

EBME 309, Spring 2003 Two-dimensional dynamics

Niels F. Otani

Associate Professor, Department of Biomedical Engineering, Case Western Reserve University

Abstract. Linear and nonlinear 2-d differential equations are studied to illustrate the behavior two-dimensional nonlinear systems. The Fitzhugh-Nagumo equations as a model of an excitable cell are used as an example.

1. Linear 2-d dynamics

Now that you have some idea of what happens in one-dimensional dynamics, let's now look at two-dimensional dynamics. First let's start with a linear 2-d system. A 2-d system means that there are now two dependent variables (in this case x and y) and therefore two, usually first-order, ordinary differential equations:

$$\frac{dx}{dt} = Ax + By \quad (1)$$

$$\frac{dy}{dt} = Cx + Dy \quad (2)$$

Here A , B , C and D are constants. These equations are linear because they have two properties:

1. If the solutions obtained when the initial conditions are $x(0) = x_0$ and $y(0) = y_0$ are $x(t)$ and $y(t)$, then the solutions obtained when the initial conditions are $x(0) = cx_0$ and $y(0) = cy_0$ are $cx(t)$ and $cy(t)$, where c is any constant.
2. If the solutions obtained from initial conditions x_{10} and y_{10} , and x_{20} and y_{20} , are $x_1(t)$ and $y_1(t)$, and $x_2(t)$ and $y_2(t)$, respectively, then the solutions from initial conditions $x_{10} + x_{20}$ and $y_{10} + y_{20}$ are $x_1(t) + x_2(t)$ and $y_1(t) + y_2(t)$.

It is always possible to convert two linear, first-order equations into a single second-order equation as follows: first notice that Eq. (2) may be written as:

$$\left(\frac{d}{dt} - D\right)y = Cx. \quad (3)$$

Note: The expression $(d/dt - D)(\text{anything})$ is shorthand for $d(\text{anything})/dt - D(\text{anything})$. We therefore operate on Eq. (1) with $(d/dt - D)$:

$$\left(\frac{d}{dt} - D\right)\frac{dx}{dt} = A\left(\frac{d}{dt} - D\right)x + B\left(\frac{d}{dt} - D\right)y. \quad (4)$$

Substituting Eq. (3):

$$\frac{d^2x}{dt^2} - (A + D)\frac{dx}{dt} + (AD - BC)x = 0. \quad (5)$$

(Notice that the coefficient of dx/dt is minus the trace of the matrix

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \quad (6)$$

while the coefficient of x is its determinant. This is not an accident.) As you've seen from previous differential equation courses, this second-order equation has solutions of the form $x(t) = (\text{const})e^{\lambda t}$. You can determine the constant λ by substituting this solution into Eq. (5). Doing the differentiations and combining terms, we have,

$$(\lambda^2 - (A + D)\lambda + (AD - BC))(\text{const})e^{\lambda t} = 0 \quad (7)$$

which is only zero when

$$\lambda^2 - (A + D)\lambda + (AD - BC) = 0 \quad (8)$$

This equation, called the *characteristic equation*, has two roots; let's call them λ_1 and λ_2 . Then both $e^{\lambda_1 t}$ and $e^{\lambda_2 t}$ are solutions. Since the system is linear, the following,

$$x(t) = C_1 \exp(\lambda_1 t) + C_2 \exp(\lambda_2 t) \quad (9)$$

must also be a solution, and in fact turns out to be the most general form of the solution. The constants C_1 and C_2 are determined by the initial conditions. The other dependent variable, $y(t)$, has the same form, but with different constants C_1 and C_2 .

