Phase plane analysis of neural nets

Bard Ermentrout

March 4, 2002

Introduction

Models of neural networks often involve the solutions to differential equations which describe the time evolution of these complex systems. The dynamical behavior of these networks ranges from the convergence to an equilibrium (generally desired in connectionist applications) to oscillatory behavior (in models of central pattern generators and bursting) through possibly chaotic behavior. There are many ways to analyze these models; the most used techniques are via simulation. In this chapter I will give an overview of an alternative technique for studying the qualitative behavior of small systems of interacting neural networks. One form that the models take is (Ellias &

\[ \tau_i \frac{dx_i}{dt} = -x_i + f_i(\sum_{j=1}^{n} w_{ij} x_j + s_i) \quad i = 1, \ldots, n \]  

(1)

where \( x_i \) represents the activity or firing rate of the \( i^{th} \) neuron, \( \tau_i \) is the time constant, \( w_{ij} \) are the connection weights, \( s_i \) are inputs, and \( f_i \) are typically saturating nonlinear functions that have the form shown in Figure 1. That is, the nonlinear functions are increasing and bounded. Some typical examples are:

\[ f(x) = \tanh(x) \]  

(2)

\[ f(x) = \tan^{-1}(x) \]  

(3)

\[ f(x) = \frac{1}{1 + \exp(-x)} \]  

(4)

Often, a slightly different form of (1) is chosen where the nonlinearities are inside the sums. The transformation from one to the other is elementary and all of the following holds for either type of model.

A complete analysis of networks of the form (1) is obviously impossible. However, if \( n \leq 2 \) then a fairly complete description of (1) can be given. Thus, the goal of this chapter is to introduce the reader to the qualitative theory of differential equations in the plane. In particular, I will analyze two neuron networks that consist of (i) two excitatory cells, (ii) two in-
hibitory cells, and (iii) an excitatory and inhibitory cell. The advantages of restricting the analysis to these small networks are the special topology of the plane, the completeness of the analysis possible, and finally the ease of exposition. Indeed, an overview of nonlinear dynamics can be obtained through these simple examples. Beer (1995) has attempted to exhaustively study the dynamics in the case $n = 2$ and gives a nearly complete overview of the possible types of behavior that can be expected. However, he does miss several interesting examples (Ermentrout, 1998, pp 371-373). Another more general approach for the analysis of large numbers of coupled systems is to use bifurcation methods which enable one to reduce the dimensionality of the resulting equations and then apply techniques such as those used here. While planar systems may seem to be a rather extreme simplification, there is some justification for it. For example, in some local cortical circuits, there is no structure in the connectivity and there are essentially two types of neurons: excitatory and inhibitory. Thus, we can view the simple planar system as representing a population of coupled excitatory and inhibitory neurons. This approach was used successfully (Pinto et al, 1996) to study cortical processing in the rodent somatosensory system and to explain effects of altering inhibitory interneurons in the hippocampus (Tsodyks et al 1997).
The approach of this chapter is not restricted to neural networks and can be applied to a variety of other systems such as positive-feedback biochemical models (Segel, 1984), activator-inhibitor systems (Edelstein-Keshet, 1988), population and disease models (Murray, 1989), and membrane models of the action potential (Rinzel and Ermentrout, 1998). The techniques are powerful and provide insights into the behavior of these systems that would otherwise only be accessible through simulation. Computational methods are a very powerful adjunct to this type of analysis and together with the qualitative analysis of this chapter enable the researcher to understand his or her system.

In the next section, I will describe a pair of neurons coupled with mutual inhibition and mutual excitation. The penultimate section is devoted to a summary of the rich behavior of an excitatory-inhibitory pair. Finally, some comments on numerical methods and software close the review. In the chapter on Bifurcation of neural nets, a systematic analysis of a particular case is given in order to illustrate alternate techniques.
Two coupled cells of the same type.

In this section, we analyze the behavior of two cells that act via mutual inhibition or mutual excitation. I will use phase plane analysis to draw a complete picture of the phase-space.

