in nearby entries or perhaps deviate significantly

from the banded structure. In fact the ill-posedness of the PDE in the case of negative moments is tantalizing evidence for interesting and complex behaviour.

8. Finite difference discretizations of the reaction-diffusion equations

8.1 Forms of the reaction-diffusion equation and their discretizations.

In Chapter 7, it was pointed out that for PDEs of essentially the same type as Cottet's reaction-diffusion equation, and therefore having no stable patterns, discretizations (by standard finite difference methods) can nevertheless have stable patterns (see [22]). A different approach to discretizations of Allen-Cahn equations is given in [7].

Let us consider the one-dimensional case of Cottet's reaction-diffusion equation, given by (see equation (6.24))

$$v_t = \frac{\gamma}{G'(v)}\left[\varepsilon^2\frac{\hat{\tau}}{2}v_{xx} + \tau_0 v - \frac{G(v)}{\gamma}\right], \tag{8.1}$$

where to simplify calculations, we have taken $\alpha = 1$ (though this is not necessary). We treat it as an approximation via finite differences to a system of ordinary differential equations. Let

$$v_{xx} \sim \frac{v_{m+1} - 2v_m + v_{m-1}}{\varepsilon^2}$$

be the finite difference approximation (stepsize ε). Then equation (8.1) becomes

$$\begin{aligned}\dot{v}_m &= \frac{\gamma}{G'(v_m)}\left[\frac{\hat{\tau}}{2}\left(v_{m+1} - 2v_m + v_{m-1}\right) + \tau_0 v_m - \frac{G(v_m)}{\gamma}\right] \\ &= \frac{\gamma}{G'(v_m)}\left[\frac{\hat{\tau}}{2}\left(v_{m+1} + v_{m-1}\right) + (\tau_0 - \hat{\tau})v_m - \frac{G(v_m)}{\gamma}\right].\end{aligned} \tag{8.2}$$

We could also use an equation in a different variable, not exactly the membrane potential (see below), that is equivalent to (8.1), namely

$$u_t = -u + g\left(\gamma\left(\varepsilon^2\frac{\hat{\tau}}{2}u_{xx} + \tau_0 u\right)\right), \tag{8.3}$$

recalling that $G = g^{-1}$. We do this to show that equation (8.2) is equivalent to an equation that arises from the dynamic Ising model of statistical physics, so that previous results for this theory can be used (see [22]). Equation (8.3) can be shown to be equivalent to (8.1) by means of the transformation

$$v = g\left(\gamma\left(\varepsilon^2\frac{\hat{\tau}}{2}u_{xx} + \tau_0 u\right)\right) \tag{8.4}$$

so that (recalling that $G = g^{-1}$)

$$\varepsilon^2\frac{\hat{\tau}}{2}u_{xx} + \tau_0 u = \frac{G(v)}{\gamma}. \tag{8.5}$$

Note also, from (8.3) and (8.4) that

$$u_t = -u + v. \tag{8.6}$$

Differentiating (8.6) twice by x gives

$$u_{xxt} = -u_{xx} + v_{xx}. \tag{8.7}$$

Then, differentiate (8.5) with respect to time, to get

$$\varepsilon^2\frac{\hat{\tau}}{2}u_{xxt} + \tau_0 u_t = \frac{G'(v)}{\gamma}v_t.$$

Thus, using (8.6), (8.7) and (8.5) again,

$$\begin{aligned} v_t &= \frac{\gamma}{G'(v)}\left[\varepsilon^2\frac{\hat{\tau}}{2}u_{xxt} + \tau_0 u_t\right] \\ &= \frac{\gamma}{G'(v)}\left[\varepsilon^2\frac{\hat{\tau}}{2}\left(-u_{xx} + v_{xx}\right) + \tau_0\left(-u + v\right)\right] \\ &= \frac{\gamma}{G'(v)}\left[\varepsilon^2\frac{\hat{\tau}}{2}v_{xx} + \tau_0 v - \frac{G(v)}{\gamma}\right], \end{aligned}$$

which is (8.1).

Discretizing (8.3) via

$$u_{xx} \sim \frac{u_{m+1} - 2u_m + u_{m-1}}{\varepsilon^2},$$

gives

$$\dot{u}_m = -u_m + g\left[\gamma\left(\frac{\hat{\tau}}{2}\left(u_{m+1} + u_{m-1}\right) + \left(\tau_0 - \hat{\tau}\right)u_m\right)\right]. \tag{8.8}$$

Thus, (8.2) and (8.8) are equivalent dynamical systems (setting v_m to the term in square brackets in (8.8) gives the equivalence directly).

It is well known that equations like (8.1) without the $\frac{\gamma}{G'(v)}$ term, i.e.

$$v_t = \varepsilon^2 \frac{\hat{\tau}}{2} v_{xx} + \tau_0 v - \frac{G(v)}{\gamma}, \tag{8.9}$$

with Neumann or periodic boundary conditions, do not have stable patterns, no matter how small we make ε. The only equilibria are flat (constant in space). This is the theory referred to in Chapter 7 [31,47]. Now, equation (8.1) has the same energy functional and equilibria as equation (8.9), as shown below. It must also, therefore, have the same stability properties of equilibria (since stability of an equilibrium is a property of the energy functional). Therefore, equation (8.1), and by equivalence (8.3), also have no stable patterns under Neumann or periodic boundary conditions.

However, previous work [22] has shown that a special case of the discrete space equation, (8.8), namely where $\tau_0 = \hat{\tau}$ and $g = \tanh$, has stable patterns. This is somewhat surprising. In the next section, we extend these results to more general forms of equation (8.8), and in particular to more than one dimension. Let us first consider the significance of these results, however.

Since for small ε, the approximation given by the discretizations above are good, solutions of the partial differential equations should stick close to the stable patterns for a long time if they start close to them. But they must eventually collapse to a flat equilibrium. This shows that there is a fundamental difference in asymptotic behaviour between the Hopfield network and the reaction-diffusion equation, no matter how good the approximation, and that Hopfield networks of

the form covered by the approximation theorems of Chapter 6 can have stable patterns. (After all, equation (8.2) or (8.8) is a Hopfield network with a simple tri-diagonal connection matrix). Remember that these simple Hopfield networks are representatives of a class of Hopfield networks, all of which are approximated by the same reaction-diffusion equation.

8.2 Lyapunov functional.

First, let us look at equilibrium equations for the discrete space equations, (8.2) and (8.8). These are the same equation written in terms of different variables, and in fact the two variables satisfy an identical equilibrium equation (though their time derivatives are different). The equilibrium equation for (8.2) is clearly

$$\frac{G(v_m)}{\gamma} = \frac{\hat{\tau}}{2}(v_{m+1} + v_{m-1}) + (\tau_0 - \hat{\tau})\, v_m. \tag{8.10}$$

The equilibrium equation for (8.8) is

$$u_m = g\left[\gamma\left(\frac{\hat{\tau}}{2}(u_{m+1} + u_{m-1}) + (\tau_0 - \hat{\tau})\, u_m\right)\right]. \tag{8.11}$$

However, using the relation $G = g^{-1}$, this can be written

$$\frac{G(u_m)}{\gamma} = \frac{\hat{\tau}}{2}(u_{m+1} + u_{m-1}) + (\tau_0 - \hat{\tau})\, u_m,$$

which is identical to (8.10).

Furthermore, equations (8.1) and (8.3), and even (8.9) have the same Lyapunov functional, namely,

$$E[v] = \int \left(\frac{\varepsilon^2}{2} v_x^2 - v^2 + Q(v)\right) dx\,, \tag{8.12}$$

where

$$Q'(v) = q(v) = \frac{2}{\hat{\tau}}\left[\frac{G(v)}{\gamma} - (\tau_0 - \hat{\tau})v\right], \tag{8.13}$$

so that (8.2) is

$$\dot{v}_m = \frac{\gamma}{G'(v_m)} \frac{\hat{\tau}}{2} \left[v_{m+1} + v_{m-1} - q(v_m) \right].$$

(Equation (8.12) is just another formulation of equation (7.2)). Their discretizations, $\mathbf{v} = (v_m)$, have the analogous Lyapunov functional

$$E[\mathbf{v}] = \sum_j Q(v_j) - \frac{1}{2} \sum_j v_j (v_{j+1} + v_{j-1}). \tag{8.14}$$

Lemma 8.1 *$E[\mathbf{v}]$ in (8.14) is a Lyapunov functional for equation (8.2) with periodic boundary conditions.*

Proof To show that (8.14) is a Lyapunov functional, it is necessary to show that equilibria are critical points of the energy surface and that energy decreases with time. Note that (8.10) is equivalent to

$$v_{m+1} + v_{m-1} = q(v_m), \tag{8.15}$$

with q given by (8.13). First,

$$\frac{d}{dv_j} E[\mathbf{v}] = q(v_j) - (v_{j-1} + v_{j+1}),$$

which is clearly zero at equilibria. Second,

$$\begin{aligned}
\dot{E}[\mathbf{v}] &= \sum_j q(v_j)\dot{v}_j - \frac{1}{2} \sum_j \left[v_j \dot{v}_{j+1} + v_{j+1}\dot{v}_j + v_j \dot{v}_{j-1} + v_{j-1}\dot{v}_j \right] \\
&= \sum_j q(v_j)\dot{v}_j - \frac{1}{2} \sum_j \left[(v_{j+1} + v_{j-1})\, \dot{v}_j \right] - \frac{1}{2} \sum_k v_{k+1}\dot{v}_k - \frac{1}{2} \sum_\ell v_{\ell-1}\dot{v}_\ell \\
&= \sum_j \left[q(v_j) - (v_{j+1} + v_{j-1}) \right] \dot{v}_j \\
&= -\sum_j \frac{2G'(v_j)}{\gamma\hat{\tau}} \dot{v}_j^2 \leq 0,
\end{aligned}$$

since $\hat{\tau}$, γ, and $G'(v_j) \geq 0$, and where we have used (8.13). □

Lemma 8.2 *$E[\mathbf{u}]$ from (8.14) is a Lyapunov functional for equation (8.8) with periodic boundary conditions.*

Proof The proof is similar to the proof of Lemma 8.1. Note that since (8.11) is equivalent to (8.10), it is also equivalent to (8.15), with $\mathbf{u}$ replacing $\mathbf{v}$. Thus,

$$\frac{d}{du_j}E[\mathbf{u}] = q(u_j) - (u_{j-1} + u_{j+1}),$$

and is zero at equilibria as before. Also,

$$\begin{aligned}\dot{E}[\mathbf{u}] &= \sum_j \left[q(u_j) - (u_{j+1} + u_{j-1})\right]\dot{u}_j \\ &= -\sum_j \frac{2}{\gamma\hat{\tau}}\left[G(\dot{u}_j + u_j) - G(u_j)\right]\dot{u}_j, \text{ by (8.8) and (8.13)}, \\ &= -\sum_j \frac{2}{\gamma\hat{\tau}}G'(\xi)\dot{u}_j^2 \le 0,\end{aligned}$$

where ξ is between u_j and $u_j + \dot{u}_j$, and we have used the mean value theorem. □