There is another way to solve Eqs. (1) and (2). We can write these equations in matrix form as,

$$\frac{d}{dt} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \quad (10)$$

In vector notation, this is:

$$\frac{d\mathbf{x}}{dt} = \mathbf{M} \cdot \mathbf{x}, \quad (11)$$

where,

$$\mathbf{x} \equiv \begin{bmatrix} x \\ y \end{bmatrix}, \quad \text{and} \quad \mathbf{M} \equiv \begin{bmatrix} A & B \\ C & D \end{bmatrix} \quad (12)$$

We now try to find solutions which are functions of t times one of the *eigenvectors* of the matrix \mathbf{M} . Recall that an eigenvector \mathbf{v}_1 of a matrix \mathbf{M} has the property,

$$\mathbf{M} \cdot \mathbf{v}_1 = \lambda_1 \mathbf{v}_1; \quad (13)$$

that is, when a matrix operates on one of its eigenvectors, the result is another vector pointing in the same direction as the eigenvector. The resultant vector is longer or shorter than the original eigenvector, scaled by a factor λ_1 which is called the *eigenvalue*. (If λ_1 is negative, then the resultant vector points in the direction opposite to the original eigenvector. It is also possible for λ_1 to be complex.) It turns out that this λ_1 is the same as the quantity λ_1 defined above as the solution to the characteristic equation (8). We will see this momentarily.

If the initial value of $\mathbf{x}(t)$ is a constant times one of these eigenvectors \mathbf{v}_1 , then initially we have,

$$\frac{d\mathbf{x}}{dt} = \mathbf{M} \cdot ((\text{const})\mathbf{v}_1) = (\text{const})\lambda_1 \mathbf{v}_1 \quad (14)$$

which means that $d\mathbf{x}/dt$ is also pointing in the direction of eigenvector \mathbf{v}_1 . It should be clear that when both a vector \mathbf{x} and its time rate of change ($d\mathbf{x}/dt$) point in the same direction, then that the vector will continue to point in that direction for all time, and the trajectory, marked by the tip of the vector, will be a *straight line*. Therefore, if initially $\mathbf{x}(t)$ points in an eigenvector direction: ($\mathbf{x}(0) = (\text{const})\mathbf{v}_1$), then we can be confident that the solution must be of the form, $\mathbf{x}(t) = f(t)\mathbf{v}_1$ for some scalar function $f(t)$. Substituting this form of the solution into Eq. (11), we have,

$$\frac{df(t)}{dt} \mathbf{v}_1 = \mathbf{M} \cdot (f(t)\mathbf{v}_1) \quad (15)$$

or, using Eq.(13),

$$\frac{df(t)}{dt} \mathbf{v}_1 = f(t)\lambda_1 \mathbf{v}_1. \quad (16)$$

We can drop the \mathbf{v}_1 from both sides of the equation to obtain

$$\frac{df(t)}{dt} = \lambda_1 f(t). \quad (17)$$

Notice what eigenvectors has allowed us to do. We have reduced our two-dimensional equations down to a single one-dimensional equation to which we know the solution: $f(t) = f_1 e^{\lambda_1 t}$! Here f_1 is a constant which depends on the initial conditions. All this, of course, also applies to any other eigenvectors. In a two-dimensional system, there are generally two eigenvectors, which we'll call \mathbf{v}_1 and \mathbf{v}_2 .

Now here's the point: as long as there are two eigenvectors pointing in different directions (which is almost always true), you can write *any* initial condition vector \mathbf{x}_0 as a *linear combination* of these eigenvectors, like this:

$$\mathbf{x}_0 = f_1 \mathbf{v}_1 + f_2 \mathbf{v}_2 \quad (18)$$

Now remember what we said about linear equations. The solution resulting from a sum of initial conditions is just the sum of the individual solutions resulting from these initial conditions. (You might have to think about that one a little. . .) The solution for initial vector $f_1 \mathbf{v}_1$ is $f_1 e^{\lambda_1 t} \mathbf{v}_1$, as we have just seen. Similarly, if the initial vector is $f_2 \mathbf{v}_2$ then the solution is $f_2 e^{\lambda_2 t} \mathbf{v}_2$. Therefore when the initial conditions are as expressed in Eq. (18), the solution must be,

$$\mathbf{x}(t) = f_1 e^{\lambda_1 t} \mathbf{v}_1 + f_2 e^{\lambda_2 t} \mathbf{v}_2 \quad (19)$$