General considerations

Before analyzing the two-component neural network, I will first give a brief description of phase-plane techniques in general. Consider a planar differential equation:

\begin{align}
    x' &= f(x, y) \tag{5} \\
    y' &= g(x, y) \tag{6}
\end{align}

At each point \((x_0, y_0)\) there is a solution \((x(t), y(t))\) such that \((x(0), y(0)) = (x_0, y_0)\) and such that the tangent to the trajectory is \((f(x(t), y(t)), g(x(t), y(t)))\). Thus, at any point in the plane, one can easily determine the direction of the trajectory by simply evaluating \(f\) and \(g\) at that point. These directions enable one to paint a qualitative picture of the dynamics of (5,6); i.e., I can determine where \(x\) and \(y\) are increasing and decreasing with time. The most crucial points are those values of \(x\) and \(y\) at which the direction of the
trajectory changes. Thus, setting \( f(x, y) = 0 \) defines a curve in the plane where \( x \) does not change and breaks the plane into regions where \( x \) is either increasing or decreasing. This curve is called the \textit{x-nullcline}. The curve \( g(x, y) = 0 \) defines the \textit{y-nullcline}. The two curves together usually break the plane into regions of 4 distinct types; (i) \( x \) and \( y \) are increasing, (ii) \( x \) and \( y \) are decreasing, (iii) \( x \) increases and \( y \) decreases and (iv) \( x \) decreases and \( y \) increases. The intersection of the two nullclines occurs at points where both \( x \) and \( y \) are not changing; that is at \textit{equilibria} or \textit{rest states} of (5,6).

The behavior of trajectories away from equilibria is straightforward and is found by simply looking at the signs of \( f \) and \( g \). Near the equilibria, one can look at the linearization of \((f, g)\) about the equilibrium. This results in a two-by-two matrix called the \textit{Jacobian}:

\[
A = \begin{pmatrix}
\frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \\
\frac{\partial g}{\partial x} & \frac{\partial g}{\partial y}
\end{pmatrix} \equiv \begin{pmatrix}
a & b \\
c & d
\end{pmatrix}
\tag{7}
\]

where the partial derivatives are evaluated at the equilibrium. The eigenvalues of \( A \) determine the behavior of the equilibria. If they both have negative real parts, the equilibrium is stable and if any have positive real parts, the equilibrium is unstable. If both are real and of the same sign, the point is called a \textit{node}. Nodes consist of infinitely many trajectories emanating from
(unstable) or entering (stable) the equilibrium. If both eigenvalues are complex, the rest state is a vortex, and trajectories spiral into (stable) or out of (unstable) the rest point. If the eigenvalues have opposite signs, the rest state is a saddle point. Then a single pair of trajectories which define the stable manifold or set, enter the rest point and a single pair of trajectories, defining the unstable manifold, leave the equilibrium. When the determinant of $A$ (i.e. $ad - bc$) is negative the rest point is a saddle; if it is positive and the trace of $A$ ($a + d$) is nonzero, the equilibrium is a node or vortex. Cases for which the real part is zero do not persist for small changes in the parameters and often indicate the appearance of new equilibria or periodic solutions. A simple necessary and sufficient criterion for linear stability is that the trace, $a + d$ be negative and the determinant of $A$, $ad - bc$, be positive. A complete description of phase-plane methods can be found in (Edelstein-Keshet, 1988) as well as most texts on ordinary differential equations.

**Crossed excitatory and inhibitory networks**

The first result I want to establish in systems that have mutual coupling of the same sign is that periodic solutions are impossible. Once this is established, then a complete characterization can be made by simply studying the
intersections of the nullclines.

**Theorem 1** Suppose that \( w_{21}w_{12} \geq 0 \). Then there are no periodic solutions to

\[
\begin{align*}
\tau_1 x_1' &= -x_1 + f(w_{11}x_1 + w_{12}x_2 + s_1) \\
\tau_2 x_2' &= -x_2 + f(w_{21}x_1 + w_{22}x_2 + s_2)
\end{align*}
\]

(8) (9)

As the proof of this theorem has been given in the previous edition of the Handbook (page 733, Theorem 1), I will not prove it.

All solutions to (8,9) are bounded, and Theorem 1 implies that trajectories are monotone so this means that all solutions tend to equilibria. This in turn means that the time constants can be set to 1 without loss of generality as the dynamics is completely trivial. The intersections of the nullclines and some observations on the signs of the coefficients in the linearized matrix based on the nullclines allows one to completely determine the number and stability type of the equilibria.