Thus, equilibria for the three equations are the same as are their stability properties. We can therefore focus our attention on equations (8.10, or equivalently, 8.15) and (8.14) to obtain results for the discretizations of the equation of Cottet (8.1), its transformed form (8.3), or the Ginzburg-Landau (or Allen-Cahn) equation, (8.9). We remark here that in the form of the Ginzburg-Landau equation usually studied,

$$v_t = \varepsilon^2 v_{xx} - f(v), \tag{8.16}$$

the reaction term, $f(v)$, is something like $\lambda(v^3 - v)$, which has the property that $|f(v)| \to \infty$ as $|v| \to \infty$. In our neural network theory we have

$$f(v) = \frac{G(v)}{\gamma} - \tau_0 v, \tag{8.17}$$

which has the property that $|f(v)| \to \infty$ as $|v| \to 1$. However, as long as initially $v \in [-v_s, v_s]$, where v_s is the positive solution to $f(v) = 0$, the solutions remain in

this interval, so the behaviour of f for $|v| > v_s$ is irrelevant. Thus, our $f(v)$ is of the type studied by Carr and Pego [8,9,10]. The space discretization of this Ginzburg-Landau equation (8.16) has equilibrium given by (8.15) with $q(v) = 2v + f(v)$. More generally, for an equation like (8.9) or (8.1),

$$q(v) = 2v + \frac{2}{\tau} f(v). \tag{8.18}$$

In the remainder of this chapter we will use two examples of this general q. We are most interested in the form given by (8.13) which comes from our neural field model, and we will refer to this as q_1. For comparison we also discuss an example for the Ginzburg-Landau equation (8.16), with $f_2(v) = \lambda(v^3 - v)$, so that

$$q_2(v) = (2 - \lambda)v + \lambda v^3 . \tag{8.19}$$

It is sometimes convenient to work with $f(v)$ and sometimes with $q(v)$ and we shall use both as appropriate.

8.3 Stability of flat equilibria.

We wish to establish that there exist stable patterns for these dynamical systems; i.e. that there are stable equilibria that are not constant in space. We examine equations that have equilibria given by (8.15) and energy functional given by (8.14). In these equations we will allow $f(v)$ to be any smooth function satisfying the conditions

$$f(-v) = -f(v), \text{ i.e. } f \text{ odd, so}$$

$$f(0) = 0 \tag{8.20}$$

$$vf''(v) > 0 \text{ for } v \neq 0.$$

From (8.18) it is clear that the same conditions must then apply to $q(v)$, since $\tau > 0$. These conditions hold for $f(v)$ given by (8.17) and $q(v)$ by (8.13) if $G = g^{-1}$ is odd

and satisfies conditions (2.4). They hold for the example in (8.19) if $\lambda > 0$. If in addition, $f'(0) < 0$, then f is of the form considered by Carr and Pego [8,9,10]. For convenience, we continue to assume periodic boundary conditions.

First, we examine the stability of the trivial equilibrium, $v \equiv 0$.

Proposition 8.3 *The trivial solution to (8.15), $v \equiv 0$, is asymptotically stable for $q'(0) \geq 2$ and unstable for $q'(0) < 2$.*

Proof The result is easy to see if we rewrite the energy functional (8.14) in the form

$$E[\mathbf{v}] = \sum_j Q(v_j) - \sum_j v_j^2 + \frac{1}{4}\sum_j \left[(v_{j+1} - v_j)^2 + (v_{j-1} - v_j)^2\right] \qquad (8.21)$$

which can be shown to be equivalent under periodic boundary conditions. Now it is clear that the last term, which gives a contribution from interactions, can only increase as any constant equilibrium is perturbed and the first two terms contain no interactions and so may be treated separately for each v_j. Letting $F(v) = Q(v) - v^2$, we have that $F'(0) = 0$ by (8.20) and $F''(0) = q'(0) - 2$. Thus, if $q'(0) > 2$, then $F''(0) > 0$ and $F(v) > 0$ for v near 0, so the 0 equilibrium is asymptotically stable. If $q'(0) = 2$, then $F''(0) = 0$ and $F'''(0) = q''(0) = 0$, since q is an odd function, but $F^{(4)}(0) = q'''(0) > 0$, since $q''(v) < 0$ for $v < 0$ and $q''(v) > 0$ for $v > 0$. Thus, the 0 equilibrium is still asymptotically stable for $q'(0) = 2$. If $q'(0) < 2$, then $F''(0) < 0$ and $F(v) < 0$ for v near 0. A perturbation that is constant in space will decrease the energy, so the 0 equilibrium is unstable. □

The proposition is illustrated by our two examples:

Example 8.1 For $q_1(v)$ from (8.13), the condition $q'(0) < 2$ for instability of the trivial equilibrium becomes $\tau_0 > \frac{\beta}{\gamma}$. (Recall that $G'(v) \geq G'(0) = \beta$). □

Example 8.2 For $q_2(v)$ given by (8.19), the trivial solution is unstable if $\lambda > 0$. □

We are really only interested in the case $f'(0) < 0$ ($q'(0) < 2$), so that the constant space solutions ($v_j = v_s$ for all j or $v_j = -v_s$ for all j) exist and the zero solution is unstable. These constant-space solutions are always stable.

Proposition 8.4 *The two solutions $v_j = v_s$ for all j, and $v_j = -v_s$ for all j, where v_s is the positive solution to $f(v) = 0$ when $f'(0) < 0$, are asymptotically stable.*

Proof As before, use the energy functional in the form (8.21). Again the last term cannot decrease when a constant equilibrium is perturbed and the first two terms may be treated separately for each v_j. With $F(v) = Q(v) - v^2$, we have

$$F'(\pm v_s) = q(\pm v_s) \mp 2v_s = \frac{2}{\hat{\tau}} f(\pm v_s) = 0,$$

and

$$F''(\pm v_s) = \frac{2}{\hat{\tau}} f'(\pm v_s) > 0\,.$$

Therefore, the energy is greater for perturbations of this equilibrium and it is asymptotically stable. □

8.4 A stable equilibrium of period 6.

Now we demonstrate the existence, under certain conditions on $q(v)$, of stable equilibria that are not constant in space. Consider a period 6 equilibrium, v^*, of the form

$$v_0 = v_4 = 0; \quad v_2 = v_3 = -v_5 = -v_6 = B > 0, \tag{8.22}$$

where B is a constant (Fig. 8.1).

Proposition 8.5 *An equilibrium of the form* (8.22) *exists if and only if*

$$q'(0) < 1, \text{ i.e. } f'(0) < -\frac{\hat{\tau}}{2}. \tag{8.23}$$

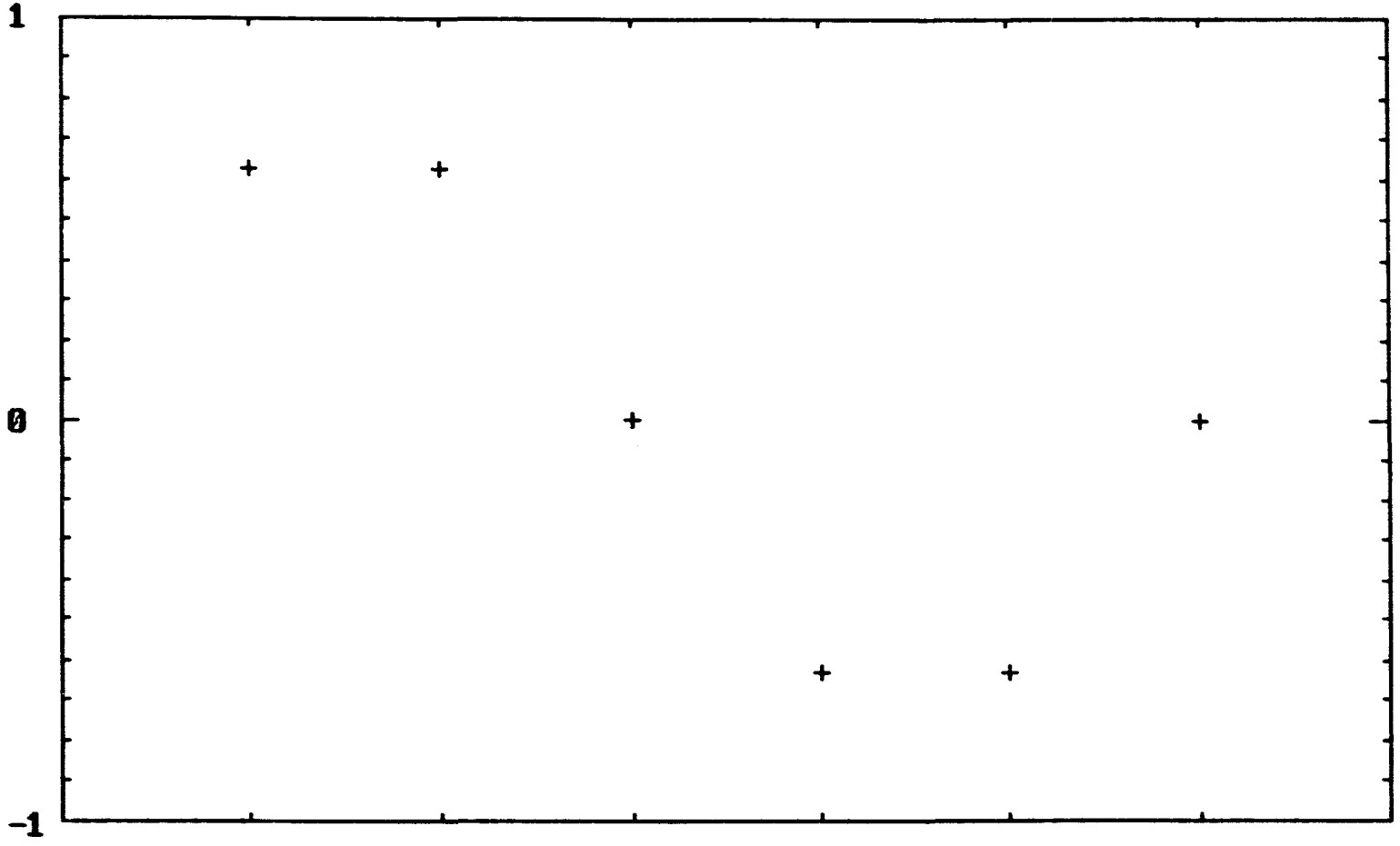

Figure 8.1 A period 6 equilibrium for a discrete space version of Cottet's equation; i.e., a solution to equation (8.10) with $\gamma = 1.2$, $\tau_0 = \hat{\tau} = 2$, $\alpha = 1$ and $G = \tanh^{-1}$.

