Part IV

All Systems

Chapter 1

Fast and Slow Times

Different physical processes operate at different time scales. Some go slow, some go fast. Usually they are not time coordinated. But when processes of different time scales start to interact in a larger scheme of things some may evolve to go in synchrony, some may retain their old pace, and some others may even manage to make their time differences more divergent. As a result, multitime processes in nature are everywhere.

Take the first mathematical model in history for a case in point. Here, for example, let M be the mass of the Sun and m be the mass of a planet. Let $\mathbf{r}_s$ and $\mathbf{r}_p$ be their displacement vectors from, say the center of the Milky Way Galaxy. Let $\mathbf{F} = \mathbf{F}(\mathbf{r}_p - \mathbf{r}_s)$ be the gravitational pull of the Sun on the planet. Then the gravity pull of the planet on the Sun is exactly the opposite, $-\mathbf{F}(\mathbf{r}_p - \mathbf{r}_s)$, and the Newtonian model for the two-body motion is

$$M\ddot{\mathbf{r}}_s = -\mathbf{F} \quad \text{and} \quad m\ddot{\mathbf{r}}_p = \mathbf{F},$$

with the double dots denoting for the accelerations. One simple rescale of the time $\tau = t/\sqrt{m}$ translates the model into

$$\mathbf{r}_s'' = -\epsilon\mathbf{F} \quad \text{and} \quad \mathbf{r}_p'' = \mathbf{F}, \quad \text{with} \quad \epsilon = m/M.$$

Here the mass ratio ϵ becomes the multitime proxy for the system. For the Earth, the mass ratio relative to the Sun is $\epsilon \sim 3 \times 10^{-6}$. For the Halley's Comet, it is a 10-order of magnitude smaller, $\epsilon \sim \times 10^{-16}$. Without any serious computation the two-time scale model immediately demolishes the geocentric paradigm of the heaven. In fact, by setting the mass ratio to zero, $\epsilon = 0$, we get the zero acceleration for the Sun. As a result the Sun's motion is stationary, moving in a constant velocity with respect to the galactic center relative to

the planet. That is, relative to the planet, the Sun is motionless. It is the planet that is doing these frantic moving around in the solar system. To ancient astronomers it was like inside a clear plastic ball tumbling down a hill to whom the Sun appeared to be rolling.

The point is clear. Typical two-body planetary models are of two time scales. The motion for the Sun is slow, that for the planet is fast. The addition of any celestial body of diverging mass ratio to the model creates another time scale. The multitime nature of the solar system is a given simply because its celestial bodies acquired their odd masses in ways not prohibited by the physical laws at the chaotic beginning of the system's creation.

Down to the microscopic realm, the firing dynamics of a neuron is of many time scales. Here, the electromagnetic force is instantaneous: wherever the force field exists charged ions must move. In comparison, the cross-membrane motion of the ions driven by the active ion pumps is slow. It takes the conversion of biochemical energy. This divergence in the time scales is clear by taking a look at such a model:

$$\begin{cases} CV_{\mathrm{C}}{}' = -[I_{\mathrm{Na}} + f_{\mathrm{K}}(V_{\mathrm{C}} - \bar{E}_{\mathrm{K}}) + A_{\mathrm{Na}} - A_{\mathrm{K}} - I_{\mathrm{ext}}] \\ A_{\mathrm{Na}}{}' = \lambda A_{\mathrm{Na}}[V_{\mathrm{C}} - \gamma(A_{\mathrm{Na}} - A_{\mathrm{K}})] \\ A_{\mathrm{K}}{}' = \lambda A_{\mathrm{K}}[-V_{\mathrm{C}} + \gamma(A_{\mathrm{Na}} - A_{\mathrm{K}})] \\ \epsilon I_{\mathrm{Na}}{}' = V_{\mathrm{C}} - \bar{E}_{\mathrm{Na}} - h_{\mathrm{Na}}(I_{\mathrm{Na}}). \end{cases} \tag{1.1}$$

Here, $V_{\mathrm{C}}, A_{\mathrm{Na}}, A_{\mathrm{K}}, I_{\mathrm{Na}}$ are, respectively, the transmembrane voltage, the sodium ion pump current, the potassium ion pump current, and the sodium ion current through its serial electrical and diffusive channels. I_{ext} is any forcing current external to the circuit if any. The potassium ion current is given by the IV-characteristic $I_{\mathrm{K}} = f_{\mathrm{K}}(V_{\mathrm{C}} - \bar{E}_{\mathrm{K}})$. The sodium current is governed by an S-shaped hysteresis: $V_{\mathrm{C}} - \bar{E}_{\mathrm{Na}} - h_{\mathrm{Na}}(I_{\mathrm{Na}}) = 0$. It is an ideal representation of the instantaneous electromagnetic force. In circuit theory it is approximated by the introduction of an extremely small parameter $0 < \epsilon \ll 1$ for the I_{Na}-equation as shown. It is the fastest equation. In comparison, the voltage equation V_{C} is the next in time scale as the capacitance C is usually small. That leaves the pump equations to have the slowest time scale as the pump parameter λ is not in the same orders of magnitude as either $1/C$ or $1/\epsilon$.

Up to the organism level, time scales diverge crossing trophics. Simple organisms tend to regenerate fast. Complex organisms further up the food chain tend to reproduce slow. The higher up the slower reproduction becomes. Conversion of mass and energy between trophics take time. Such properties can be

easily captured by models. For example, in this three-trophic food chain model

$$\begin{cases} \dot{X} = bX - mX^2 - \dfrac{a_1 X}{1 + a_1 h_1 X} Y \\ \dot{Y} = b_1 \dfrac{a_1 X}{1 + a_1 h_1 X} Y - d_1 Y - m_1 Y^2 - \dfrac{u_2 Y}{1 + u_2 v_2 Y} Z \\ \dot{Z} = r_2 \dfrac{u_2 Y}{1 + u_2 v_2 Y} Z - d_2 Z - m_2 Z^2 \end{cases} \tag{1.2}$$

with $X,\ Y,\ Z$ being the populations for a prey, a predator, and a top-predator, the change of parameters below

$$\begin{aligned} &\tau = \frac{b_1}{h_1} t,\ x = \frac{X}{K},\ y = \frac{Y}{Y_0},\ z = \frac{Z}{Z_0}, \\ &\text{with } K = \frac{b}{m},\ Y_0 = bh_1 K,\ Z_0 = b_1 b v_2 K \\ &\beta_1 = \frac{1}{a_1 h_1 K},\ \beta_2 = \frac{1}{u_2 v_2 Y_0},\ \delta_1 = \frac{d_1 h_1}{b_1},\ \delta_2 = \frac{d_2 v_2}{r_2}, \\ &\mu_1 = \frac{m_1 h_1 Y_0}{b_1},\ \mu_2 = \frac{m_2 v_2 Z_0}{r_2},\ \varepsilon = \frac{b_1}{b h_1},\ \zeta = \frac{r_2 h_1}{v_2 b_1}. \end{aligned} \tag{1.3}$$

transforms the dimensional model into this dimensionless one

$$\begin{cases} \varepsilon \dot{x} = x \left(1 - x - \dfrac{y}{\beta_1 + x} \right) \\ \dot{y} = y \left(\dfrac{x}{\beta_1 + x} - \delta_1 - \mu_1 y - \dfrac{z}{\beta_2 + y} \right) \\ \dot{z} = \zeta z \left(\dfrac{y}{\beta_2 + y} - \delta_2 - \mu_2 z \right) \end{cases} \tag{1.4}$$

The parameters ε and ζ are the birth ratios of adjacent trophic species. The prey dynamics is fast if ε is small. The top-predator dynamics is slow if ζ is small, and so on. If the top-predator is a pathogen or a species having a lot alternative food sources, then ζ can be large, making it a fast variable relative to the predator species. In other words, various time scales can be captured by varying magnitudes of the birth rate ratios ε, ζ. Singularly perturbed models like this one arise naturally from ecological modeling.