Recall \( f \) is increasing and bounded. Without loss of generality, one can assume the minimum of \( f \) is 0 and the maximum is 1. \( f \) is invertible and the inverse is also monotone with asymptotes at 0 and 1. The formula for the
$x_1$-nullcline is

$$x_2 = (-w_{11}x_1 - s_1 + f^{-1}(x_1))/w_{12}$$

(10)

The $x_2$-nullcline satisfies:

$$x_1 = (-w_{22}x_2 - s_2 + f^{-1}(x_2))/w_{21}$$

(11)

If $h(x) = (-w_s x - s + f^{-1}(x))/w_c$, then $h$ is monotone if $w_s$ is either positive or small and negative. However if $w_s$ is large enough, $h$ develops a kink and is “cubic” shaped (see Figure 2a.) If $w_c$ is positive (mutual excitation) then $h \to -\infty$ as $x \to 0$ and $h \to \infty$ as $x \to 1$ (cf figure 2a.) When $w_c$ is negative (mutual inhibition), the asymptotes are switched (figure 2b.) Finally, the stimulus parameter $s$ merely translates the nullclines up and down in the case of the $x_1$-nullcline and left-right for the $x_2$-nullcline. The phase plane is easy to construct with these observations.

In both cases, there can be up to 9 different equilibria and that there is always at least one. Figure 3 shows some typical configurations for mutually inhibitory interactions. To assess the stability of the equilibria, one need only look at the positions of the nullclines at the equilibria. Refering to (7), I will use the nullclines to determine the signs and relative magnitudes of the entries in this matrix. For mutually inhibiting cells, the following are necessary and sufficient for stability:
i The slope of both nullclines is negative through an equilibrium.

ii The slope of the $x_1$ nullcline is steeper than the $x_2$ nullcline through the equilibrium.

If either of these is violated, the equilibrium is unstable.

For mutually excitatory cells, the conditions for stability are:

i The slope of both nullclines is positive through an equilibrium.

ii The slope of the $x_1$ nullcline is steeper than the $x_2$ nullcline through the equilibrium.

Tangential intersections are saddle-nodes and as some parameter varies will lead to either two new equilibria or the disappearance of the pair. The matrix $A$ has a zero eigenvalue when there are tangential crossings for then the slopes of the nullclines are the same. That is $-a/b = -c/d$ so $ad - bc = 0$. A bit of counting shows that when there are nine equilibria, four are stable nodes, four saddles, and one unstable node. As the parameters vary, a pair of equilibria is lost, a saddle-point and either a stable node or the unstable node, leaving seven equilibria. Further losses of equilibria (or gains, up to a maximum of nine) are obtained as the parameters vary ending in the minimum of a single globally stable equilibrium.
When there are several stable equilibria, it is important to determine what
initial conditions lead to which of the equilibria. The set of all initial data
that tend to a particular equilibrium point is called the domain of attraction
of the equilibrium point. For the present networks, this is very easy to
determine geometrically. Figure 3 depicts a network of mutually inhibitory
cells that has 5 equilibria labeled \((a - e)\). The above discussion allows one
to conclude that \((a, c, e)\) are stable nodes and \((b, d)\) are saddle-points. Each
saddle-point has associated with it one positive eigenvalue and one negative
eigenvalue. Corresponding to this negative eigenvalue is the stable manifold
for the saddle point. It consists of the set of all initial conditions that tend
to the equilibrium point as \(t \to \infty\). For two-dimensional neural nets, this is
a one-dimensional set and is then often called a separatrix. I have drawn it
for each of the two saddle points in Figure 3 as the dashed lines pointing into
\((b)\) and \((d)\). These curves divide the two-dimensional plane into 3 regions
which I have labeled \(A, C, E\). All initial data in \(A\) tends to equilibrium point
\(a\) and so on. Thus, although the saddle points are unstable, their stable
manifolds provide the boundaries which determine the final states of the
network given the initial state. From this description the reader should be
able to construct complete qualitative pictures for other configurations of the
nullclines for mutually excitatory or inhibitory nets.

Summarizing, I have used phase plane analysis to show that for a pair of coupled neurons with mutual excitation or inhibition, the only stable solutions are equilibria. The stable manifolds of the saddle-points divide the plane into domains of attraction for each of the stable equilibria. All equilibria are approached monotonically and there can be up to 4 stable steady states.

A pair of excitatory and inhibitory cells

In many regions of cortex and in fact throughout the central nervous system, many of the coupled excitatory and inhibitory cells comprise a local neural network. These networks have been the subject of numerous mathematical and computational investigations (Ermentrout and Cowan, 1979; Pinto et al, 1996; Ellias and Grossberg, 1976; Wilson and Cowan, 1972; Beer, 1995) One can view such systems as either two neurons acting in isolation (a difficult experiment to imagine) or more reasonably as a two-layer network with spatially homogeneous activity. Then each component is the activity of a pool of cells rather than the activity of a single cell.
I will consider a network of the form:

\[ x_1' = -x_1 + f(w_{11}x_1 - w_{12}x_2 + s_1) \]  \hspace{1cm} (12)