Proof An equilibrium must satisfy (8.15) at every point. Here, we always have $v_3 + v_5 = q(v_4) = q(0) = 0$, by (8.20), so the equilibrium exists exactly when it is possible to find a B such that $B = v_2 + v_4 = q(v_3) = q(B)$ (the other cases are taken care of by the symmetries in (8.22) and in q). That is, the equilibrium exists when $q'(0) < 1$, by (8.20) applied to q. Also, $B = q(B)$ is equivalent to $B = 2B + \frac{2}{\hat{\tau}} f(B)$ or $f(B) = -\frac{\hat{\tau}}{2} B$, which has a positive solution when $f'(0) < -\frac{\hat{\tau}}{2}$. □

Example 8.1 (continued) Condition (8.23) implies $\tau_0 - \frac{\hat{\tau}}{2} > \frac{\beta}{\gamma}$ for $q_1(v)$ as in (8.13), which is not possible if $2\tau_0 \leq \hat{\tau}$, so the condition for existence of this period 6 equilibrium is in this case $2\tau_0 > \hat{\tau}$ and $\gamma > \frac{\beta}{\tau_0 - \hat{\tau}/2}$. □

Example 8.2 (continued) Condition (8.23) implies $\lambda > \frac{\hat{\tau}}{2}$ for $q_2(v)$ as in (8.19). □

We can determine the stability of this equilibrium from the energy functional (8.14) by means of the matrix of second derivatives $\frac{\partial^2 E}{\partial v_i \partial v_j}$. The equilibrium v^* is asymptotically stable if and only if this matrix evaluated at v^* is positive definite, since this ensures that the energy is greater in a neighbourhood of the equilibrium (for Lyapunov's stability theorem, see, e.g. [33]). This matrix is

$$P = \frac{\partial^2 E}{\partial v_i \partial v_j}[v^*] = \begin{bmatrix} q'(-B) & -1 & 0 & 0 & 0 & -1 \\ -1 & q'(0) & -1 & 0 & 0 & 0 \\ 0 & -1 & q'(B) & -1 & 0 & 0 \\ 0 & 0 & -1 & q'(B) & -1 & 0 \\ 0 & 0 & 0 & -1 & q'(0) & -1 \\ -1 & 0 & 0 & 0 & -1 & q'(-B) \end{bmatrix}. \tag{8.24}$$

Proposition 8.6 *P is positive definite if and only if*

$$q'(0) > 0 \tag{8.25a}$$

and

$$q'(\pm B) - \frac{2}{q'(0)} - 1 \geq 0. \tag{8.25b}$$

Proof Let $\mathbf{x}$ be an arbitrary vector in $\mathbf{R}^6$. Then

$$\mathbf{x}^t P\mathbf{x} = q'(-B)\left[x_1^2 + x_6^2\right] + q'(B)\left[x_3^2 + x_4^2\right] + q'(0)\left[x_2^2 + x_5^2\right]$$
$$- 2\left[x_6x_1 + x_1x_2 + x_2x_3 + x_3x_4 + x_4x_5 + x_5x_6\right].$$

Now if $q'(0) < 0$ then we can take $x_1 = x_3 = x_4 = x_6 = 0$ and $x_2 = x_5 \neq 0$ to make $\mathbf{x}^t P\mathbf{x} < 0$ so P is not positive definite. If $q'(0) = 0$ then we may take $x_1 = x_3 = x_4 = x_6 = 1$, say, and $x_2 = x_5 > \frac{q'(B)-1}{2}$ so that

$$\mathbf{x}^t P\mathbf{x} = 2q'(-B) + 2q'(B) - 2\left[4x_2 + 2\right] = 4\left[q'(B) - 1 - 2x_2\right] < 0$$

and again, P is not positive definite. If $q'(0) > 0$, we can rewrite $\mathbf{x}^t P\mathbf{x}$ as

$$\begin{aligned}\mathbf{x}^t P\mathbf{x} = {} & q'(0)\left[x_2 - \frac{1}{q'(0)}(x_1 + x_3)\right]^2 + q'(0)\left[x_5 - \frac{1}{q'(0)}(x_4 + x_6)\right]^2 \\ & + (x_3 - x_4)^2 + (x_1 - x_6)^2 + \frac{1}{q'(0)}\left[(x_1 - x_3)^2 + (x_4 - x_6)^2\right] \\ & + \left[q'(-B) - \frac{2}{q'(0)} - 1\right]\left[x_1^2 + x_6^2\right] + \left[q'(B) - \frac{2}{q'(0)} - 1\right]\left[x_3^2 + x_4^2\right]\end{aligned} \tag{8.26}$$

to see that P is positive definite when $q'(\pm B) - \frac{2}{q'(0)} - 1 > 0$. If this quantity is ≤ 0, we can take $x_1 = x_3 = x_4 = x_6$ and $x_2 = x_5 = \frac{2}{q'(0)}x_1$ to make $\mathbf{x}^t P\mathbf{x} \leq 0$ for $\mathbf{x} \neq 0$, so P is not positive definite. □

Thus the period 6 equilibrium exists and is asymptotically stable exactly when $0 < q'(0) < 1$ and $q'(\pm B) - \frac{2}{q'(0)} - 1 > 0$. In terms of f, conditions (8.25) are

$$1 + \frac{2}{\hat{\tau}}f'(B) > \frac{1}{1 + \frac{1}{\hat{\tau}}f'(0)}, \tag{8.27a}$$

$$f'(0) > -\hat{\tau}. \tag{8.27b}$$

Example 8.1 (continued) For equation (8.10), we have seen that the period 6 equilibrium exists only when $\tau_0 - \frac{\hat{\tau}}{2} > \frac{\beta}{\gamma}$ and stability requires at least the equivalent

of (8.27b), which is $\tau_0 - \hat{\tau} < \frac{\beta}{\gamma}$. Thus, a necessary condition for the existence of a stable period 6 equilibrium for large γ (i.e. as $\gamma \to \infty$) is

$$\frac{\hat{\tau}}{2} < \tau_0 \leq \hat{\tau}\,.$$

If we consider the case $\tau_0 = \hat{\tau} = 2$ and $\beta = 1$, for example, (8.27b) is automatically satisfied and (8.27a) becomes

$$\frac{G'(B)}{\gamma} - 1 > 2\gamma\,. \tag{8.28}$$

Since, by definition, $q_1(B) = B$, we have (with $\tau_0 = \hat{\tau} = 2$, $\beta = 1$) $f(B) = -B$ or $\frac{G(B)}{\gamma} = B$ so that $\gamma = \frac{G(B)}{B}$ and B goes from 0 to 1 as γ goes from 1 to ∞. Hence, with these parameters, we may express (8.28) as

$$\frac{BG'(B)}{G(B)} - 1 > \frac{2G(B)}{B}\,. \tag{8.29}$$

If $G(v) = \tanh^{-1} v$, for example, this condition becomes

$$\frac{1}{1-B^2} > \frac{\tanh^{-1} B}{B} + 2\left(\frac{\tanh^{-1} B}{B}\right)^2, \tag{8.30}$$

and it can be shown by asymptotic analysis [22] that this is true for large enough B (or γ). A numerical calculation shows that it is true for $\gamma > 1.8576$. □

Example 8.2 (continued) For $q_2(v)$ given by (8.19), condition (8.27) becomes

$$1 + \frac{2\lambda}{\hat{\tau}}\left(3B^2 - 1\right) > \frac{1}{1 - \frac{\lambda}{\hat{\tau}}}\,, \quad \lambda < \hat{\tau}\,.$$

In this case, we can find B in terms of λ as follows:

$$q_2(B) = B \Rightarrow f_2(B) = -\frac{\hat{\tau}}{2}B \Rightarrow \lambda\left(B^3 - B\right) = -\frac{\hat{\tau}}{2}B \Rightarrow B^3 = \left(1 - \frac{\hat{\tau}}{2\lambda}\right)B$$

$$\Rightarrow B^2 = 1 - \frac{\hat{\tau}}{2\lambda}\,.$$

Using this, our stability condition becomes

$$\frac{4\lambda}{\hat{\tau}} - 2 > \frac{1}{1 - \frac{\lambda}{\hat{\tau}}},$$

where the denominator on the right is positive since $\lambda < \hat{\tau}$. Thus,

$$2\left(1 - \frac{\lambda}{\hat{\tau}}\right)\left(\frac{2\lambda}{\hat{\tau}} - 1\right) > 1$$

or

$$4\left(\frac{\lambda}{\hat{\tau}}\right)^2 - 6\frac{\lambda}{\hat{\tau}} + 3 < 0.$$

But this quadratic inequality is not satisfied for any real $\frac{\lambda}{\hat{\tau}}$, showing that the period 6 equilibrium is never stable for f_2. □

8.5 Large scale stable patterns.

The existence of a stable period 6 equilibrium for some $q(v)$ is of limited interest in itself, since we want to consider ε to be small, and as $\varepsilon \to 0$, the grid shrinks and the period 6 equilibrium oscillates at very high frequency (and thus, in a sense, consists only of transition layers). However, there can also exist stable equilibria of arbitrarily large period, for appropriate $q(v)$. This is demonstrated by the following series of propositions.

Proposition 8.7 *If* $0 < q'(0) < 2$ *and* q *satisfies* (8.20) *then there exists a periodic solution to* (8.15) *of period* N *for all even* $N \geq 6$,

$$N > \frac{2\pi}{\cos^{-1}\left(\frac{q'(0)}{2}\right)}. \tag{8.31}$$

Proof Let $N \geq 6$ be even. Let $\phi(n) = \sin\left(\frac{2\pi n}{N}\right)$. We carry out the proof by an iteration. Define an initial vector (of period N) as

$$v_n^{(0)} = \delta\phi(n) \text{ for } n = 0, 1, \ldots, N-1, \tag{8.32}$$

where the constant $\delta > 0$ is to be chosen (e.g. Fig. 8.2a).