Not only multitime is a ubiquitous property of models but also a welcome relief for analysis. Its presence in a model almost always makes the model easier to understand than without. In fact, as a common simplification practice, one often judiciously sets some of the small parameters to their singular value zero to obtain a model smaller in dimension. In this chapter, we will take the approach to one extreme to show how multitime scales can be utilized to study differential equation models.

1.1 Anatomy of Fast-Slow System

After a mathematical model is obtained, it is usually not feasible, if not altogether impossible, to its structures in their essential elements, free of complications from other parts whose effects are not to enlighten but to obscure. Further mathematical idealizations are desirable and often follow. Such exercises usually result in forms considered canonical to the underpinning physical processes. The ideal forms may be minimal in ingredient or minimal in structure. The field of chaos offers a perfect example for this line of thinking.

A system of differential equations may have a structure as simple as having a stable equilibrium, or as mild as having a steady state oscillation, or as wild as having an unpredictable trajectory trapped inside a chaotic attractor. Simple and mild structures are also outnumbered by their chaotic kind. In the zoo of dynamical systems, the tamed species may manage to fill up an exhibition hall, but the wild kind can take up the whole park.

1.1.1 Complicated Look, Simplicity At Heart

Otto E. Rössler is a prolific German biochemist, a theorist. One year after he was introduced to the field of chaos he found in 1976 a simple way to construct differential equations of desired features. According to him, "(each of his artificially constructed systems) consists of (1) an ordinary 2-variable chemical oscillator and (2) an ordinary single-variable chemical hysteresis system." He believes that "according to the same dual principle, many more analogous systems can be devised, no matter whether chemical, biochemical, biophysical, ecological, sociological, economic, or electronic in nature." For implementation, his approach is both systematic and ad-hoc at the same time: "the described system is just one out of a huge variety of possible combinations of an oscillator, on the one hand, and a switching system, on the other." A chaotic attractor now bears his name.

Here is a deceptively 'complicated' system we cooked up using Rössler's recipe.

$$\begin{aligned} \dot{x} &= (y+a)[b] + (a-y)[-c] \\ \epsilon\dot{y} &= (a^2-y^2)[y+a-m(x+a)] - dy, \end{aligned} \tag{1.5}$$

with a, b, c, d, ϵ being positive parameters. When Rössler found his attractor, he opted for a final simple form rather than a simple geometric structure as if he still wanted to solve his equations exactly. We will let go completely the futility to find closed solutions of arbitrary differential equations as well the illusion

that simple equations ought be easier to solve or easier to understand. We will instead go after straight for the simplest geometric structures of equations which may look at a glance complex.

The one enduring element of Rössler's Dual Principle is the use a small parameter ϵ as the coefficient of the derivative of one of the variables:

$$\dot{x} = f(x, y), \ \epsilon\dot{y} = g(x, y) \tag{1.6}$$

where the time variable is t and $\dot{} := d/dt$ denotes the time derivative. System like this is referred to as a *singularly perturbed system* and the parameter ϵ is called the *singular parameter* if ϵ is very small. The immediate important fact about such a system is the fact that if the right hands of the equations are comparable in magnitude, $f \sim g$, then the change in x is slow and the change in y is fast because the rate of change for y is very large as $\dot{y} = g/\epsilon \sim \pm\infty$. All the properties of the system are dependent on the interplays between the variables' different rate responses to the time. Because of this two-time scale characteristic variable x is called *the slow variable* and y is called *the fast variable*.

1.1.2 Slow Time For Slow System, Fast Time For Fast System

There are two different ways to write the system. The form above is referred to as *the slow form* and the time t is *the slow time*. This characterization is viewed from the vantage of the slow variable x and the slow time scale t. By changing the slow time variable t to the fast time variable $\tau = t/\epsilon$, with $' = d/d\tau = \epsilon d/dt$, the slow form is changed to the *fast form* of the system:

$$x' = \epsilon f(x, y), \ y' = g(x, y). \tag{1.7}$$

The characterization is viewed from the vantage of the fast y and the fast time variable τ. In either form, the x-equation is called *the slow equation* or *the slow system* if x is a vector variable. Similarly , the y-equation is *the fast equation* or *the fast system*.

For higher dimensional systems, multiple time scales may involved. For example, a system of this form $\dot{x} = f, \epsilon_1\dot{y} = g, \epsilon_2\dot{z} = h$ with $0 \ll \epsilon_2 \ll \epsilon_1$ has x as the slow variable, y the fast variable, and z the fastest variable, and so on.

The dynamics of such multitime systems progress in distinct phases in general with different subsets of variables dominating the dynamics in different stages. They take in turns to showcase their unique geometric properties. When the right hand functions are meticulously crafted the dynamics often invoke the imagery of a symphony with fast, slow, and in-between movements, in succession by players playing their exquisite roles, in turns or together in harmony.

1.1.3 Slow Motion On Slow Manifold

Here is how the prototypical system (1.5) does it. Consider first the slow form of the system. Take the limit to $\epsilon = 0$, the equations become

$$\begin{cases} \dot{x} = f(x,y) = (y+a)[b] + (a-y)[-c] \\ 0 = g(x,y) = (a^2 - y^2)[y + a - m(x+a)] - dy. \end{cases}$$

The second equation is reduced to an algebraic equation rather than remains as a differential equation. In concept and in practice in general it is easier to solve algebraic equations than differential equations. More importantly, the full 2-dimensional system is reduced to the 1-dimensional fast equation on the zeroes of g, referred to as *the slow manifold*, which, depending on the number of variables, can be a set of curves, or a set of surfaces, or hyper-surfaces.

By setting $d = 0$ we obtain explicitly three branches of the manifold: $y = \pm a$ and $y = -a + m(x + a)$, all are lines, purposefully selected and constructed to create a specific configuration of the manifold to be explained below. In particular, the third solution is made to cut through the first two branches at these intercepts

$$(x,y) = (-a,-a) \text{ and } (x,y) = ((2-m)a/m, 1)$$

which can be easily controlled by the parameters a and m.

The inclusion of the parameter $d > 0$ is to create what Rössler called a switch. It is one of many ways to construct such switches, a convenient way nonetheless. What it does is to generate a z-shaped smooth curve for the slow manifold, which we will call it a *z-switch*, or a switch for short following Rössler. Instead of forming a sharp corner point, each of the knee point of the Z is smooth. They are called *the turning points* of the slow manifold. The switch depends on d continuously, with the curve closely tracking the respective line branches of $g = 0$ at $d = 0$. The following three paragraphes give a more detailed description, which can be skipped without any loss of following up the rest discussion.

For $d \neq 0$, the solution set of $g(x,y) = 0$ breaks up in two structurally different configurations as shown in Fig.1.1. The inclusion of the term $-dy$ is also made by design. For $d > 0$, all is required is to have the value of g at the intercept (a,a) positive $g(a,a) = -da < 0$ and the other intercept value negative, i.e. $g(a,a) = -da < 0$, $g(-a,-a) = -d(-a) > 0$. For $d < 0$, the signs reverse themselves. The resulting configuration of the slow manifold for $d > 0$ contains a z-shaped curve in one continuous piece, which engineers also like to

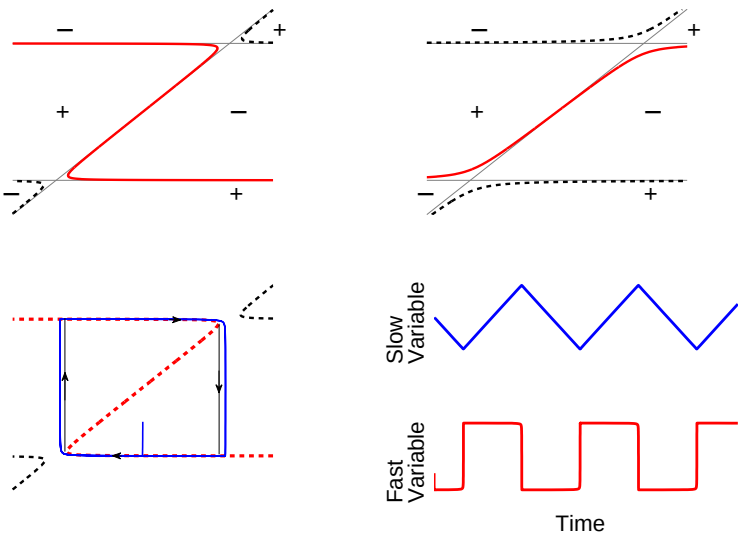


Figure 1.1: Top: two configurations of the equilibrium solutions of the fast subsystem. $\pm$ are the signs of g off the slow manifold. Bottom left: the singular limit cycle at $\epsilon = 0$ and a perturbed orbit for $0 < \epsilon \ll 1$. Bottom right: the time series of the perturbed solution.

call it a *hysteresis*. The morphosis of the z-switch in parameter d is continuous. The curve converges to the three line branches of $g = 0$ as d goes to zero.