\[ x_2' = (-x_2 + f(w_{21}x_1 - w_{22}x_2 + s_2))/\tau \]  \hspace{1cm} (13)

where all of the weights are nonnegative. I have introduced a time constant for the inhibitory neurons because one cannot expect them to have the same temporal behavior as the excitatory cells. The Jacobian matrix \( A = [\partial x'_j/\partial x_i] \) at an equilibrium point, \((\bar{x}_1, \bar{x}_2)\) has coefficients:

\[ a = -1 + w_{11}f'(w_{11}\bar{x}_1 - w_{12}\bar{x}_2 + s_1) \]  \hspace{1cm} (14)

\[ b = -w_{12}f'(w_{11}\bar{x}_1 - w_{12}\bar{x}_2 + s_1) < 0 \]  \hspace{1cm} (15)

\[ c = (w_{21}f'(w_{21}\bar{x}_1 - w_{22}\bar{x}_2 + s_2))/\tau > 0 \]  \hspace{1cm} (16)

\[ d = (-1 - w_{22}f'(w_{21}\bar{x}_1 - w_{22}\bar{x}_2 + s_2))/\tau < 0. \]  \hspace{1cm} (17)

It is clear that all of the coefficients except for \( a \) have a fixed sign independent of the parameters. If \( w_{11}f' > 1 \) then \( a > 0 \) and the system is called an activator-inhibitor system since \( x_1 \) activates both itself and \( x_2 \) while \( x_2 \) inhibits everything to which it connects. Activator-inhibitor models occur ubiquitously in biology and their dynamics is rich and varied. (See, for example, the chapter by Murray). Oscillations, excitability, and multiple steady states are among the possible behaviors of these networks. Since a
very complete analysis of these systems as applied to neural excitation is
given in Rinzel and Ermentrout (1998), I only sketch some of the dynamic
behavior possible for this network.

The qualitative behavior of any planar model can be understood by combi-
nating nullcline analysis with local stability analysis of the equilibria which
depends on the coefficients of the Jacobian $A$. The neural model studied in
Rinzel and Ermentrout (1998) has exactly the same nullcline structure and
has a Jacobian matrix with the same structure as the neural net model.
Hence, I will only outline the dynamics of this system; details can be ex-
tracted from the aforementioned article.

It is instructive to first consider the effects of parameters on the shapes
of the nullclines. A typical nullcline configuration is shown in figure 4 for
$(12, 13)$. The $x_2$-nullcline is always monotonically increasing; $w_{21}$ sharpens
it while $w_{22}$ makes it shallower and $s_2$ shifts it left and right. As described
in §2, the effect of $w_{11}$ is to kink the $x_1$-nullcline while $w_{12}$ makes it less
kinked. $s_1$ shifts it up and down. Finally, the parameter $\tau$ has no effect
on the nullclines but dramatically alters the dynamics and stability of the
equilibria. Changing $\tau$ has no effects on the determinant of $A$ (so a saddle
point cannot become a node) but it can switch the sign of the trace of $A$ and
so change a point from a stable node to an unstable node.

The positions of the nullclines make it clear that there can be up to five equilibria and at least one equilibrium point. Furthermore any equilibria that occur on the “unkinked” part of the \(x_1\)-nullcline are necessarily asymptotically stable since then \(a < 0\) in Equation (14). Thus, the trace, \(a + d < 0\) and the determinant, \(ad - bc > 0\). If \(w_{11}\) is sufficiently small so that the \(x_1\)-nullcline is monotone, then there is only one equilibrium point and it is globally stable. This statement follows from the facts that all solutions are bounded and an application of Bendixson’s negative criterion (Edelstein-Keshet, 1988) which eliminates periodic orbits when \(a + d < 0\). Any time the inhibitory nullcline has a lesser slope than the excitatory nullcline, the equilibrium is a saddle point. These considerations along with the discussion above show how the parameters affect the local existence and stability of various rest states. The global dynamics is much more complicated since one cannot eliminate the possibility of limit cycle solutions.

**Excitability**

One important difference between networks consisting of one excitatory and one inhibitory layer and the networks described in §2 is the possibility of
excitable dynamics. As was shown in §2, trajectories of the activity of cells are necessarily monotone. Thus if, say, \( x_1 \) is increasing then it can never decrease again. However, in mixed networks, no such restriction occurs and it is possible for \( x_1 \) to initially increase before decreasing again. In particular, a network is said to be excitable if there is a unique globally stable rest state with the following property. Small perturbation from rest decay monotonically back to rest but perturbations larger than some threshold continue to grow before decaying back to the stable rest state (see Figure 5.) There are at least two qualitatively different types of excitability for networks with the present structure. In type I excitability, there are three equilibria while in type II, there is one. These two cases are described in Rinzel and Ermentrout (1998). In the context of neural networks, this type of behavior has been called an active transient. It can be viewed as a transient excitatory activity due to a stimulus that is eventually quelled by the inhibitory interneuronal feedback.