Not only is this eigenvector-eigenvalue analysis a method of solution, it is also telling you something important. It says that two-dimensional linear equations really only have two types of behavior, one associated with each of the two eigenvectors. Any behavior you see coming from a 2-d linear system is actually just the mathematical sum of these two behaviors. Furthermore, these two behaviors are independent of each other. One can be growing exponentially, while the other is decaying, for example. These two behaviors *characterize* the system; in fact, "eigen" means "characteristic" in German, if my information is right. These behaviors, by the way, are called *eigenmodes* of the system, quite appropriately.

When both of the eigenvalues have negative real parts, then both of the exponentials appearing in Eq. (19) are decreasing in magnitude, and thus $\mathbf{x}(t)$ must be "falling" into the origin. Such behavior is called *stable* behavior, and the system itself is referred to as *stable*. If *either* of the eigenvalues has a positive real part, then the corresponding exponential is growing with time. In this case, unless the corresponding coefficient (f_1 or f_2) is identically 0, $\mathbf{x}(t)$ will grow in magnitude, and move away from the origin. This behavior is termed *unstable*.

Note, by the way, that is possible for both the eigenvalues and eigenvectors to be complex. When this happens, we will not be able to find eigenvector directions

in the real phase plane, and therefore we will not be able to find any straight-line trajectories. In this case, the orbits tend to spiral around the origin. The notion of stability still applies to this case. In stable systems, the trajectories spiral in; in unstable systems, the trajectories spiral out.

So how do we solve for these eigenvalues and eigenvectors? Writing out Eq. (13), we have,

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} v_x \\ v_y \end{bmatrix} = \lambda \begin{bmatrix} v_x \\ v_y \end{bmatrix} \quad (20)$$

where the column vector appearing on both sides stands for either of the eigenvectors, and λ stands for the corresponding eigenvalue. Writing this equation out as two scalar equations, and then writing it back in matrix form again, we find that Eq. (20) may be written as,

$$\begin{bmatrix} A - \lambda & B \\ C & D - \lambda \end{bmatrix} \begin{bmatrix} v_x \\ v_y \end{bmatrix} = 0 \quad (21)$$

This equation only has a non-zero solution for the column vector \mathbf{v} when the determinant of the matrix is zero; that is, when,

$$(A - \lambda)(D - \lambda) - BC = 0 \quad (22)$$

Notice that this is just the characteristic equation (Eq. (8)) again. The solutions are just going to be λ_1 and λ_2 , which demonstrates that the two definitions of λ we have been using are the same.

To obtain the eigenvector corresponding to one of these eigenvalues, simply substitute that eigenvalue into Eq. (21) and solve for the components of \mathbf{v} . In solving for \mathbf{v} , you will find that one of the equations is redundant, a consequence of the fact that the determinant of the matrix is zero. You handle this by choosing one of the components of \mathbf{v} arbitrarily, say equal to 1, and then solving for the remaining component(s). This arbitrariness corresponds to the fact that the length of the eigenvector is arbitrary—only the direction the eigenvector points is important. In other words, if \mathbf{v} is an eigenvector of \mathbf{M} with eigenvalue λ , then so is $a\mathbf{v}$ for any constant a . This follows from the fact that \mathbf{M} is linear.

2. Nonlinear 2-d dynamics

Now that we know something about how two-dimensional linear systems behave, let's look at nonlinear 2-d systems. We'll use as our example the *Fitzhugh-Nagumo equations*, which describe the behavior of an excitable

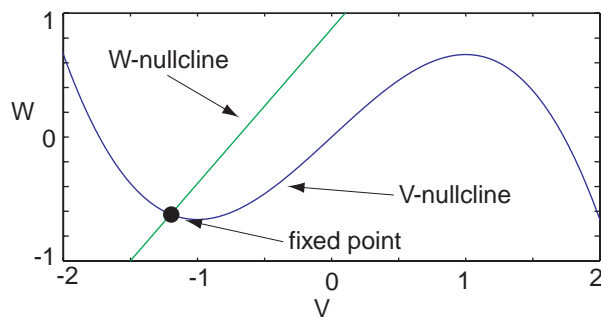


Figure 1. The fixed point and the V - and W -nullclines for the Fitzhugh-Nagumo equations.