For $d > 0$ the solution set of $g = 0$ also contains two remnants, one to the right of the upper right knee and one to the left of the lower left knee of the z-switch. As it will become clear soon the two remnants do not play any essential role for all the intents and purposes. In other words, if the limit cycle dynamics is a suit made from the fabric of the 2-dimensional phase space, then these two corner remnants are part of the throw-away scrap. As a result we will ignore them and pretend the z-switch is the whole zeroes of g from now on. In fact, for some more sophisticated choices of g a z-switch comprises all the zeroes of g without any remnants.

For $d = 0$ the turning points become the intercepts of the horizontal lines and the slanted line. Such an intersection by two different branches of the solution $g = 0$ at an angle is called a *transcritical point*. They are fairly common for population models. Like the turning points they too act like switches, but with their own peculiarities which we will discussion in greater details later.

This typically completes the description of the fast equation $\epsilon\dot{y} = g$ at the singular value $\epsilon = 0$. And we will assume without a loss of generality that the slow manifold is a z-switch as depicted in Fig.1.1.

Now look at the slow x-equation of the slow form at the singular limit $\epsilon = 0$. The immediate revelation is that it is a 1-dimensional differential equation constrained on the slow manifold $g = 0$. It will soon become clear that only the dynamics of the slow equation on the top and bottom branches of the z-switch matter. As noted before the once 2-dimensional full system is only 1-dimensional on the z-switch. Moreover, depending on the dynamics we want to achieve, we can construct the function f on each of horizontal branches accordingly, and do so independently from each other.

In fact, by design f is made to be a linear combination of two arbitrary functions f_{-1} and f_{+1}:

$$f(x, y) = (y + a)f_{+1}(x, y) + (a - y)f_{-1}(x, y).$$

When restricted to the bottom branch $y = -a$ of the d-limiting z-switch, $\dot{x} = f = 2af_{-1}$, the f_{+1} term is conveniently killed off. Similarly, on the top branch $y = +a$, $\dot{x} = f = 2af_{+1}$, completely independent from the bottom branch assignment. For the particular system (1.5), we assign $f_{+1} = b > 0$ so that every trajectory, $(x(t), a)$, of the reduced slow equation moves to the right on the top branch. On the bottom branch, $\dot{x} = 2af_{-1} = -2ac < 0$ for which every trajectory moves to the left, again by design.

1.1.4 Fast Transition Between Slow Manifolds

Next, consider the fast form (1.7). At the singular limit $\epsilon = 0$,

$$x' = 0 \text{ and } y' = g(x, y).$$

It too is reduced from a full 2-dimensional system to a 1-dimensional equation. In this case, the slow variable frozen at its initial $x(t) \equiv x_0$ because $x' = 0$, as if x is a parameter for the fast equation. The dynamics of the fast equation is completely determined by the equilibrium set $g(x, y) = 0$, i.e. the slow manifold: if $g > 0$ at an initial point, $y(\tau)$ moves up, and if $g < 0$ at an initial point, $y(\tau)$ moves down, often converges to a point on the slow manifold.

Depending on the location of x_0 the number of equilibrium points of the fast y-equation changes (ignoring the remnants of $g = 0$ as discussed above). Let a_- denote the x-coordinate of the left turning point of the z-switch and a_+ the right turning point of the z-switch. Then for $x_0 < a_-$ the y-equation

has one equilibrium point on the top branch of the z-switch. For $x_0 > a_+$ the equation has one equilibrium point on the bottom branch of the switch. For $a_- < x_0 < a_+$ the equation has three equilibrium points: one on each of the top, bottom, and middle branch of the z-switch. For any initial y_0 not on the slow manifold, the trajectory of the y-equation can only move up or down, away or toward the equilibrium point(s), depending completely on the sign of $g(x_0, y_0)$.

Here is another easy handling of g: by multiplying -1 to g if necessary we can make the top and bottom branches of the equilibrium points on the z-switch either attractive or repulsive for the fast y-equation. All it takes is to make sure $g < 0$ for points right of the z-switch if we want y to move down, which in turn automatically forces $g > 0$ to the left since the z-switch is where $g = 0$. It suffices to only pick one point (x_0, y_0) far above the top branch $y = a$ so that $g(x_0, y_0) < 0$ by the so-called one-point testing rule.

1.1.5 Slow And Fast Concatenation

As if in an orchestra the individual players are now assigned their parts, placed wherever needed be, and directed to do whatever instructed. The show is ready to start. Once it is switched on by a small positive value $\epsilon > 0$, regardless whichever form, slow or fast, any trajectory that starts near the z-switch is attracted to one limit cycle. It consists of four pieces of movement, alternating in fast and slow tempo progressions, and dominated in turn and in tandem by the slow x-equation and the fast y-equation.

Start from any point near the z-switch. Say near the bottom branch. Whether it is near, on, or far away from the z-switch, it is quickly attracted to the bottom branch the z-switch. Because the slow variable barely moves, the trajectory moves nearly vertical. This behavior changes only when the trajectory enters an ϵ-neighborhood of the slow manifold $g = 0$. In which the minuteness of g balances out the smallness of ϵ so that magnitudes of the rates $\dot{x} = f$ and $\dot{y} = g/\epsilon$ become comparable, and only in which the trajectory can appear to turn this or that way with comparable speeds for both variables. Subsequently, the trajectory will be forcefully pinned inside the neighborhood, tightly against the bottom branch the z-switch because of the strong contracting action of the fast y-equation. It is inside such a neighborhood, the dynamics of the fast equation takes over. In this case, because $\dot{x} = f(x, y) \sim 2af_{-1} = -2ac < 0$ near the bottom branch of the switch, the trajectory moves leftward, slowly.

Eventually, the trajectory enters a neighborhood of the left turning point the z-switch, moving leftward still as $\dot{x} < 0$. Once it clears, say, a $\sqrt{\epsilon}$-neighborhood

of the turning point, g is not small enough to counter balance the divisor ϵ in the rate $\dot{y} = g/\epsilon$. As a result, the fast dynamics takes over. The trajectory moves quickly upward because $\dot{y} = g > 0$, racing toward the top branch of the z-switch. The trajectory looks like a vertical line during this phase as the slow variable x again barely moves.

Once it enters an ϵ-neighborhood of the top branch of $g = 0$, the strong effect of $1/\epsilon$ to the y-variable is again neutralized, and the movement on the top branch is taken over by the slow motion of the x-equation. This time, the trajectory is tugged rightward since $\dot{x} \sim f(x, a) > 0$ as $y \sim a$, hugging ever so tightly the z-switch like it did along the bottom branch. This phase lasts until the trajectory enters and then leaves a $\sqrt{\epsilon}$-neighborhood of the right turning point of the z-switch. A fast phase transition crisis takes place similar to what has happened at the left turning point. This time instead of going up it is crashing down toward the bottom branch of the z-switch because $\dot{y} = g/\epsilon < 0$ to the right of the switch. This sets up another repeat of the roundabout around the z-switch. The limit cycle is thus materialized.