Now iterate according to

$$v_n^{(m+1)} = q^{-1}\left(v_{n+1}^{(m)} + v_{n-1}^{(m)}\right), \tag{8.33}$$

where we have used the fact that $q'(0) > 0$ to ensure that q is strictly increasing so that q^{-1} exists and is also increasing. If $v_n^{(m)} \geq v_n^{(m-1)}$ for all n, then

$$v_n^{(m+1)} \geq q^{-1}\left(v_{n+1}^{(m-1)} + v_{n-1}^{(m-1)}\right) = v_n^{(m)}$$

for all n. So an initially increasing sequence must continue to increase. Similarly, an initially decreasing sequence must continue to decrease. Now, for the initial vector in (8.32), the points $v_n^{(m)}$ where $n = \frac{kN}{2}$, i.e. multiples of $N/2$, will be zero and will remain zero due to the symmetries in the vector and the iteration. Thus, we need only show that in the positive parts of the initial vector, each point increases on the first iteration (and in the negative parts each point decreases) to get a monotone increasing (monotone decreasing) sequence of points for each n. That is, we need

$$v_n^{(0)} = \delta\phi(n) \leq q^{-1}\left(\delta\left(\phi(n-1) + \phi(n+1)\right)\right) = q^{-1}\left(v_{n-1}^{(0)} + v_{n+1}^{(0)}\right) = v_n^{(1)}.$$

I.e.

$$q\left[\delta\phi(n)\right] \leq \delta\left(\phi(n-1) + \phi(n+1)\right).$$

There exists a δ such that this relation is satisfied as long as the slope of the function on the right hand side, considered as a function of δ, is greater than the slope of the function on the left at $\delta = 0$. That is, the inequality can be satisfied if

$$q'(0)\phi(n) < \phi(n-1) + \phi(n+1).$$

Now expanding the sine functions gives

$$\phi(n-1) + \phi(n+1) = 2\phi(n)\cos\left(\frac{2\pi}{N}\right),$$

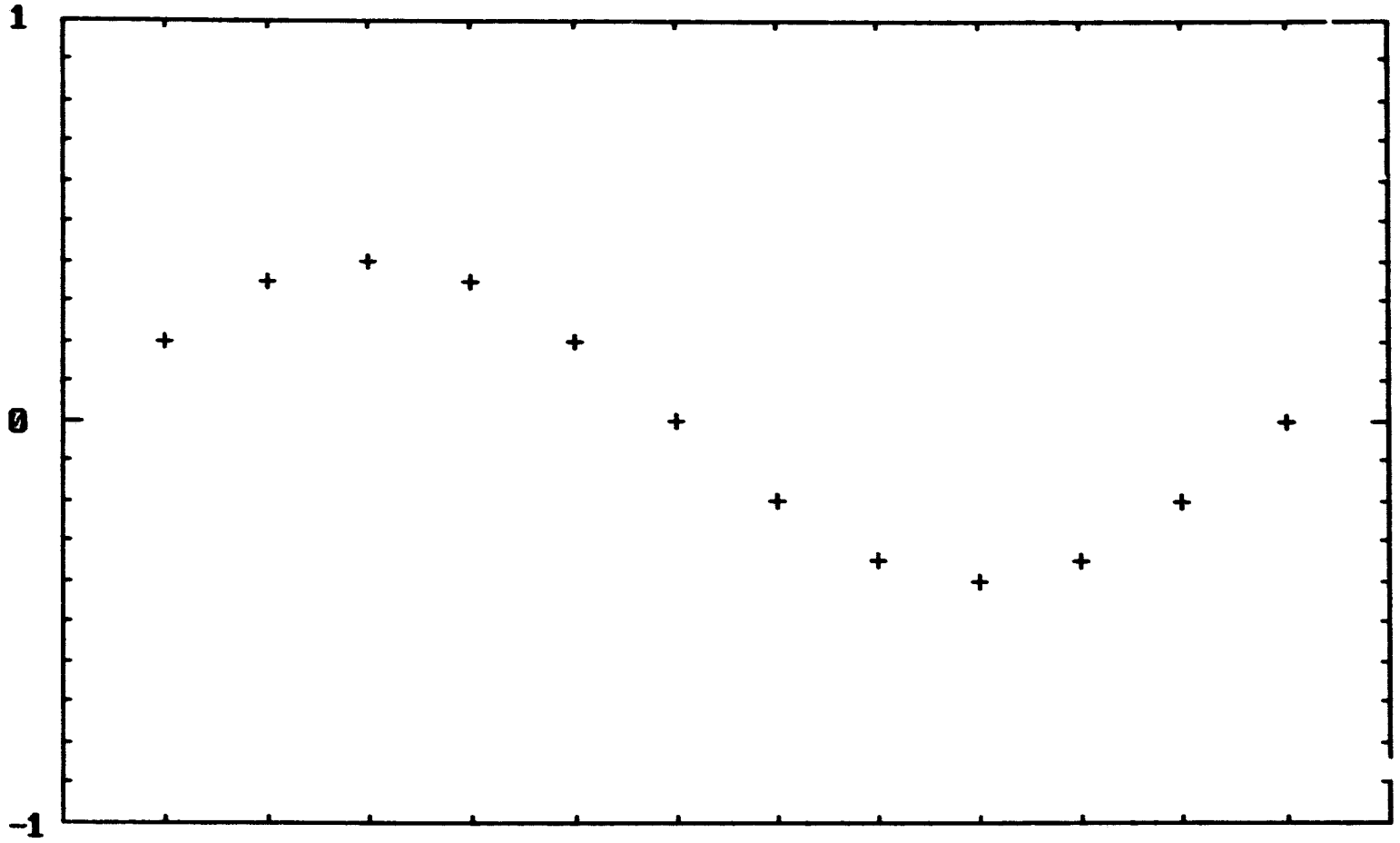

Figure 8.2a Initial data for iteration to a period 12 equilibrium (i.e., solution to equation (8.10)) with parameters as in Fig. 8.1. We use $\delta\phi(n) = 0.4 \sin\left(\frac{\pi n}{6}\right)$.

so the condition becomes

$$q'(0) < 2\cos\left(\frac{2\pi}{N}\right),$$

or equivalently,

$$N > \frac{2\pi}{\cos^{-1}\left(\frac{q'(0)}{2}\right)}.$$

If this is satisfied, we get a monotone increasing sequence $v_n^{(m)}$ for each n, $0 < n < \frac{N}{2}$. Recall that $q(v_s) = 2v_s$, where v_s is the positive solution to $f(v) = 0$ (see 8.18). So $v_s = q^{-1}(2v_s)$. Since q and therefore q^{-1} are increasing functions, we have by the iteration scheme (8.33) that $\left|v_n^{(m+1)}\right| < v_s$ as long as $\left|v_{n-1}^{(m)}\right| < v_s$ and $\left|v_{n+1}^{(m)}\right| < v_s$. Thus, if we take δ small enough so that $\left|v_n^{(0)}\right| < v_s$ for all n, then $\left|v_n^{(m)}\right| < v_s$ for all n and m. Then each monotone increasing sequence must converge (to v_n, say). For $\frac{N}{2} < n < N$, each sequence is monotone decreasing and converges to $v_n = -v_{N-n}$. The resulting vector v^* satisfies the equilibrium equation (8.15) at every point and is therefore a solution (e.g. Fig. 8.2b). □

We will require some properties of these equilibria. First, it is evident by the method of construction that these equilibria have symmetries. In particular,

$$v_0 = v_{\frac{N}{2}} = v_N = 0,$$

$$v_n = -v_{\frac{N}{2}+n},$$

and

$$v_n = v_{\frac{N}{2}-n}.$$

Now define

$$\Delta^2 v_i = v_{i-1} + v_{i+1} - 2v_i.$$

From equation (8.15), it is clear that for any equilibrium,

$$\Delta^2 v_i = q(v_i) - 2v_i = \frac{2}{\tau} f(v_i),$$

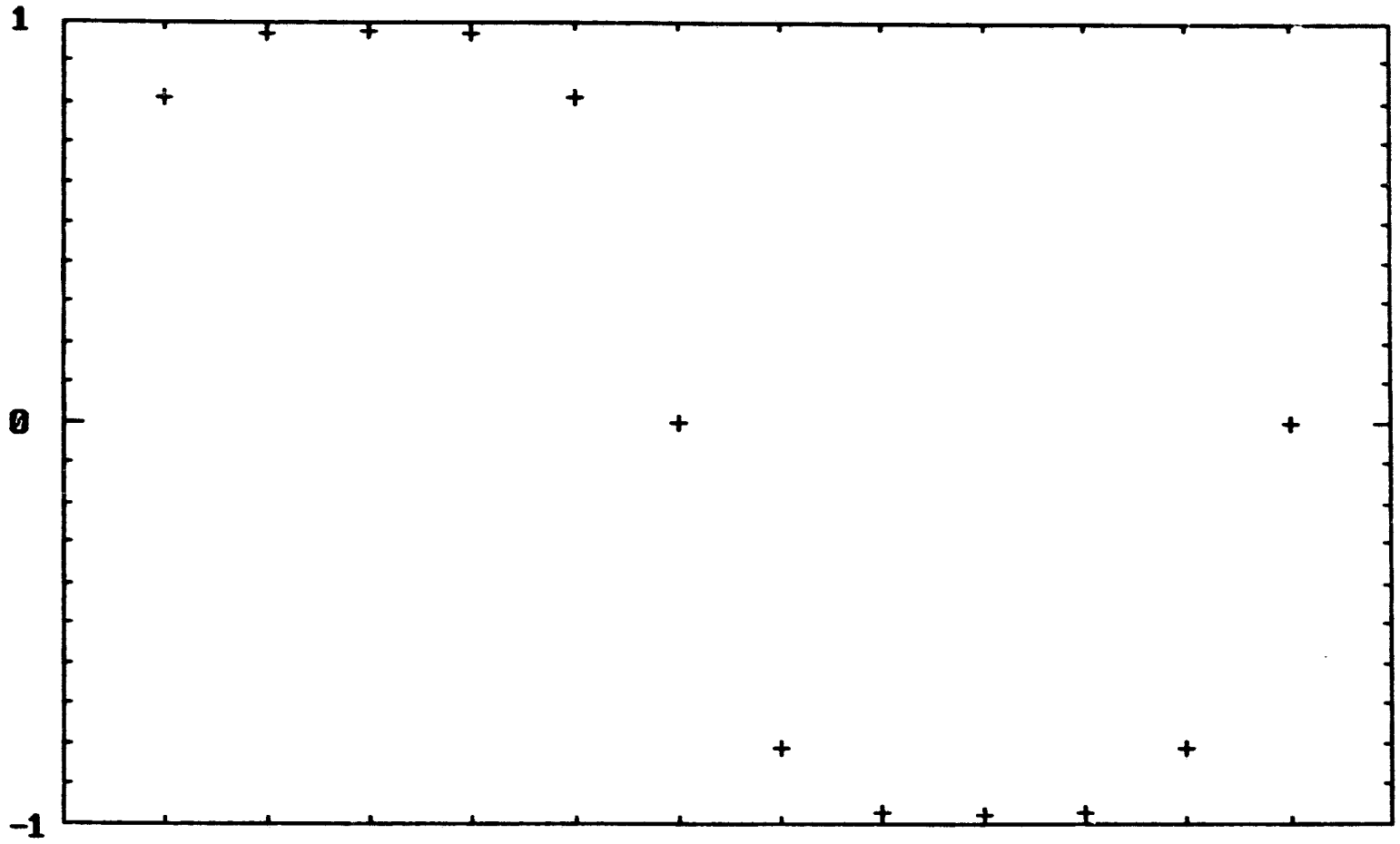

Figure 8.2b The period 12 equilibrium resulting from the iteration. This is a solution to equation (8.10) with parameters as in Fig. 8.1.

which is negative for $0 < v_i < v_s$ and positive for $-v_s < v_i < 0$. Thus, an equilibrium must be concave down where it is positive and concave up where it is negative. This concavity also implies that the equilibria found in Proposition 8.7 increase to a maximum and decrease to zero again from v_0 to $v_{\frac{N}{2}}$. Thus,

$$v_n \geq v_1$$

for $1 < n < \frac{N}{2}$ and in particular

$$v_2 > v_1$$

for even $N \geq 8$.