cell, like a neuron or cardiac cell. The equations are:

$$\frac{\partial V}{\partial t} = \frac{1}{\epsilon} \left(V - \frac{V^3}{3} - W \right) \equiv f(V, W) \quad (23)$$

$$\frac{\partial W}{\partial t} = \epsilon(V - \gamma W + \beta) \equiv g(V, W) \quad (24)$$

where V is the membrane voltage and W is a quantity which is associated with the recovery of the cell after it fires. Typical values for the parameters are $\epsilon = 0.2$, $\gamma = 0.8$, and $\beta = 0.7$. You can think of Eq. (23) as the current equation, with the capacitive charging term (proportional to dV/dt) on the left and the nonlinear channel currents on the right. From the form of Eq. (24) you can see that the recovery of the cell is being represented as a gate opening equation with voltage-dependent rates.

Of course, one way to figure out how a system of differential equations behaves is to model them on the computer, as we are doing in EBME 359. There are however, reasons to supplement the simulations with pencil-and-paper analysis. The most important reason to look at the equations analytically is to acquire some understanding for how the equations produce the behavior that you observe in the simulations. Another reason, of course, is to check your simulations.

A natural way to analyze two-equation systems is to plot their trajectories on a two-dimensional plane whose axes are the two dependent variables. Such a study is called *phase plane analysis*. In the case of the Fitzhugh-Nagumo equations (23) and (24), the two dependent variables are V and W , so that we are plotting trajectories in V - W space.

We start the analysis by plotting so-called nullclines. The V -nullcline is simply defined to be that set of points for which $dV/dt = 0$. Defining $f(V, W)$ to be the right-hand side of Eq. (23), the V -nullcline is just that set of

points (V, W) which satisfies the condition $f(V, W) = 0$. From Eq. (23), it should be clear that we plot the V -nullcline simply by plotting the function $W = V - V^3/3$.

Similarly, the W -nullcline is defined as the locus of points for which $g(V, W)$, the right-hand side of Eq. (24), is zero. From the form of the right-hand side of Eq. (24), the W -nullcline is given by the function, $W = (V + \beta)/\gamma$. The V and W -nullclines for the Fitzhugh-Nagumo equations are shown in Fig. 1.

At every point where the two nullclines intersect (see Fig. 1), both dV/dt and dW/dt are zero. This means that neither V nor W can change if the system is started at one of these points, so the system will stay at these points forever. These points are therefore called *fixed points*. For the Fitzhugh-Nagumo equations, there is only one fixed point located at approximately $(V, W) = (-1.2, -0.62)$.

The next step is to consider the *stability* of these fixed points. The question to be answered here is, if the system is started not at the fixed point, which we will refer to as (V_0, W_0) , but is instead started very close to the fixed point, will the system move toward the fixed point, or will it try to escape? The set of points which tend to approach the fixed point is called the *attractive basin* of the fixed point. We can then restate the question like this: does the attractive basin completely surround the fixed point? If so, the fixed point is referred to as *stable*; otherwise it is *unstable*. The attractive basin must completely surround the fixed point; otherwise we would be able to find a point arbitrarily close to the fixed point which is not ‘‘attracted’’ to the fixed point.