Limit cycle around a hysteresis of a fast-slow system has been known and well-understood since the 1920s. We know for example the limit cycle orbit is continuously dependent on the singular parameter ϵ. As $\epsilon \to 0$, the perturbed orbit of $\epsilon > 0$ converges to the so-called *singular orbit*, which is the concatenated loop of two limiting fast orbits and two slow orbits. Which piece to start is unimportant. Start for instance the fast orbit of the fast equation from the left turning point of the z-switch. It moves up and ends on the top branch of the slow manifold. It forms a T with the top branch. We call the point a *junction point*. Its x-coordinate is the same as the turning point $x = a_-$ and its y-coordinate is from the solution $g(a_-, y) = 0$. Next in the concatenation is the slow orbit of the slow equation from the previous junction point to the right turning point whose x-coordinate is a_+. Then the descending fast orbit of the fast y-equation from the right turning point to its corresponding junction point on the bottom branch of the slow manifold. Lastly, the slow orbit from the right junction point to the left turning point to complete the singular orbit loop in concatenation. The perturbed limit cycle with $\epsilon > 0$ closely tracks along this singular limit cycle as shown in Fig.1.1.

The figure also depicts the time series of the two variables. It characterizes such multi-time scale systems in the time profile. Usually you cannot tell the singular perturbation nature of a system from the slow variable or experiment data. It can behave just like from a regular system. But the dynamics of the fast variable cannot be more revealing: there will always be fast rises and falls,

followed by mild, slow movements, not necessarily like the horizontal lines for the mockup system. The take-away message is if one sees a fast transition in a datum, then the underline process must be of different time scale.

1.2 Systems By Custom Design

The types of 3-dimensional structures that can be realized by Rössler's dual principle are only limited by one's imagination. Whatever dynamics one can dream up the chance is there is a way.

Collected below are some examples. Each makes use of a particular type of switch. Instead of a curve, all switches are surfaces in the 3-dimensional space. They are the slow manifolds of the full systems. And all are defined as the equilibrium points of the fast equation. The switches are either shaped like a Z or have something to do with it. One type is to have two z-switches to share a common middle branch, one z-switch glued on top of another. One type is to rotate a z-curve around an axis to get a rotationally symmetrical z-switch surface. One type is to have a regular z-curve to shrink in size until it disappear on the slow manifold switch.

The reduced slow dynamics on the switches are all 2-dimensional. Like in the mockup 2-dimensional system above, the right-hand side of the slow equations can be prescribed arbitrarily on separate attracting branches of the switches. For most cases we will use the simplest linear systems possible. They take the form $\dot{x} = a(x - \bar{x}) + b(y - \bar{y})$, $\dot{y} = c(x - \bar{x}) + d(y - \bar{y})$ with $(\bar{x}, \bar{y})$ being an equilibrium point of the slow system anywhere we wish to put. We can choose the constants a, b, c, d so that the equilibrium $(\bar{x}, \bar{y})$ is a sink, a source, a saddle, or a spiral kind of all the above. For some systems we also add a cubic nonlinearity to control the outward spiral by a limit cycle. Writing such a slow-subsystem in the polar coordinate centered at the equilibrium point, it reads $\dot{r} = \alpha r(1 - r^2/R^2)$, $\dot{\theta} = \beta$ with R being the radius of the limit cycle. It is essentially a linear system with an outward spiral equilibrium point. Only in two examples the reduced the slow system are truly nonlinear. Even in these cases the functional forms are textbook type examples for some typical 2-dimensional dynamics, which are considered simple for experts. More specifically, they are the so-called Hamiltonian systems of the form

$$\dot{x} = \frac{\partial H}{\partial y}, \quad \dot{y} = -\frac{\partial H}{\partial x},$$

with $H(x, y)$ a function of two variables. We use such systems because their solutions are the level curves of function $H(x, y)$, called the Hamiltonian. So if

we know H we know the equation's solution. Thus, depending on the type of orbits we like to have, we go out to look for the right kind of functions H whose level curves fit the description.

1.2.1 Fold At The Interface Of Fast And Slow Flows

For a fast-slow 2-dimensional full system, there are not many ways for the dynamics to be more complex than the limit cycle type described above. One way is to make one of the junction point an equilibrium point of the slow equation. The limiting orbit is a *singular homoclinic orbit* and for each small ϵ one can adjust the z-switch a bit if necessary to obtain a homoclinc orbit for the perturbed system, and orbit that converges to the equilibrium point in both forward and backward time. The other alternative orbit is obtained by making the junction point an equilibrium point on both branches of the slow manifold. Such an orbit is a heteroclinic orbit.

There are a whole lot more one can do for the 2-dimensional slow subsystems. Lining up equilibrium points with junction points is certainly one simple trick in the book. Here we want to describe two complications that are unique for slow systems 2-dimensional or higher. They are depicted in Fig.1.2. They are referred to as the K-fold points. There are two kinds.

The first kind is the junction point type. In this case the K-fold point is in the interior of the slow manifold. The flows near it are regular. It is not an equilibrium point of the slow equations. Away from the branch there is a turning edge of the slow manifold. The corresponding junction points form a curve through the K-fold point. From one side of the backbone of the K, the slow orbits flow to one side of the K and from the other side, the slow orbits move to the opposite side of the K. The slow orbit forms an elbow at the point against the backbone. In the depicted junction K-point, the flows in black are down on the slow manifold, the flows in blue are atop of the black, and the flows in red are atop of the blue. In fact, this is exactly what must happen when the perturbation is switched on with $\epsilon > 0$. That is, carried on by the flows near the slow manifold, one side of the junction edge will be folded underneath the other side of the backbone. Hence the moniker junction point K-fold.

The other type of K-fold point is also on a branch of the slow manifold. But instead in the interior this point is on the turning edge of the slow manifold, which the the backbone of the K. The slow flows also form an elbow at the point. On one side of turning edge the slow flows rush on and make the singular fast jump away from the slow manifold. One the other side of the backbone of

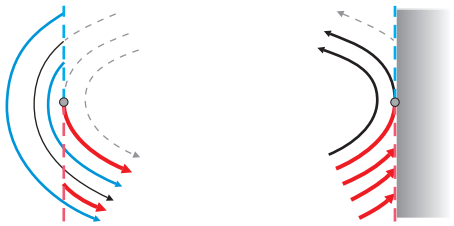


Figure 1.2: Left: Junction K-fold point. Right: Turing K-fold point.

the K the slow flows stay on the slow manifold. But they are not accessible from the interior side of the elbow.

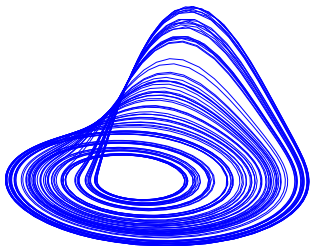
Interesting dynamics happen if the K-fold points are reinjected back into the slow manifolds they belong. In both cases the flows of the full systems will be folded up and if they are also stretched in the backbone direction of the K, chaotic dynamics usually ensue.

1.2.2 Fold The Fold

Fig.1.3 shows two chaotic attractors: the original Rössler attractor and one constructed having a turning K-fold point:

$$\begin{cases} \dot{x} = \zeta(z+r) + (r-z)[\alpha(x-2)h - \beta y] \\ \dot{y} = (z+r)[b + \frac{c-b}{-2-a}(x-a) - y] + (r-z)[\beta(x-2) + \alpha y h] \\ \epsilon\dot{z} = (r^2 - z^2)[z + r - m(x+r)] - dz \\ h = 1 - [(x-2)^2 + y^2]/R^2 \end{cases} \tag{1.8}$$

The xy-equations are the slow equations and the z-equation is the fast equation. The slow manifold, consisting of the equilibrium points $\dot{z} = 0$ of the fast equation, has a fixed z-shaped in the xz-variables for every value of the y variable. On the top of the z-switch the slow orbits flow to the top turning edge, and then fall down to the bottom branch. On the bottom branch, trajectories spiral outward. The turning K-fold attractor was made to simulate the Rössler attractor. The Rössler attractor and the mockup attractor are structurally similar. The chaos is solely due to the presence of a turning K-fold point. The Rössler is deceptively simple as it is hard to deconstruct it down to its components. The mockup system looks complicated but it is easy to understand every component of the equations and to build the chaos bottom up by strategically placing a K-fold turning point.