Furthermore, we know that f has a unique minimum on $(0, v_s)$, say at $v = \eta$, by (8.20) and since $f'(0) < 0$ (i.e. $q'(0) < 2$). For $v_i \geq \eta$, $f(v_i)$ is an increasing function of v_i. Thus, for an equilibrium, $v_i > v_j \geq \eta$ implies $f(v_i) > f(v_j)$ and therefore

$$\Delta^2 v_i > \Delta^2 v_j.$$

Proposition 8.8 *The solution to* (8.15) *of period* $N \geq 8$ *given by Proposition 8.7 exists and is stable if*

$$0 < q'(0) < 2\cos\left(\frac{2\pi}{N}\right) \tag{8.34}$$

and

$$q'(v_1^*) > \frac{2}{q'(0)} + 1\,. \tag{8.35}$$

Proof As for the period 6 equilibrium, stability is demonstrated by showing that the matrix of second partial derivatives, $P = \frac{\partial^2 E}{\partial v_i \partial v_j}[v^*]$, where v^* is the equilibrium, is positive definite. The matrix will be similar to that in (8.24), having $q'(v_i^*)$ on the diagonal in the i^{th} position and -1 in adjacent positions. Letting $\mathbf{x}$ be an arbitrary vector in $\mathbf{R}^N$, we have

$$\mathbf{x}^t P \mathbf{x} = \sum_{i=1}^{N} q'(v_i^*)x_i^2 - 2\sum_{i=1}^{N} x_i x_{i+1}\,. \tag{8.36}$$

In order to see when this must be positive, we need to express it as a sum of squares with positive coefficients, as in (8.26). The interaction terms in the last sum can be handled by including terms like $(x_i - x_{i+1})^2$ for each adjacent pair, and then the extra $2x_i^2$ for each point will have to be subtracted from $q'(v_i^*)x_i^2$. However, this will not work for the points where $v_i^* = 0$, i.e. when $i = \frac{kN}{2}$, for some integer k, since the equilibrium only exists for $0 < q'(0) < 2\cos\left(\frac{2\pi}{N}\right) < 2$, so that $(q'(0) - 2)\, x_i^2 < 0$. Thus, we handle the points $i = \frac{kN}{2}$ and their interactions with adjacent points separately to get the equivalent expression

$$\begin{aligned} \mathbf{x}^t P \mathbf{x} = & \sum_{i=\frac{kN}{2}} \sum_{j\sim i} \frac{q'(0)}{2} \left[x_i - \frac{2}{q'(0)} x_j \right]^2 + \frac{1}{2} \sum_{i \neq \frac{kN}{2}} \sum_{\substack{j\sim i \\ j \neq \frac{kN}{2}}} [x_i - x_j]^2 \\ & + \sum_{i=\frac{kN}{2}} \sum_{j\sim i} \left[q'(v_j^*) - \frac{2}{q'(0)} - 1 \right] x_j^2 + \sum_{|v_j^*| > |v_1^*|} \left[q'(v_j^*) - 2 \right] x_i^2 , \end{aligned} \tag{8.37}$$

where the sums over $j \sim i$ mean sums over the immediate neighbours of i. Thus, the second sum above is over all adjacent pairs where neither is a zero point and the last sum is over all points aside from the zero points and those adjacent to them. Now by (8.20), $vq''(v) > 0$, so that $q'(v)$ increases with $|v|$ and so (8.37) is positive if (8.35) is satisfied and $q'(v_2^*) > 2$. This last condition follows from (8.34) and (8.35), however. Note that $q'(v_2^*) > q'(v_1^*)$ since $|v_2^*| > |v_1^*|$, which is a property of the equilibria from the discussion following Proposition 8.7. Also, (8.34) gives $\frac{2}{q'(0)} > \frac{1}{\cos\left(\frac{2\pi}{N}\right)}$. So

$$q'(v_2^*) > q'(v_1^*) > \frac{1}{\cos\left(\frac{2\pi}{N}\right)} + 1 \geq 2 .$$

Thus, conditions (8.34) and (8.35) give the result. □

Example 8.1 (continued) For $q_1(v)$ given by (8.13) with $\hat{\tau} = \tau_0 = 2$ and $\beta = 1$, conditions (8.34) and (8.35) become

$$0 < \frac{1}{\gamma} < 2\cos\left(\frac{2\pi}{N}\right), \tag{8.38}$$

$$\frac{G'(v_1^*)}{\gamma} - 1 > 2\gamma . \tag{8.39}$$

The first of these (8.38) is always true with $\gamma > 1$, say (for $N \geq 6$). □

In order to establish the existence of these stable patterns for this example, it is necessary to show that (8.39) can also be satisfied. We do this with the help of some earlier results from [22]. First, we look at the difference between two equilibria, i.e. two solutions to (8.15).

Lemma 8.9 *Let* $\mathbf{u}$ *and* $\mathbf{v}$ *be solutions of* (8.15) *and let* $\mathbf{z} = \mathbf{u} - \mathbf{v}$. *Suppose* $u_m, v_m \geq \eta$ *and* $z_m \geq 0$ *for some* m, *then* $z_{m+1} \geq 2z_m - z_{m-1}$.

Proof From the properties of equilibria discussed after Proposition 8.7, $z_m > 0$ implies that $\Delta^2 u_m > \Delta^2 v_m$. That is, $u_{m+1} + u_{m-1} - 2u_m \geq v_{m+1} + v_{m-1} - 2v_m$, which when rearranged, gives the desired result. □

Lemma 8.10 *Let* $\mathbf{u}$ *and* $\mathbf{v}$ *be solutions of* (8.15) *and let* $\mathbf{z} = \mathbf{u} - \mathbf{v}$. *Suppose* $u_1, \ldots, u_m \geq \eta$, $v_1, \ldots, v_m \geq \eta$ *and* $z_1, \ldots, z_m \geq 0$. *Then*

$$z_{m+1} \geq (m+1)z_1 - mz_0 .$$

Proof Let k be an integer such that $1 \leq k \leq m-1$. Note that from Lemma 8.9, we have $-z_m \geq -\frac{1}{2}z_{m+1} - \frac{1}{2}z_{m-1}$. Using this,

$$(k+1)z_m - kz_{m-1} \geq (k+1)z_m - \frac{k}{2}z_m - \frac{k}{2}z_{m-2} = \left(\frac{k}{2}+1\right)z_m - \frac{k}{2}z_{m-2}$$

and then using Lemma 8.9 again on the first term,

$$(k+1)z_m - kz_{m-1} \geq \left(\frac{k}{2}+1\right)(2z_{m-1} - z_{m-2}) - \frac{k}{2}z_{m-2} = (k+2)z_{m-1} - (k+1)z_{m-2}.$$

Now apply this result $m-1$ times to $z_{m+1} \geq 2z_m - z_{m-1}$, with $k = 1, 2, \ldots, m-1$, in turn to get the result. □

Corollary 8.11 *If $z_1 \geq 0$ and $z_0 = 0$ in Lemma 8.10, then $z_2, \ldots, z_m \geq 0$ is automatic, since $z_{k+1} \geq (k+1)z_1$ for each k, $1 \leq k \leq m-1$. Therefore, the result holds for z_{m+1}, i.e.*

$$z_{m+1} \geq (m+1)z_1 .$$

□

Now suppose that we have two solutions to (8.15), one of even period M, call it **u**, and one of larger even period, say $N > M$, call it **v**. Suppose also that $u_0 = v_0 = 0$. We claim that $v_1 > u_1$ and therefore, that v_1 is an increasing function of N, at least when $u_1, v_1 \geq \eta$.

Proposition 8.12 *Let* **u** *and* **v** *be solutions to* (8.15) *of even period M and N respectively, with $6 \leq M < N$. Let $u_0 = v_0 = 0$. Suppose $u_1, v_1 \geq \eta$. Then $v_1 > u_1$.*

Proof Since $u_1 \geq \eta$ and **u** has even period M, we have $u_i \geq \eta$ for $1 \leq i < \frac{M}{2}$ (this is a property of the equilibria from the discussion following Proposition 8.12). Similarly, $v_i \geq \eta$ for $1 \leq i < \frac{N}{2}$. Now suppose that $u_1 \geq v_1$, so that $z_1 = u_1 - v_1 \geq 0$. Then we can apply Corollary 8.11 with $m = \frac{M}{2} - 1$, to show that $z_{M/2} \geq \frac{M}{2} z_1 \geq 0$. However, $u_{M/2} = 0$ and $v_{M/2} > 0$ so $z_{M/2} < 0$ and we have a contradiction. Thus $u_1 < v_1$ and since this is true for arbitrary even periods $M, N \geq 6$, v_1 is an increasing function of the period, N. □

This result can be applied to stability of periodic equilibria as follows. Since $q'(v)$ is an increasing function for positive v (8.20), $q'(v_1^*)$ is an increasing function of the period N by Proposition 8.12. Thus, if for some N, condition (8.35) is satisfied, then it will also be satisfied for all larger N. Also, increasing N increases the upper bound on $q'(0)$ in condition (8.34). So if, for a particular q, the existence of a stable equilibrium of even period N can be established, then the equilibria of larger even

period also exist and are stable.

Example 8.1 (continued) For the particular q given by equation (8.13), i.e. q_1, conditions (8.34) and (8.35) become (8.38) and (8.39) but these are true for period $N = 6$ as shown in the previous section. Thus the equilibria for all even periods $N \geq 6$ exist and are stable for q_1.

Example 8.2 (continued) For q_2 given by (8.19), there was no stable period 6 equilibrium so it would be necessary to find one of larger period to get the large scale stable patterns in this case.

8.6 Patterns in d-dimensions.

The above results can be extended to two or more dimensions without much difficulty. There is, of course, a larger choice of patterns that can be examined for stability. For example, we can obtain a d-dimensional analogy to Proposition 8.7, for $0 < q'(0) < 2d$ either by starting with an initial function

$$\phi(\mathbf{n}) = \prod_{r=1}^{d} \sin\left(\frac{2\pi n_r}{N}\right) \tag{8.40}$$

for $\mathbf{n}$ a grid point with coordinates n_r, which in 2 dimensions will produce a checkerboard pattern of positive and negative square regions, or with

$$\phi(\mathbf{n}) = \sin\left(\frac{2\pi \sum_{r=1}^{d} n_r}{N}\right), \tag{8.41}$$

which in 2 dimensions will produce a pattern of diagonal ridges and valleys. In either case the condition for existence of this equilibrium is

$$q'(0) < 2d\cos\left(\frac{2\pi}{N}\right) \text{ or } N > \frac{2\pi}{\cos^{-1}\left(\frac{q'(0)}{2d}\right)}. \tag{8.42}$$

And in the case of the second type of pattern described above, the condition for stability of the equilibrium (denoted $\mathbf{v}^*$) is

$$\begin{aligned}\mathbf{x}^t P\mathbf{x} &= \sum_i q'(0)x_i^2 - 2\sum_i \sum_{j\sim i} x_i x_j \\ &= \sum_{v_i^*=0} \sum_{j\sim i} \frac{q'(0)}{2d}\left[x_i - \frac{2d}{q'(0)}x_j\right]^2 + \frac{1}{2}\sum_{v_i^*\neq 0}\sum_{\substack{j\sim i \\ v_j^*\neq 0}} [x_i - x_j]^2 \\ &\quad + \sum_{|v_i^*|=|v_1^*|}\left[q'(v_i^*) - \frac{2d}{q'(0)} - d\right]x_j^2 + \sum_{|v_i^*|>|v_1^*|}[q'(v_i^*) - 2d]\,x_i^2.\end{aligned}$$

This is analogous to (8.26) and (8.37). So a sufficient condition for stability is that the equilibrium exists and

$$q'(v_1^*) > \frac{2d}{q'(0)} + d, \tag{8.43a}$$

$$q'(v_2^*) > 2d. \tag{8.43b}$$

Again, for large N we expect these conditions to be satisfied. Other types of patterns are, of course, possible.