There is a simple way to determine the stability of a fixed point which works for any number of system equations. We start by linearizing the equations around the fixed point. For our present two-equation example, (Eqs. (23) and (24)), linearizing around (V_0, W_0) yields,

$$\begin{aligned} \frac{dV}{dt} &= f(V_0, W_0) + \frac{\partial f}{\partial V}(V_0, W_0) (V - V_0) \\ &\quad + \frac{\partial f}{\partial W}(V_0, W_0) (W - W_0) \end{aligned} \quad (25)$$

$$\begin{aligned} \frac{dW}{dt} &= g(V_0, W_0) + \frac{\partial g}{\partial V}(V_0, W_0) (V - V_0) \\ &\quad + \frac{\partial g}{\partial W}(V_0, W_0) (W - W_0) \end{aligned} \quad (26)$$

In this expression, $f(V_0, W_0)$ and $g(V_0, W_0)$ are of course, zero, since that’s how V_0 and W_0 were defined. Let’s define $\delta V \equiv V - V_0$, and $\delta W \equiv W - W_0$ to be, respectively, the displacements of V and W away from the

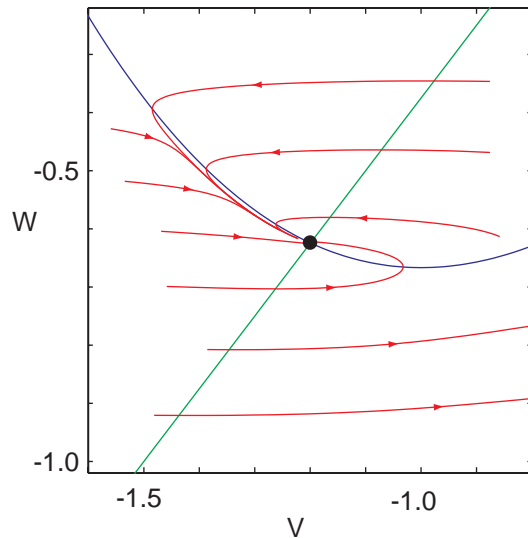


Figure 2. A few selected trajectories close to and around the fixed point of the Fitzhugh-Nagumo equations. The arrows show the direction of flow.

fixed point. The vector $(\delta V, \delta W)$ is called the *perturbation* of V and W around the fixed point (V_0, W_0) . Now, $dV/dt = d(\delta V)/dt$ since V_0 is a constant, and similarly for dW/dt , so Eqs. (25) and (26) become,

$$\frac{d\delta V}{dt} = \frac{\partial f}{\partial V}(V_0, W_0) \delta V + \frac{\partial f}{\partial W}(V_0, W_0) \delta W \quad (27)$$

$$\frac{d\delta W}{dt} = \frac{\partial g}{\partial V}(V_0, W_0) \delta V + \frac{\partial g}{\partial W}(V_0, W_0) \delta W \quad (28)$$

These two equations thus describe the evolution of perturbation of (V, W) around the fixed point. We can also write these equations in matrix form:

$$\frac{d\delta \mathbf{V}}{dt} = \mathbf{M} \cdot \delta \mathbf{V} \quad (29)$$

where

$$\mathbf{M} = \begin{bmatrix} \frac{\partial f}{\partial V}(V_0, W_0) & \frac{\partial f}{\partial W}(V_0, W_0) \\ \frac{\partial g}{\partial V}(V_0, W_0) & \frac{\partial g}{\partial W}(V_0, W_0) \end{bmatrix} \quad (30)$$

and

$$\delta \mathbf{V} = \begin{bmatrix} \delta V \\ \delta W \end{bmatrix} \quad (31)$$

You may recognize the matrix in Eq. (30) as being the Jacobian matrix.

Keep in mind that all these linearized equations are approximations: in the Taylor series for $f(V, W)$ and

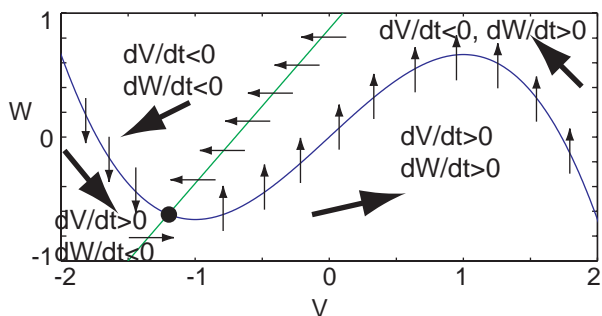


Figure 3. Illustration of the flow direction across nullclines, and rough flow direction in the region between nullclines. The sign of the time-derivatives of V and W in each region are also indicated.