The Rössler Attractor

$$\begin{cases} \dot{x} = -y - z \\ \dot{y} = x + ay \\ \dot{z} = b + z(x - c) \end{cases}$$

with parameter values $a = 0.2, b = 0.2, c = 5.7$.

Turning K-Fold Attractor:

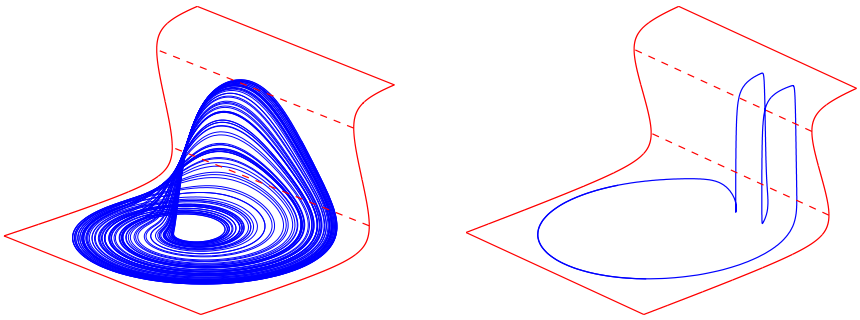


Figure 1.3: Bottom left: Parameter values of Eq.(1.8): $a = 0.5, b = 0, c = -1, m = 0.4, r = 0.5, R = 6, \alpha = 0.5, \beta = 2, d = 0.08, \zeta = 10, \epsilon = 0.1$. Right: Same values except for $\epsilon = 0.001$ at which the effect of the fast variable's switching becomes more dominant.

Fig.1.4 shows the type of chaotic attractors that uses junction K-fold points. It belongs to this custom-made system:

$$\begin{cases} \dot{x} = \lambda z(x - a) + (3 - z) \times [\alpha(x + 0.5)h + \beta y] \\ \dot{y} = \mu z(y + b) - (3 - z) \times [-\beta(x + 0.5) + \alpha y h] \\ \epsilon \dot{z} = z(3 - z)(x + z - \frac{3}{2}) + d(x - c) \\ h = 1 - [(x + \frac{1}{2})^2 + y^2]/R^2. \end{cases} \tag{1.9}$$

The construction of the fast z-switch is similar to the previous example. In this case we kept the K-fold turning point away from the attractor on the lower branch of the z-switch. On the top branch we carefully fan out the flow and direct it to the turning fold to create a junction K-fold point on the bottom branch of the switch. The result is a pure junction K-fold attractor. The figure shows the top view of the attractor.

Junction K-Fold Attractor:

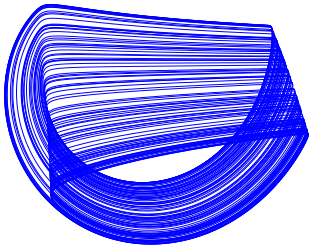


Figure 1.4: Parameter values of Eq(1.9): $a = -5.5, b = 0, c = 0.5, R = 6, \alpha = 6.2, \beta = 20, \lambda = -1.2, \mu = 0.87, d = 0.015, \epsilon = 0.006$.

1.2.3 Fraternal And Paternal Twin Spirals

Fig.1.5 shows two more uses of the z-switch. In both cases the slow flows on the top branch and the bottom branch of the z-switch are spirals. Both are generated by this construction:

$$\begin{cases} \dot{x} = (z+2)(\lambda(x-k)h - \mu y) + (2-z)(\alpha(x-2)h - \beta y) \\ \dot{y} = (z+2)(\mu(x-k) + \lambda y h) + (2-z)(\beta(x-2) + \alpha y h) \\ \epsilon\dot{z} = (4-z^2)[z+2-m(x+2)] - dz \\ h = 1 - \frac{(x-k)^2+y^2}{R^2} \end{cases} \tag{1.10}$$

For parameter $R = \infty$ both spirals are the logarithmic spirals and the slow equations are near linear. For parameter $\mu > 0$ both spirals are going counterclockwise. For $\mu < 0$ each rotates in a different direction. One spiral center is at $(x, y) = (2, 0)$ and the other is at $(x, y) = (k, 0)$ for the slow subsystem. Parameter m can be used to control the top turning edge of the z-switch and parameter k can be used to control the location of the top center. Each attractor contains two turning K-fold points and two junction K-fold points, one pair on each of the branches of the z-switch.

Double-Spiral Attractors:

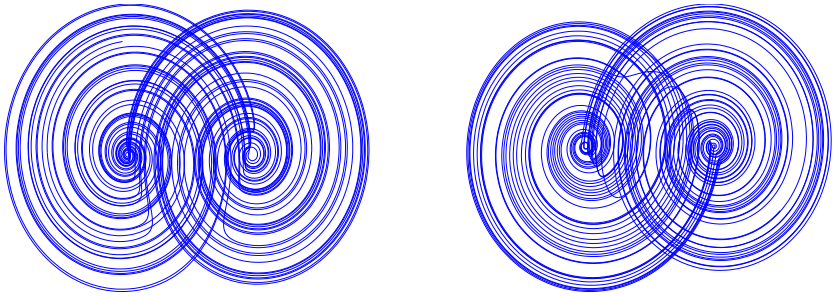


Figure 1.5: Left: Parameter values of Eq.(1.10): $a = 7, b = 1.435, d = 0.08, R = 6, \alpha = 0.5, \beta = 4, k = -3.3, m = 1.543, \lambda = 0.58, \epsilon = 0.08, \mu = -4$. The left spiral is clockwise. Right: The same except $\mu = 4$. The left spiral reverses itself in direction.

1.2.4 Toroid Braid

For Fig.1.6, the slow manifolds are slightly more intricate. The made-up system is

$$\begin{cases} \dot{x} = z(\lambda x - \mu y) + (2-z)\big[\alpha x\big(1 - \frac{x^2+y^2}{R^2}\big) - \beta y\big] \\ \dot{y} = z(\mu x + \lambda y) + (2-z)\big[\beta x + \alpha y\big(1 - \frac{x^2+y^2}{R^2}\big)\big] \\ \epsilon\dot{z} = z\big[(2-z)[a(z-2)^2 + b] - cx\big]\big[z + m(x^2+y^2) - h\big] \\ \qquad\qquad -d(z-1) \end{cases} \tag{1.11}$$

For the fast equation of the first construction, we consider first the case with these parameter values: $a = c = d0,\ b = 1$:

$$\epsilon\dot{z} = z(2-z)\big[z + m(x^2+y^2) - h\big].$$

The slow manifold consists of a horizontal top branch $z = 2$, a horizontal bottom branch $z = 0$ just like the various z-switches above. Unlike the others the slanted middle plane is replaced by the paraboloid

$$z = h - m(x^2+y^2).$$

It intersects the top branch along a circle $x^2 + y^2 = (h-2)/m$. It intersects the bottom branch along a bigger circle $x^2 + y^2 = h/m$. If you take a look at the cross section of the branches on the xz-plane, there are two z-switches imbedded, one is the mirror image of the other. If you rotate the cross section

Toroid Attractor:

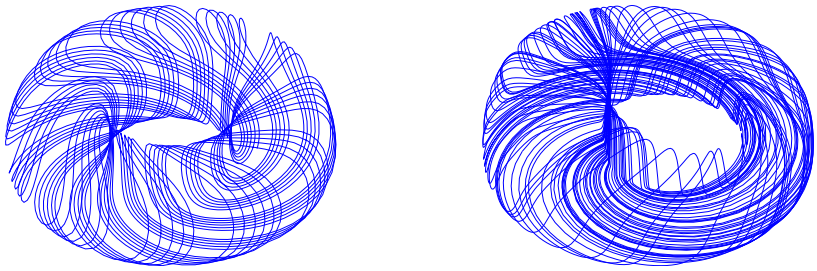


Figure 1.6: Left: Parameter values of Eq.(1.11): $a = c = 0, b = 1, d = 0.1, m = 0.1, h = 3.6, R = 10, \lambda = -4, \mu = -4, \alpha = 2, \beta = 4, \epsilon = 0.1$. Right: $a = 3, b = 0.8, c = 0.1, d = 1, m = 0.05, h = 3.312, R = 10, \lambda = -2, \mu = 1, \alpha = 2.8, \beta = 5, \epsilon = 0.1$.

around the z-axis, you will get the full slow manifold. Also, when the parameter $d > 0$ is turned on, you will get the two mirror double z-switch on the xz-cross section, and the full slow manifold of interest is the curve's rotation about the z-axis. We can call it a toroid-switch instead.

As before the slow equations on the top and bottom branches are linear systems of equations. On the top branch, all orbits are pulled inward the center $(0, 0)$. On the bottom, all orbits are pushed outward away from the center. Rotation of the slow orbits is added by using nonvanishing values of μ and β. As shown in the figure, an invariant toroid is obtained. Orbits on the invariant toroid can be periodic, quasi-periodic, but no chaos. No K-fold points of either kind are present.

The terms $a(z-2)^2 + b$ and $-cx$ are purposely included so that the toroid switch is not rotational symmetric. The top turning edge shapes like an egg instead of circle with $a = c = 0$. When projected down to the bottom branch as the junction points, the asymmetric curve will create a junction K-fold point when meeting the symmetric slow flows on the bottom branch of the toroid switch. Chaos breaks out because of the existence of the K-fold point. In other words, the once smooth invariant toroid is folded as a place. Orbits from the attractor pile on top of each other near the K-fold point. The attractor on the right of Fig.1.6 is one example of such folded toroid attractors.

Double-Toroid Attractor:

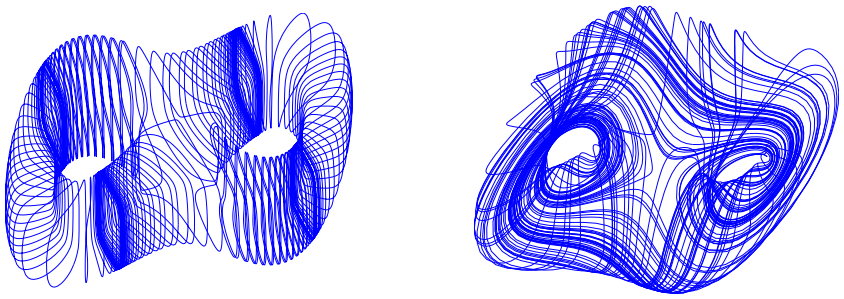


Figure 1.7: Left: Parameter values of Eq.(1.12): $a = b = 1, d = 0.06\alpha = \cos\theta, \beta = \sin\theta, \theta = .08, \epsilon = 0.06$. Right: chaotic attractor for a modified system.

1.2.5 Wind Around Conjoined Toroid

For the system of Fig.1.7, the dynamics look more complicated. However the design idea is the same as the toroid attractor above. The terms used are more complex than the toroid attractor, but they are the simplest kind to realize the double toroid construction:

$$\begin{cases} \dot{x} = a(1.5-z)\frac{x(x^2-8)}{16} + bz\big[\alpha(\alpha x+\beta y)\frac{8-(\alpha x+\beta y)^2}{16} - \beta(\beta x-\alpha y)\big] \\ \dot{y} = a(1.5-z)y + bz\big[\beta(\alpha x+\beta y)\frac{8-(\alpha x+\beta y)^2}{16} + \alpha(\beta x-\alpha y)\big] \\ \epsilon\dot{z} = z(1.5-z)\big[z-1+H(x,y)\big] + d(\frac{3}{4}-z) \end{cases} \tag{1.12}$$

As for the slow manifold, or the double-toroid switch, the top and bottom branches are horizontal plane. Instead of the paraboloid which only one hump, we use a surface having two humps, which is represented by the equation

$$z = 1 - H(x,y), \text{ with } H = \frac{(x-y)^4}{64} - \frac{x^2-y^2}{2} + \frac{(x+y)^2}{4}.$$

Therefore, instead one hole this surface cuts out two holes on the top branch $z = 1.5$. As for the toroid switch, only one hole is cut out by the surface. In terms of level curves, the $z = 1.5$ level curve consists two closed curves and the $z = 0$ level curve contains only one. The d-term is strategically place to break up the branches in the right ways so that the top and bottom cutouts become turning points of the double-toroid switch.

For the slow equations on the top and bottom branches, the requirements are far more different than what we have encountered so far. Linear systems

are not enough anymore. For example, on the bottom branch $z = 0$, the slow equations are

$$\dot{x} = a(1.5)x(x^2 - 8)/16, \qquad \dot{y} = a(1.5)y$$

for which there are three equilibrium points $(0, 0), (\pm\sqrt{8}, 0)$. It is made so that the $(0, 0)$ equilibrium point is a saddle, and the other two equilibrium points are sources. The slow equations on the top branch is made the same except in two ways. First the flow's direction is reversed so that two equilibrium points on the wing are two sinks. Second, the vector field is slightly rotated by θ angle, which explains the parameters $\alpha = \cos\theta, \beta = \sin\theta$. One invariant double toroid is created as shown. Six orbits are shown. Each converges to one of the four limit cycles.

Lastly, the double-toroid attractor can be made chaotic by creating K-fold turning points and K-fold junction points. The system of equations is the same as the double toroid equations except that the bottom slow equations are replaced by this perturbed Hamiltonian system instead:

$$\dot{x} = a(1.5 - z)\left(-\frac{\partial H}{\partial y}\right), \qquad \dot{y} = a(1.5 - z)\left(\frac{\partial H}{\partial x} + cy\right), \text{ with } z = 0,$$

where H is the same function used to construct the double-toroid switch. The trajectories of the unperturbed equations with $c = 0$ are the level curves of H. The perturbed Hamiltonian flow is enough to break up some inherent symmetry of the double-toroid system. From the chaotic attractor one can see at least two K-fold junction points, one from each of the inner junction curves.

1.2.6 Fast-Slow Butterfly

The French mathematician Henri Poincaré is perhaps the first person to encounter the phenomenon of chaos in a mathematical model. That was in the 1880s. The American meteorologist Edward Lorenz is considered the founding father of the chaos theory by his discovery of the Lorenz attractor. That discovery was made in 1963. His equations are simple but extremely hard to analyze. Here is the last fast-slow contraption of this section:

$$\begin{cases} \epsilon\dot{x} = (y - p) - 5(x^3 - (z - q)x) \\ \dot{y} = \frac{(x+1)^2}{4} f(y, z, a_1, b_1, c_1, \theta_1) - \frac{(x-1)^2}{4} f(-y, z, a_2, b_2, c_2, \theta_2) \\ \dot{z} = \frac{(x+1)^2}{4} g(y, z, a_1, b_1, c_1, \theta_1) + \frac{(x-1)^2}{4} g(-y, z, a_2, b_2, c_2, \theta_2) \end{cases} \tag{1.13}$$

with the ingredient functions given by

$$\begin{bmatrix} f(y,z,a,b,c,\theta) \\ g(y,z,a,b,c,\theta) \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \phi(\bar{y},\bar{z},a,b,c) \\ \psi(\bar{y},\bar{z},a,b,c) \end{bmatrix}$$
$$\phi(y,z,a,b,c) = ay + bz - ay^2$$
$$\psi(y,z,a,b,c) = by + az - \tfrac{3}{2}by^2 - (\tfrac{3}{2}ay + c)z$$

and the rotated slow variables

$$\begin{bmatrix} \bar{y} \\ \bar{z} \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} y \\ z \end{bmatrix}$$

We now break it down on how his butterfly effect is realized.