**Periodic solutions**

These occur generally (although not strictly) when there is a single rest state on the middle branch and it is unstable. This point must necessarily be a node and the boundedness of solutions thus implies that a limit cycle exists.
If some parameter (say $\tau$) is varied in such a way as to make the unique equilibrium go from a stable point to an unstable point (without introducing any new rest states) then a Andronov-Hopf bifurcation generically occurs and this implies that a periodic solution exists near the rest state. For planar systems, easily checked necessary conditions for a Andronov-Hopf bifurcation are that the determinant of $A$ remain positive and the trace change from negative to positive as the parameter is varied. If this new limit cycle is unstable, then there can be regimes in parameter space where there are two stable behaviors: (i) a stable rest state and (ii) a stable large amplitude periodic solution (see Figure 6.) This is known as bistability.

Other behavior

In addition to excitability, multiple equilibria, oscillations, and bistability, other types of dynamic behavior can be found in these simple models. Infinite period oscillations and homoclinic trajectories can be obtained in some parameter regimes. (A homoclinic trajectory is one that leaves a saddle point from one side and enters it from another and can occur as the period of a limit cycle tends to infinity.) Homoclinics are important since they separate qualitatively different types of behavior. Furthermore, when one periodically
stimulates a system with homoclinics, it is possible to obtain complex irregular behavior called \textit{chaos} which cannot occur in planar systems without forcing (see Guckenheimer and Holmes, 1983.)

There are many other pictures possible with this simple model and I urge the reader to explore the phase plane dynamics of this excitatory inhibitory net. Phase plane methods provide a powerful analytic and qualitative tool for studying small neural networks. When combined with sophisticated numerical tools, a complete understanding of the global dynamics is possible.

In systems with more than two-components, it is difficult to make any general comments on behavior. For symmetrically coupled networks with no self-connections, a complete analysis can be given (Hopfield, 1982). Weakly coupled systems of intrinsically oscillatory networks can be analyzed with the techniques described in the chapter by Kopell. Bifurcation methods and averaging techniques can often be used to reduce higher-dimensional systems to a simpler set of equations that is in a much lower dimension. (See Hoppensteadt & Izhikevich, 1997, as well as their chapter in this Handbook.)
Numerical methods

Computers are a valuable adjunct in the exploration of systems of differential equations. For this article, I have used a program called XPPAUT that is available for both Windows 95/NT/98 computers (Winpp) and Unix workstations. Both are available through http://www.pitt.edu/~phase in the To get global pictures of the dynamics as one or two parameters are varied, a powerful numerical package written by Doedel et al (1997) called AUTO can be used. A version is available at http://indy.cs.concordia.ca/auto/.

References


E. Doedel, A. Champneys, T. Fairgrieve, Y. Kuznetsov, B. Sandstede, and XJ Wang, (1997) AUTO97: Continuation and bifurcation software for ordinary differential equations (with HomCont), Computer science, Concordia University, Montreal, Canada.


Figure 1: Typical nonlinear input-output function of a single model neuron.
Figure 2: Nullcline shape for (a) mutual excitation and (b) mutual inhibition.
Figure 3: Phase plane for two mutually inhibitory neurons. Nullclines are solid lines and the stable manifolds of the saddle points $b$, $d$ are shown dashed. $a$, $c$, $e$ are stable nodes with domains of attraction $A$, $C$, $E$ respectively. Insets show other some other possible nullcline configurations.
Figure 4: Typical phase plane for an excitatory-inhibitory pair. Nullclines and a typical trajectory are shown. There is a unique globally attracting equilibrium point.
Figure 5: Excitable dynamics. $\bar{x}_1$ is the globally stable rest state and $x_1^*$ is the threshold. Trajectory $a$ is subthreshold and $b$ is superthreshold.
Figure 6: Phase plane for bistable regime of parameters. Nullclines are shown as well as the stable periodic orbit (dark line), the unstable periodic orbit (dashed line) and representative trajectories (thin lines.) Grey area denotes domain of attraction for the fixed point. The rest of the plane is attracted to the stable periodic orbit.