$g(V, W)$, we threw away higher order terms in $\delta\mathbf{V}$ in order to obtain Eqs. (25) and (26). The approximation is good as long as the magnitude of $\delta\mathbf{V}$ is small; that is, as long as we stay close to the fixed point.

Notice that Eq. (29) is identical in form to Eq. (11) which describes a 2-d linear system. What this means is that, near a fixed point, a nonlinear system behaves, to a good approximation, like a 2-d linear system! The displacement $\delta\mathbf{V}$ and the fixed point play, respectively, the roles of \mathbf{x} and the origin in the linear system. And of course, just as in the linear system, the eigenvalues and eigenvectors of the Jacobian matrix \mathbf{M} completely characterize the behavior of the trajectories. Again, all this applies only to the immediate vicinity of the fixed point. Once you are farther away, the system is nonlinear, and all bets are off.

As an example, consider the fixed point of the Fitzhugh-Nagumo equations shown as the intersection of the V - and W -nullclines in Fig. 1. When the partial derivatives in Eq. (30) are evaluated at the fixed point (at approximately $(V_0, W_0) = (-1.2, -0.62)$) and the characteristic equation solved, we find the real parts of both λ_1 and λ_2 are negative. The trajectories therefore should be attracted to this fixed point. This can be verified using a computer to solve the differential equations. Some of these trajectories are shown in Fig. 2. Notice that all nearby trajectories end up falling into the fixed point. The two trajectories at the bottom of the Figure are not close enough, however, and continue off to the right.

Returning to our discussion of nullclines, we find that nullclines can also provide some insight into what trajectories look like away from the fixed points. In our Fitzhugh-Nagumo example (Eqs. (23) and (24)), each nullcline divides the phase-plane into two regions, one

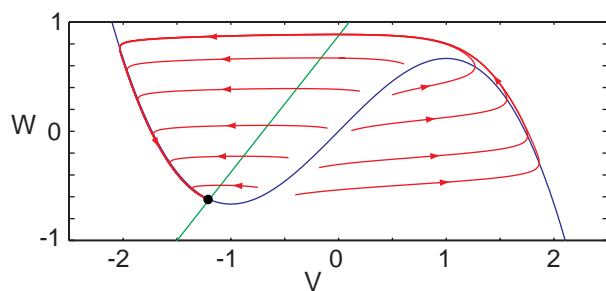


Figure 4. Typical trajectories and their relation to the nullclines in the Fitzhugh-Nagumo equations.

above the nullcline and one below it as shown in Fig. 3. The V -nullcline, for example, divides the phase-plane into the following two regions: one for which $f(V, W)$, and therefore dV/dt , is greater than zero, and for which $dV/dt < 0$. You can tell which is which by noting what happens to $f(V, W)$ when the (V, W) is pushed vertically upward from the V -nullcline. W is thus increased while V is held fixed. This point (V, W) automatically lies in the region above the V -nullcline, and since $f(V, W) = 0$ on the nullcline, if increasing W decreases $f(V, W)$, then we must have $dV/dt < 0$ above the V -nullcline. Similarly, we also have $dV/dt > 0$ below the V -nullcline. In Eq. (23), if ϵ is positive, the function $f(V, W) = (1/\epsilon)(V - V^3/3 - W)$ does indeed decrease as W increases. We can also use the W -nullcline in the same way: in our case $g(V, W) = \epsilon(V - \gamma W + \beta)$ decreases as W increases (for positive ϵ and γ), so $dW/dt < 0$ above the W -nullcline, and $dW/dt > 0$ below it. The V and W nullclines together divide the phase-plane into four regions, each with a different set of signs for dV/dt and dW/dt . For example, the lower-right region, being below both nullclines, has both $dV/dt > 0$ and $dW/dt > 0$. From this, we know that all trajectories in this region have to be moving up and to the right, as indicated by the big arrow in Fig. 3. You can also often get some additional help by studying your functions $f(V, W)$ and $g(V, W)$ carefully. For our example, we note that the size of $f(V, W)$ and $g(V, W)$ differ quite substantially from one another. Since for the default parameters for our example, $\epsilon = 0.2$, $f(V, W)$, being proportional to $1/\epsilon$, is typically quite large for any point (V, W) some respectable distance away from the V -nullcline, while $g(V, W)$, and therefore dW/dt , being proportional to ϵ , are typically quite small. This means that we expect to see all of our trajectories nearly horizontal anywhere away from the V -nullcline, as shown in Fig. 4.