For visualization reason only, x is designated the fast variable, y, z the slow variables. The slow manifold is a surface that is rather easy to understand. It is a cusped z-switch – whose turning edges meet at a point, forming a cusp when looked from the x-axis, and hence the term for the switch. Beyond the cusp point the once top and bottom branches of the switch merge into one smooth sheet. These descriptions can be easily verified by looking at the cross section curves for fixed z-values. In fact, the manifold in the sectional curve form is a cubic polynomial, $y = p + 5(x^3 - (z-q)x$, in x having precisely two turning points in the fast flow direction when $z > q$. For $z < q$ the turning points disappear and the two branches merge into one continuum sheet.

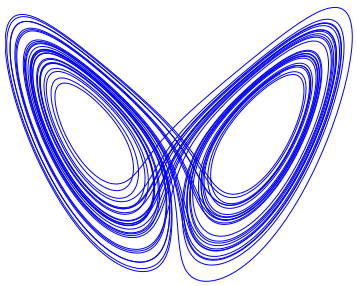
For the slow equations in variables y, z, their constructions only approximate what we like to have on the cusp switch. In particular, we want to place the wings of the butterfly on the switch, one on the back sheet near $x = -1$ and one on the front sheet near $x = 1$. Since the front and back branches are not the constant planes $x = \pm 1$, the slow equations,

$$\mathbf{v}' = \frac{(x+1)^2}{4}\mathbf{F}_{+1}(\mathbf{v}) + \frac{(x-1)^2}{4}\mathbf{F}_{-1}(\mathbf{v}), \text{ with } \mathbf{v} = \begin{bmatrix} y \\ z \end{bmatrix},$$

are not constructed with the absolute precision we have achieved for the other examples. For example, the frozen equations with $x = +1$, $\mathbf{v}' = \mathbf{F}_{+1}(\mathbf{v})$ is only an approximate of the actual restricted vector field on the curvy front sheet near $x = 1$.

Because of the inherent symmetry of the two wings of the butterfly, the vector fields $\mathbf{F}_{+1}$ and $\mathbf{F}_{-1}$ are reflective symmetric in the y variable. That is,

$$\mathbf{F}_{+1}(\mathbf{v}) = \begin{bmatrix} f(y,z) \\ g(y,z) \end{bmatrix} \text{ if and only if } \mathbf{F}_{-1}(\mathbf{v}) = \begin{bmatrix} -f(-y,z) \\ g(-y,z) \end{bmatrix}.$$



Lorenz Attractor:

$$\begin{cases} \dot{x} = \sigma(y - x) \\ \dot{y} = x(\rho - z) - y \\ \dot{z} = xy - \beta z \end{cases}$$

parameter values: $\sigma = 10$, $\rho = 28, \beta = 8/3$.

Fast-Slow Lorenz Attractor:

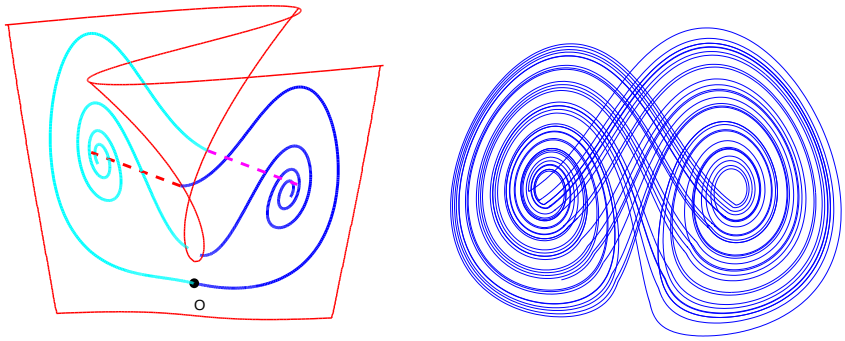


Figure 1.8: Left: Cusp-switch and singular orbits. Right: Parameter values of Eq.(1.13): $a_1 = a_2 = 0.5, b_1 = b_2 = -4, c_1 = c_2 = -1, \theta_1 = \theta_2 = \pi/4, p = 0, q = 0.1, \epsilon = 0.005$.

This explain the general form of the constructed slow equations, and the reason that we only need to specify the vector field (f, g).

This vector field is the rotation of the specific vector field (ϕ, ψ) by angle θ. That is what the rotation matrixes in cos, sin and in $\bar{y}$, $\bar{z}$ are about. Thus, our deconstruction comes to the last phase to explain the 2-dimensional system of equations $\dot{y} = \phi(y, z), \dot{z} = \psi(y, z)$, omitting the dependency on the parameters. For people familiar with some exotic 2-dimensional systems, not constructed by singular perturbations, this type of systems does not invoke surprises. One can check that without the perturbation terms with $a = c = 0$, it is an Hamiltonian system $\dot{y} = \partial H/\partial z$, $\dot{z} = -\partial H/\partial y$ with $H = \frac{b}{2}\left(z^2 + y^3 - y^2\right)$. All solutions are the level curves of H. In particular, $(0, 0)$ and $(2/3, 0)$ are two equilibrium points with the former a saddle and the latter a center. The zero level curve

$z^2 + y^3 - y^2 = 0$ contains also a homoclinic orbit to the origin, going around the center equilibrium point, counterclockwise. The perturbations are added to break this homoclinic orbit and to control the eigenvalues at the origin which are $a \pm b$.

Upon a $\theta = 45^\circ$ rotation of the homoclinic loop and everything else going with the vector field, the front wing to the butterfly is in place. With the reflection the back wing is also done. Both are held together at the equilibrium point $(0, 0, 0)$ for the full system. The cusp switch and the orbits on it as depicted in Fig.1.8 are the actual objects of the constructed system. It does not show any K-fold turning point above the cusp point. Although there are plenty K-fold junction points but the attractor does not seem to use any. The cusp point, the holding equilibrium point, the eyes of the wings are the elements responsible for the butterfly effect. These features are decidedly distinct from the double-spiral attractors even though their appearances make them look alike if the vantage point is right.

The one property that is included in all working definitions of chaos is the one that Lorenz emphasized the most with his discovery, namely the sensitive dependence on initial conditions. Like many discoveries made throughout the history, his came also with a pinch of serendipity. It happened when Lorenz set his equations running to double check an early numerical run, walked away for a cup of coffee, and only to find when he returned the new result was miles away from the original simulation. He found that any small perturbation to an initial state would result in the system bouncing between the two wings of the attractor in different orders. Since his model was related to weather prediction, this so-called sensitive dependence on initial condition property gave birth to the now frequently quoted butterfly effect metaphors. That the flap of a butterfly's wing in Brazil can set off a tornado in Texas is one of the variations.