8.7 Discretizations of the general reaction-diffusion equation.

In equation (6.23) we can discretize using

$$\begin{aligned}&[\mu(x_m, y)v(y)]_{yy} \\ &\quad \sim \frac{1}{\varepsilon^2}\left[\mu(x_m, y_{m-1})v(y_{m-1}) + \mu(x_m, y_{m+1})v(y_{m+1}) - 2\mu(x_m, y_m)v(y_m)\right]\end{aligned}$$

and if we denote $\mu(x_m, x_n)$ by $\mu_{m,n}$ then the finite difference equation becomes

$$\dot{v}_m = \frac{\gamma}{G'(v_m)}\left[\frac{\hat{\tau}}{2}\left(\mu_{m,m-1}v_{m-1} + \mu_{m,m+1}v_{m+1}\right) + (\tau_0 - \hat{\tau})\mu_{m,m}v_m - \alpha\frac{G(v_m)}{\gamma}\right]. \tag{8.44}$$

The equilibrium equation becomes

$$\frac{\hat{\tau}}{2}\left(\mu_{m,m-1}v_{m-1}+\mu_{m,m+1}v_{m+1}\right)+(\tau_0-\hat{\tau})\mu_{m,m}v_m=\alpha\frac{G(v_m)}{\gamma}$$

and while the trivial equilibrium ($v\equiv 0$) still exists, the constant space equilibria do not, as each point is now modified differently by μ. If $\hat{\tau}=\tau_0=2$ and $\alpha=\beta=1$, this simplifies to

$$\mu_{m,m-1}v_{m-1}+\mu_{m,m+1}v_{m+1}=\frac{G(v_m)}{\gamma}$$

which might be amenable to the same type of analysis as in the previous sections. This remains to be pursued.

8.8 Stable patterns and metastability.

Cottet's equation is an approximation in continuous space to a system of ordinary differential equations having large scale stable patterns. The approximation results of Chapter 6 suggest that Cottet's equation must have solutions approximating these stable states for some time, although they cannot be stable themselves. The Hopfield networks that approximate Cottet's equation can, of course, also have stable patterns. The standard discretizations of Cottet's equation analyzed in this chapter are, in fact, simple (and symmetric) examples of the Hopfield network equations. Discretizations of the more general reaction-diffusion equation discussed in Chapter 6 potentially allow a richer range of behaviours.

Formally, the central difference approximation for the v_{xx} term in the one-dimensional equation has error $O(\varepsilon^2)$, so $\varepsilon^2 v_{xx}$ is an $O(\varepsilon^4)$ approximation to $(v_{m+1}+v_{m-1}-2v_m)$. In a fixed time interval, the solution to the reaction-diffusion equation should then approximate the solution to the system of ODEs with error of $O(\varepsilon^4)$. Thus, for small ε, solutions to Cottet's equation that start near a stable pattern of

the system of ODEs should stay near it for some time. The length of time may be fixed and ε chosen small enough so that solutions to Cottet's equation are forced to change very little for long times. This is at least suggestive of the metastability that occurs in the analysis of the Ginzburg-Landau or Allen-Cahn type of equation as studied in [8,9,10].

Random initial conditions for the discrete space equations may still lead to metastable patterns of transition layers since the stable patterns seem to depend on equal spacing of layers. Figs. 8.3a,b show an example of the short term behaviour of solutions to equations (8.2) with random initial conditions. Fig. 8.4 shows an example for the generalized equation (8.44) with similar initial conditions.

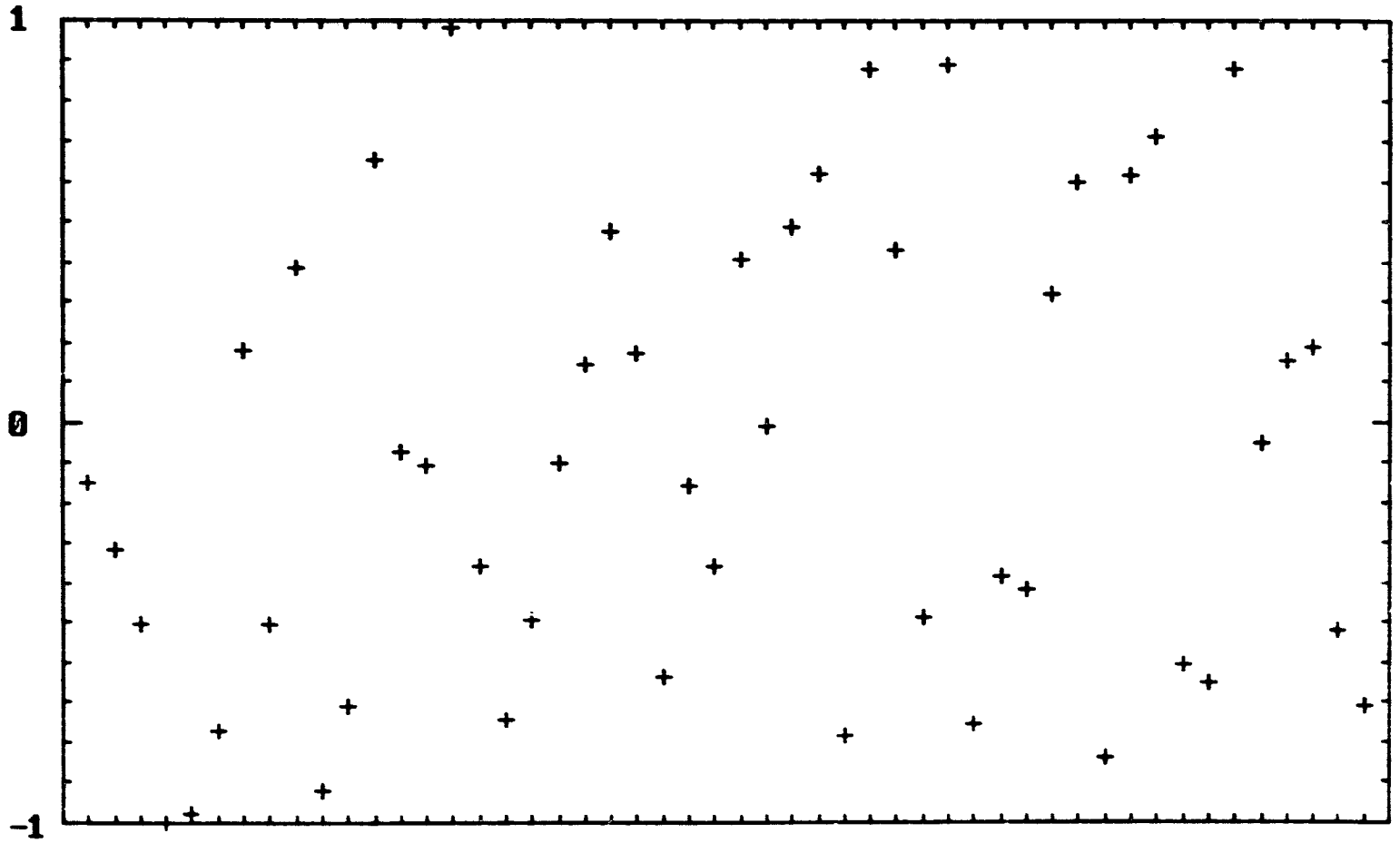

Figure 8.3a Random period 50 initial data for Cottet's equation discretized (8.2) with parameters as in Fig. 8.1.

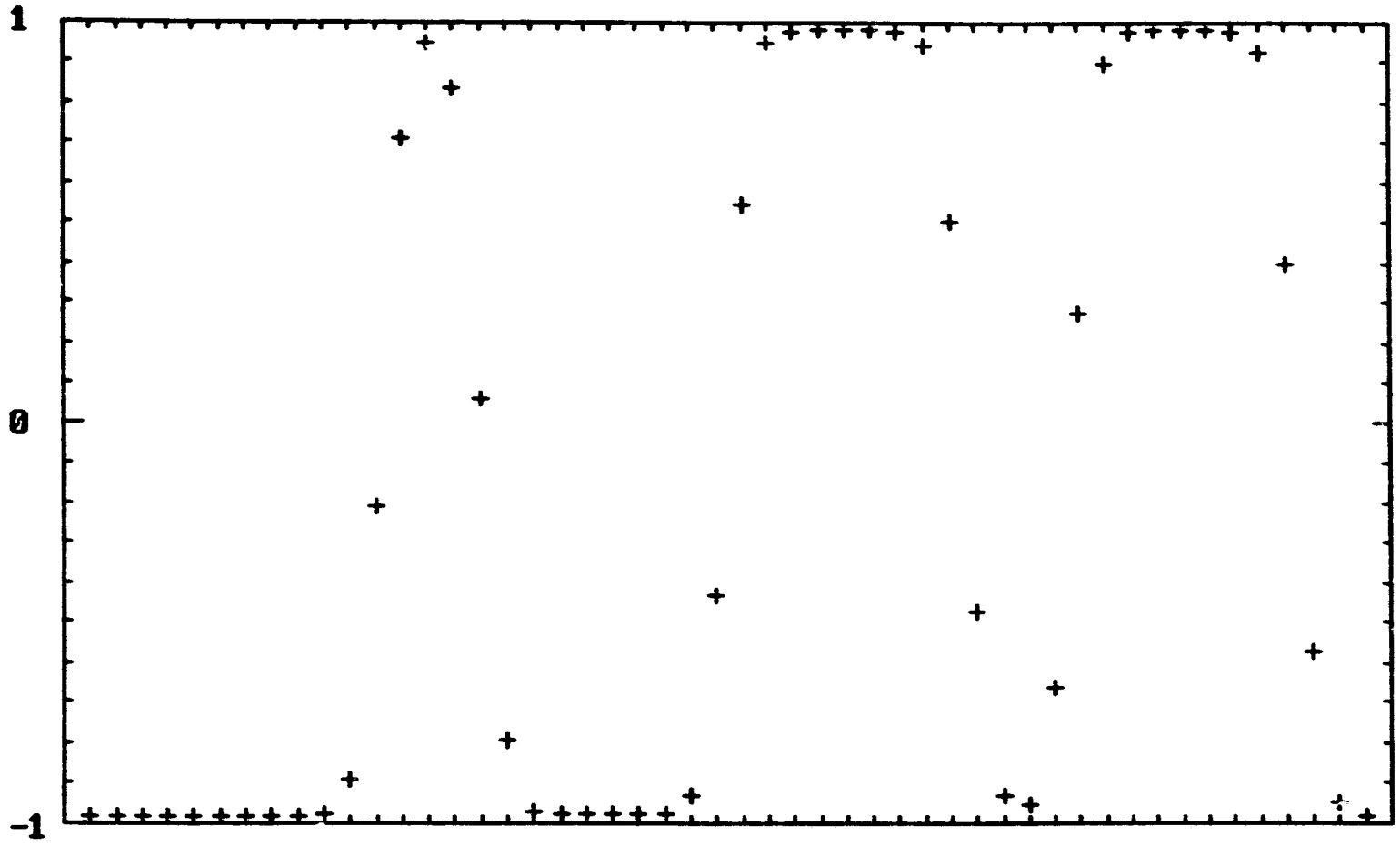

Figure 8.3b Transition layers formed at $t = 5$ from the random initial data in Fig. 8.3a.