The other bit of information the nullclines can give us has to do with what the trajectories are doing as they cross over the nullclines. By using the fact that the $dV/dt = 0$ is zero on the V -nullcline, it is clear that whatever trajectories cross the V -nullcline must cross it exactly vertically, as shown in Fig. 3. You can tell whether trajectories cross from the bottom up or top-down from the sign of dW/dt at the point of crossing, which in turn depends on what side of the W -nullcline your crossing occurs. Similarly, the trajectories must cross the W -nullcline horizontally.

All this information goes a long way towards constructing solutions to your system of differential equations, all without the use of a computer. For the Fitzhugh-Nagumo equations, it is possible to reproduce all the important features of both V and W as functions of time. While you won't always be this successful, you can see the benefits of using nullclines for both checking and, more importantly, *explaining* your computer solution.

One of the features which is well-explained by this nullcline method is the existence of a threshold firing voltage in the Fitzhugh-Nagumo equations. It should be clear that the increasing portion of the V -nullcline plays an important role. As is evident in Fig. 4, if the system is started with initial conditions V and W such that the point (V, W) lies just to the right of this portion of the V -nullcline, we can see from the above arguments that the trajectory will track quite rapidly to the right until the decreasing portion of the V -nullcline is crossed. The trajectories in this new region must move up and to the left, so the system point is forced to hug the V -nullcline and move slowly upward and leftward. At the maximum of the V -nullcline, the system point “falls off” the nullcline and moves rapidly leftward. It then crosses the other decreasing segment of the V nullcline and moves slowly back towards the fixed point. If, on the other hand, the system is started with (V, W) just to the left of the increasing portion of the V -nullcline, the behavior is quite different—the system point moves immediately to the left, and eventually returns to the fixed point. No firing occurs. We see that the increasing segment of the V -nullcline is the voltage threshold. We also note that V increases along this segment as W increases; that is, the threshold voltage increases as W increases. Here is another feature we are able to discover through the use of nullclines. Since W increases immediately after the system fires, this increase in threshold with increasing W means that the system has a higher firing threshold immediately after firing. This is a feature typical of both cardiac and nerve cells and reflects

the so-called *refractory* nature of cells which have just fired.

So far we have seen a number of features often present in dynamical systems. These include fixed points, both stable and unstable, and excitable (i.e., firing) behavior. Another common feature is the *limit cycle*, which is a trajectory which closes on itself. In essence, a limit cycle is a ring-shaped trajectory. A system which exhibits limit cycle behavior is called a *cyclical* system. Heart pacemaker cells, such as those located in the sinus node, are examples of cyclical systems.

Like fixed points, limit cycles can be either stable or unstable. If points nearby the limit cycle tend to approach the limit cycle, circling around and around in the phase plane, coming closer and closer to the limit cycle, then the limit cycle is stable. If nearby points tend to spiral away, the limit cycle is unstable.

The ability of fixed points and limit cycles to attract neighboring trajectories to themselves has led them to be called *attractors*. You might ask whether there are other types of attractors other than fixed points and limit cycles. In fact, there are. They are called *strange attractors*, a term which is well deserved. They are typically quite bizarre objects and are often chaotic.