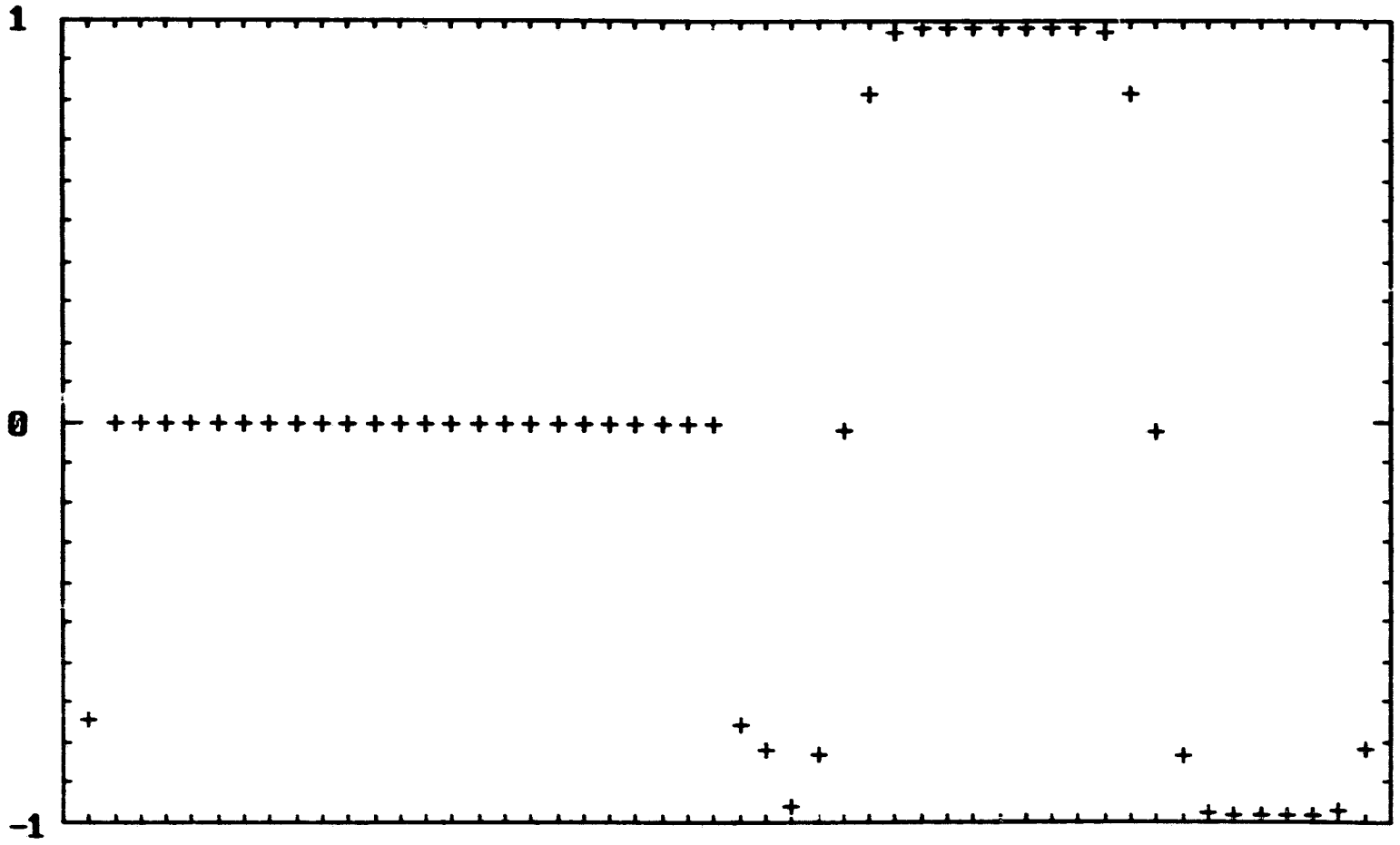

Figure 8.4 Solution for the general reaction-diffusion equation discretized (8.44), at $t = 10$, with $\mu(x_i, x_j) = \frac{j}{50}$, $j = 1, \ldots, 50$, for all i, other parameters as in Fig. 8.1 and random initial data.

9. Conclusions

The main results have been summarized already in Section 1.4. Here we discuss briefly some of the implications of these results, both from the neural network point of view and from the mathematical point of view.

9.1 Implications for neural network theory.

Some of the implications of these results for neural networks are as follows.

1. Connectivities that allow approximation by Cottet's reaction-diffusion equation imply that the network dynamics will be fairly simple. There is an energy functional, so behaviour must be convergent. Regions that initially have predominantly high or predominantly low activity will be enhanced (i.e. will be forced to a fully active or fully inactive state) with sharp transitions between them. Metastable solutions will be as good as stable ones for neural networks. This has useful applications in image processing for smoothing while simultaneously enhancing contrast. Connectivities that do not fit Cottet's equation but allow approximation by our more general reaction-diffusion equation will potentially have more complex behaviour. For example, solutions may decay to zero in some regions while forming transition layers in others, possibly with varying time scales. The range of possibilities is still expected to be somewhat limited. In some cases, there will not be an energy functional, so there is the possibility of non-convergent behaviour. However, the evidence for ubiquitous complex behaviour in biological neural systems leads us to draw the conclusion that the fine structure of neural connections *is* significant in their operation. The belief expressed by Traub and Miles that "the brain cares which cells are active, not just how many are active" [66, p.204] seems justified.

2. The theorems for approximation of Hopfield nets by reaction-diffusion equa-

tions suggest the value of significant inhibition for interesting (non-PDE-like) behaviour. They also suggest that the kind of close interplay between excitatory and inhibitory neurons that exists in Kwan's model (Chapter 4) is a key to chaotic behaviour. Finally, they suggest the possibility that long-distance (macroscopic) connections between neurons could lead to different behaviours.

3. Input-driven bifurcation allows transition from chaotic to more regular behaviour and can be used as a way of 'recognizing' and responding to certain inputs. 'Unrecognized' inputs result in continued chaotic behaviour. When the input is removed the network returns to a chaotic state. This can be done in various models, including one based on the Lorenz system of equations and a Hopfield network with a particular type of structure of alternating excitatory and inhibitory neurons. The latter appears more amenable to analysis, since the inputs required to remove or introduce chaos can be controlled more easily, and since there is no concern for boundedness of solutions when many units are joined together as there is for the Lorenz systems.

4. Unsupervised networks can by implemented in a number of ways. The chaotic systems of Chapter 4 run continuously and simply respond to input. This avoids the conventional idea of externally setting initial data and evolving to a fixed state and is thus more natural. Learning has not been explored as yet for these systems. The integro-differential equation model developed in Chapter 5 also operates in an unsupervised manner and does allow at least a simple form of learning. It also allows a natural distinction to be made between the learning and retrieval processes via duration of input.

9.2 Mathematical implications.

Some of the mathematical implications of these approaches are as follows.

1. Mathematically, the relationship between discrete space and analogous continuous space systems is of great interest and of some subtlety. It is evident from the research conducted here that smooth connectivity functions such as are required for the PDE approximation theorems of Chapter 6 prevent some interesting types of behaviour. If more interesting behaviour is desired, then either a discrete space model should be used or perhaps a continuous space model with discontinuous connectivites such as the characteristic functions of Chapter 5.

2. Continuous systems with metastable solutions may closely approximate discrete space systems with stable solutions. This is true for both standard finite difference discretizations of reaction-diffusion equations as discussed in Chapter 8, and Hopfield network discretizations. Since the reaction-diffusion equations also have metastable solutions that are not near stable states of discretizations, these discretizations can also be expected to have 'metastable' solutions.

3. Sustained external forcing terms provide one straightforward way to introduce bifurcations into dynamical systems so that dynamics can be changed from chaotic to regular. The particular regular behaviour resulting may depend on the forcing term.

9.3 Further directions.

This research has involved several approaches to neural network theory, explored in varying depths.

The approximation of Hopfield network dynamics by reaction-diffusion equa-

tions was developed in depth, but some gaps remain. For the sake of completeness, it is desirable to deal more rigorously with the problem of convergence of the approximation in the region of transition layers. The question is to determine how much the deviations from the approximation can spread outside these transition layers and how fast. The approximation is not completely conclusive until this is done (although we have strong reasons to suspect that the effect is very slow). Furthermore, the general reaction-diffusion equation that we have derived is interesting in its own right and deserves further analysis. The role of the dependence on the space variable, x, in equation (6.16) is not yet clear. The study of this equation could shed more light on the corresponding Hopfield network dynamics and could potentially have applications in image processing.

However, the results suggest that in order to develop new neural network models that have the properties set forth in Chapter 3, we should look at discrete space models with significant inhibition.

Thus, our tentative inquiries into the chaotic models of Chapters 4 and the integro-differential equation models of Chapter 5 could be pursued much further. In particular, the model of Kwan and his collaborators from Section 4.4 with the additional insights provided here shows some promise. We believe that it is of value to continue the program commenced here, to design more natural neural network models involving unsupervised dynamical behaviour, unsupervised learning and input-driven bifurcation from chaotic to regular behaviour. The discretizations of Cottet's reaction diffusion equations also seem to form long-lasting states with transition layers which eventually collapse (as in Figs. 8.3a,b). Discretizations of the more general reaction-diffusion equation derived here appear to behave similarly, but on regions where μ is small, the solution decays to zero (as in Fig. 8.4). More analysis of these equations could shed light on the analysis of the generalized

reaction-diffusion equation and the Hopfield nets it approximates and may have applications in image processing extending those of Cottet's equation.

Bibliography

[1] R. A. Adams, *Sobolev Spaces*, Academic Press, New York, 1975.

[2] S. M. Allen and J. W. Cahn, *A microscopic theory for antiphase boundary motion and its application to antiphase domain coarsening*, Acta Metallurgica, 27 (1979), pp. 1085-1095.

[3] S.-i. Amari, *Mathematical theory of self-organization in neural nets*, in Organization of Neural Networks: Structures and Models, W. Von Seelen, G. Shaw and U. M. Leinhos, eds., VCH publishers, New York, 1988, pp. 399-413.

[4] S.-i. Amari, *Dynamical stability of formation of cortical maps*, in Dynamic Interactions in Neural Networks: Models and Data, M. A. Arbib and S. Amari, eds., Springer-Verlag, Berlin, New York, 1989, pp. 15-34.

[5] D. J. Amit, *Modeling Brain Function: The World of Attractor Neural Networks*, Cambridge University Press, Cambridge, 1989.

[6] S. Becker, *Unsupervised learning procedures for neural networks*, International Journal of Neural Systems, 2 (1991), pp. 17-33.

[7] J. W. Cahn, S.-N. Chow and E. S. Van Vleck, *Spatially discrete nonlinear diffusion equations*, The Rocky Mountain Journal of Mathematics, to appear.

[8] J. Carr and R. L. Pego, *Very slow phase separation in one dimension*, in PDEs and Continuum Models of Phase Transitions, Lecture Notes in Physics 344, M. Rascle, *et al.*, eds., Springer-Verlag, Berlin, New York, 1989, pp. 216-226.

[9] J. Carr and R. L. Pego, *Metastable patterns in solutions of* $u_t = \varepsilon^2 u_{xx} - f(u)$, Communications on Pure and Applied Mathematics, 42 (1989), pp. 523-576.

[10] J. Carr and R. Pego, *Invariant manifolds for metastable patterns in* $u_t = \varepsilon^2 u_{xx} - f(u)$, Proceedings of the Royal Society of Edinburgh, Section A, 116 (1990), pp. 133-160.

[11] A. H. Cohen, *Effects of oscillator frequency on phase-locking in the lamprey central pattern generator*, Journal of Neuroscience Methods, 21 (1987), pp. 113-125.

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

[13] A. C. C. Coolen and Th. W. Ruijgrok, *Image evolution in Hopfield networks*, Physical Review A, 38 (1988), pp. 4253-4255.

[14] G.-H. Cottet, *Modèles de réaction-diffusion pour des réseaux de neurones stochastiques et déterministes*, Comptes Rendus de l'Académie des Sciences, Paris, 312-I (1991), pp. 217-221.

[15] G.-H. Cottet, *Diffusion approximation on neural networks and applications for image processing*, in Proceedings of the Sixth European Conference on Mathematics in Industry, F. Hodnett, ed., B.G. Teubner, Stuttgart, 1992.

[16] G.-H. Cottet and L. Germain, *Image processing through reaction combined with nonlinear diffusion*, Mathematics of Computation, 61 (1993), pp. 659-673.

[17] P. Degond and S. Mas-Gallic, *The weighted particle method for convection-diffusion equations. Part 1: The case of an isotropic viscosity*, Mathematics of Computation, 53 (1989), pp. 485-507.

[18] J. S. Denker, *Neural network models of learning and adaptation*, Physica D, 22 (1986), pp. 216-232.

[19] B. Derrida, E. Gardner and A. Zippelius, *An exactly soluble asymmetric neural network model*, Europhysics Letters, 4 (1987), pp. 167-173.

[20] R. L. Devaney, *An Introduction to Chaotic Dynamical Systems* (2nd. ed.), Addison-Wesley, New York, 1989.

[21] N. Dimopoulos, *A study of the asymptotic behavior of neural networks*, IEEE Transactions on Circuits and Systems, 36 (1989), pp. 687-694.

[22] R. Edwards, *A system of nonlinear differential equations arising from the dynamic Ising model*, MSc thesis, Heriot-Watt University, Edinburgh, 1990.

[23] G. B. Ermentrout and N. Kopell, *Frequency plateaus in a chain of weakly coupled oscillators. I.*, SIAM Journal on Mathematical Analysis, 15 (1984), pp. 215-237.

[24] N. W. Evans, R. Illner and H. C. Kwan, *On information-processing abilities of chaotic dynamical systems*, Journal of Statistical Physics, 66 (1992), pp. 549-561.

[25] G. B. Folland, *Real Analysis: Modern Techniques and their Applications*, Wiley, New York, 1984.

[26] W. J. Freeman, *Mass Action in the Nervous System*, Academic Press, New York, 1975.

[27] W. J. Freeman, *Simulation of chaotic EEG patterns with a dynamic model of the olfactory system*, Biological Cybernetics, 56 (1987), pp. 139-150.

[28] S. Grossberg, *Nonlinear neural networks: principles, mechanisms, and architectures*, Neural Networks, 1 (1988), pp. 17-61.

[29] D. O. Hebb, *The Organization of Behaviour*, Wiley, New York, 1949.

[30] O. Hendin, D. Horn and M. Usher, *Chaotic behavior of a neural network with dynamical thresholds*, International Journal of Neural Systems, 1 (1991), pp.

327-335.

[31] D. Henry, *Geometric Theory of Semilinear Parabolic Equations*, Springer-Verlag, Berlin, New York, 1982.

[32] G. E. Hinton and T. J. Sejnowski, *Learning and Relearning in Boltzmann Machines*, in Parallel Distributive Processing: Explorations in the Microstructure of Cognition (vol. 1), D. E. Rumelhart and J. L. McClelland, eds., MIT Press, Cambridge, Mass., 1986, pp. 282-317.

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

[34] J. J. Hopfield, *Neural networks and physical systems with emergent collective computational abilities*, Proceedings of the National Academy of Sciences, U.S.A., 79 (1982), pp. 2554-2558.

[35] J. J. Hopfield, *Neurons with graded response have collective computational properties like those of two-state neurons*, Proceedings of the National Academy of Sciences, U.S.A., 81 (1984), pp. 3088-3092.

[36] J. J. Hopfield and D. W. Tank, *Computing with neural circuits: A model*, Science, 233 (1986), pp. 625-633.

[37] D. Horn and M. Usher, *Neural networks with dynamical thresholds*, Physical Review A., 40 (1989), pp. 1036-1044.

[38] D. Horn and M. Usher, *Excitatory-inhibitory networks with dynamical thresholds*, International Journal of Neural Systems, 1 (1990), pp. 249-257.

[39] D. Kleinfeld and H. Sompolinsky, *Associative neural network model for the generation of temporal patterns: Theory and application to central pattern generators*, Biophysical Journal, 54 (1988), pp. 1039-1051.

[40] N. Kopell and G. B. Ermentrout, *Symmetry and phaselocking in chains of weakly coupled oscillators*, Communications on Pure and Applied Mathematics, 39 (1986), pp. 623-660.

[41] H.-O. Kreiss and J. Lorenz, *Initial-Boundary Value Problems and the Navier-Stokes Equations*, Academic Press, New York, 1989.

[42] H. C. Kwan, *Network relaxation as behavioral action*, Technical Reports on Research in Biological and Computational Vision, No. RBCV-TR-88-26, University of Toronto, 1988.

[43] H. C. Kwan, T. H. Yeap, B. C. Jiang and D. Borrett, *Neural network control of simple limb movements*, Canadian Journal of Physiology and Pharmacology, 68 (1990), pp. 126-130.

[44] O. A. Ladyženskaja, V. A. Solonnikov and N. N. Ural'ceva, *Linear and Quasi-*

linear Equations of Parabolic Type, American Mathematical Society, Providence, RI, 1968.

[45] R. P. Lippman, *An introduction to computing with neural nets*, IEEE ASSP Magazine, 4 (no. 2, April, 1987), pp. 4-22.

[46] S. Mas-Gallic and P. A. Raviart, *A particle method for first-order symmetric systems*, Numerische Mathematik, 51 (1987), pp. 323-352.

[47] H. Matano, *Asymptotic behavior and stability of solutions of semilinear diffusion equations*, Publications of the Research Institute for Mathematical Sciences, Kyoto University, 15 (1979), pp. 401-454.

[48] A. N. Michel and J. A. Farrell, *Associative memories via artificial neural networks*, IEEE Control Systems Magazine, 10 (no. 3, April, 1990), pp. 6-17.

[49] M. Minsky and S. Papert, *Perceptrons*, MIT Press, Cambridge, Mass., 1969.

[50] Parisi, G., *Asymmetric neural networks and the process of learning*, Journal of Physics A, 19 (1986), pp. L675-L680.

[51] A. J. Priesol, D. S. Borrett and H. C. Kwan, *Dynamics of a chaotic neural network in response to a sustained stimulus*, Technical Reports on Research in Biological and Computational Vision, No. RBCV-TR-91-38, Department of Computer Science, University of Toronto, 1991.

[52] P. A. Raviart, *An analysis of particle methods*, in Numerical Methods in Fluid Dynamics, F. Brezzi, ed., Springer-Verlag, Berlin, New York, 1985, pp. 243-324.

[53] S. Renals, *Chaos in neural networks*, in Neural Networks: EURASIP Workshop 1990, Sesimbra, Portugal, February 15-17, 1990, Proceedings, L. B. Almeida and C. J. Wellekens, eds., Springer-Verlag, Berlin, New York, 1990, pp. 90-99.

[54] S. Renals and R. Rohwer, *A study of network dynamics*, Journal of Statistical Physics, 58 (1990), pp. 825-848.

[55] L. G. Reyna and M. J. Ward, *Resolving weak internal layer interactions for the Cahn-Allen Equation*, European Journal of Applied Mathematics, to appear.

[56] F. Rosenblatt, *Principles of Neurodynamics*, Spartan Books, Washington D. C., 1962.

[57] Th. W. Ruijgrok and A. C. C. Coolen, *Generalized Hopfield models for neural networks*, Acta Physica Polonica B, 21 (1990), pp. 379-389.

[58] D. E. Rumelhart, G. E. Hinton and R. J. Williams, *Learning internal repre-*

sentations by error propagation, in Parallel Distributive Processing: Explorations in the Microstructure of Cognition (vol. 1), D. E. Rumelhart and J. L. McClelland, eds., MIT Press, Cambridge, Mass., 1986, pp. 318-362.

[59] D. E. Rumelhart and J. L. McClelland, eds. *Parallel Distributive Processing: Explorations in the Microstructure of Cognition* (2 vols.), MIT Press, Cambridge, Mass., 1986.

[60] C. A. Skarda and W. J. Freeman, *How brains make chaos in order to make sense of the world*, Behavioral and Brain Sciences, 10 (1987), pp. 161-195.

[61] H. Sompolinsky, A. Crisanti and H. J. Sommers, *Chaos in random neural networks*, Physical Review Letters, 61 (1988), 259-262.

[62] C. Sparrow, *The Lorenz Equations: Bifurcations, Chaos, and Strange Attractors*, Springer-Verlag, New York, Berlin, 1982.

[63] S. I. Sudharsanan and M. K. Sundareshan, *Equilibrium characterization of dynamical neural networks and a systematic synthesis procedure for associative memories*, IEEE Transactions on Neural Networks, 2 (1991), pp. 509-521.

[64] B. Tirozzi and M. Tsodyks, *Chaos in highly diluted neural networks*, Europhysics Letters, 14 (1991), pp. 727-732.

[65] G. Toulouse, *Perspectives on neural network models and their relevance to neurobiology*, Journal of Physics A, 22 (1989), pp. 1959-1968.

[66] R. D. Traub and R. Miles, *Neuronal Networks of the Hippocampus*, Cambridge University Press, Cambridge, 1991.

[67] M. J. Ward, *Metastable patterns, layer collapses, and coarsening for a one-dimensional Ginzburg-Landau equation*, Studies in Applied Mathematics, to appear.