1.3 Fast-Slow Chaos of Models

We can learn a lot from the multitime scale construction above and use the lessons to understand mathematical models for physical processes. We learned that unique to systems of three dimension and higher the turning-point K-fold and the junction-point K-fold are two primary mechanisms for chaos generation. We also learned that if it were not because of various nonlinear switch surfaces on which the slow processes of a system dominate there would be no reoccurring of the folded fast and slow flows, and the dynamics would be less than chaotic. Here below let's take a look at a few models for which the learned lessons are

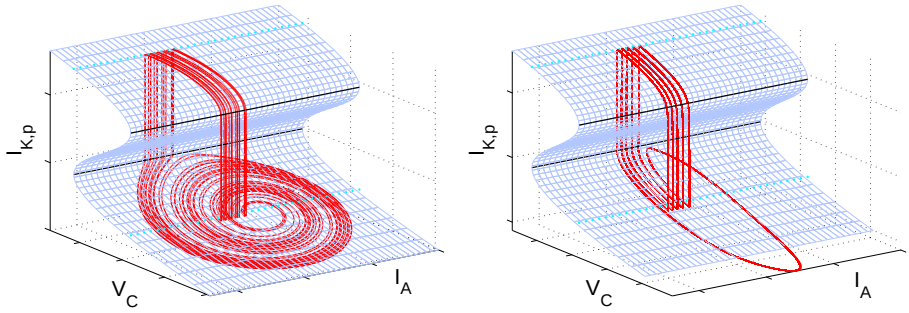


Figure 1.9: Left: Parameter values of Eq.(1.1): $E_{\rm Na} = 0.5, E_{\rm K} = -0.7, r_1 = 10, r_2 = -12.2, i_1 = 0.24, i_2 = 0.43, G_1 = 15, G_2 = -0.37, V_1 = -1, V_2 = -0.02, \lambda = 0.055, C = 2.2, R = 0.18, \epsilon = 0.001$, and zeroes for other parameters. Right: The same values except for $C = 0.1$.

particularly insightful.

1.3.1 Neuronal Models

Fig.1.9 shows a chaotic attractor for the neuronal circuit model Eq.(1.1). Variable $A_{\rm K}$ is the ϵ-fast variable and the others are the ϵ-slow variables. The Z-switch is the hysteresis surface of the $A_{\rm K}$-equation. It is the ϵ-slow manifold. For healthy neurons, the capacitance C is small for normal cell operations. This condition creates another time scale amongst the ϵ-slow variables for which the membrane potential is a C-fast variable, enhancing various cell acuities. In abnormal situations such as demyelination the cross-membrane capacitance degrades to become large. The cell loses a fast time scale and there is no further temporal distinction between the cross-membrane potential and the ion pump current variables. The cell dynamics becomes chaotic. The simulation shows that a turning-point K-fold on the lower branch of the Z-switch leads to the type of chaotic attractor similar to the Rössler attractor. It also shows that for a small capacitance the model is in an expected normal state of spike burst. The K-fold point is always present for the range of the capacitance. It takes the combination of both the point and the capacitance's degradation for the chaotic state of the cell to occur.

1.3.2 Food Chain Models

The attractor on the left of Fig.1.10 is for the food chain model Eq.(1.4). It is for the case in which the prey X reproduces fast. They are modeled by small ε values. Between the predator Y and the top-predator Z the latter is the slowest, modeled by small ζ values. The X-nullcline surface is the slow manifold. It consists of the prey extinction plane $X = 0$ and the cylindrical parabola, i.e. the predator mediated prey equilibria. The top ridge on the parabola separates the surface into two parts – the capacity branch part and the Allee branch. The capacity branch and extinction branch are the attracting branches of the slow manifold and the Allee branch is the unstable middle branch. Together they makeup a variant z-shaped switch for the fast prey species, if the parabola fold occurs in the first octant phase space.

On the capacity branch of the X-nullcline surface, the fast flows hit a junction K-fold. To the top-predator's higher population side of the K-fold, the predator population is depressed still even though it is supported by the prey's capacity mass. To the top-predator's lower concentration side of the K-fold, the predator prospers, heading up in population, all taking place on the ε-slow manifold. The creation of the z-shaped switch, i.e. the turning edge at the top of the X-nullcline surface, is primarily due to an ineffective predator in the middle, not good at preying on the prey. In particular, having a too large prey handling-time parameter h_1 will create the undesirable bump on X-capacity surface, thus the z-shaped switch. Above the prey capacity ridge, the prey population collapses with higher predator concentration. That collapse leads to a prolonged decline in the predator population which eventually allows the prey to rebound. It is during the fast recovering phase the populations can hit a traffic jam near the junction K-fold point. The chaotic dynamics is due to the creation of the z-switch and the presence of the junction K-fold point along the rebounding line. When the predator is selected to become more effective and efficient this type of population chaos shall disappear.

The tea-cup attractor on the right side of Fig.1.10 is for the same type of food chains except that the top-predator also preys on the prey X, i.e. an intraguild predator. The dimensional model is given below

$$\begin{cases} \dot{X} = bX - mX^2 - \dfrac{a_1 X}{1 + a_1 h_1 X} Y - \dfrac{u_1 X}{1 + u_1 v_1 X + u_2 v_2 Y} Z \\ \dot{Y} = b_1 \dfrac{a_1 X}{1 + a_1 h_1 X} Y - d_1 Y - m_1 Y^2 - \dfrac{u_2 Y}{1 + u_1 v_1 X + u_2 v_2 Y} Z \\ \dot{Z} = \dfrac{r_1 u_1 X + r_2 u_2 Y}{1 + u_1 v_1 X + u_2 v_2 Y} Z - d_2 Z - m_2 Z^2. \end{cases}$$

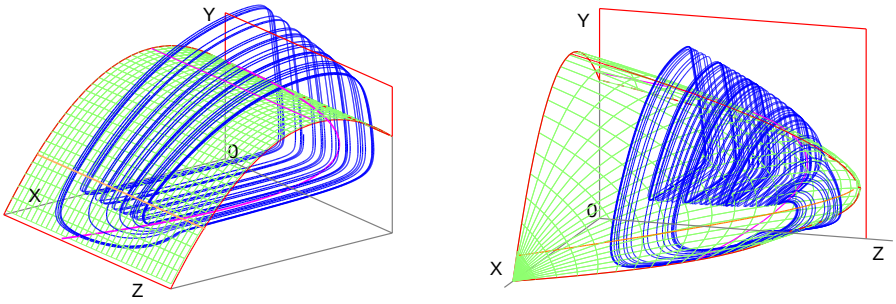


Figure 1.10: Left: Parameter for Eq.(1.4): $\beta_1 = 0.33, \beta_2 = 0.1, \delta_1 = 0.1, \delta_2 = 0.64, \mu_1 = \mu_2 = 0, \zeta = 0.33, \varepsilon = 0.1$. Right: Parameter for Eq.(1.14): $\beta_1 = 0.33, \beta_2 = 0.1, \delta_1 = 0.1, \delta_2 = 0.6, \mu_1 = \mu_2 = 0, u = 1, v = 0.1, w = 0.1, \zeta = 0.3, \epsilon = 0.2$.

It is the same as the previous food chain model except for a joint Holling Type II predation functional for the top-predator on both the prey and predator. With the same parameter scaling and the additional ones below

$$u = \frac{u_1 Z_0}{b u_2 v_2 Y_0}, \ v = \frac{u_1 v_1 K}{u_2 v_2 Y_0}, \ w = \frac{r_1 u_1 K}{r_2 u_2 Y_0}$$

the dimensional model is transformed into the following dimensionless form

$$\begin{cases} \varepsilon \dot{x} = x \left(1 - x - \dfrac{y}{\beta_1 + x} - \dfrac{uz}{b_2 + vx + y} \right) \\ \dot{y} = y \left(\dfrac{x}{\beta_1 + x} - \delta_1 - \mu_1 y - \dfrac{z}{\beta_2 + vx + y} \right) \\ \dot{z} = \zeta z \left(\dfrac{wx + y}{\beta_2 + vx + y} - \delta_2 - \mu_2 z \right) . \end{cases} \tag{1.14}$$

The key difference between the previous tri-trophic model and the intraguild model lies in the X-nullcline surface, the ε-slow manifold. Unlike the tri-trophic model whose X-nullcline does not depend on the top-predator population, the intraguild model does. As one can see from the phase portrait the higher the top-predator population the smaller the prey's capacity surface. The surface tapers as the top-predator's population increases until vanishes altogether. Above such a high top-predator population the prey population collapses. Again the creation of the top fold on the prey nullcline surface is due to an ineffective predator. That coupled with a junction K-fold point on the prey's capacity surface

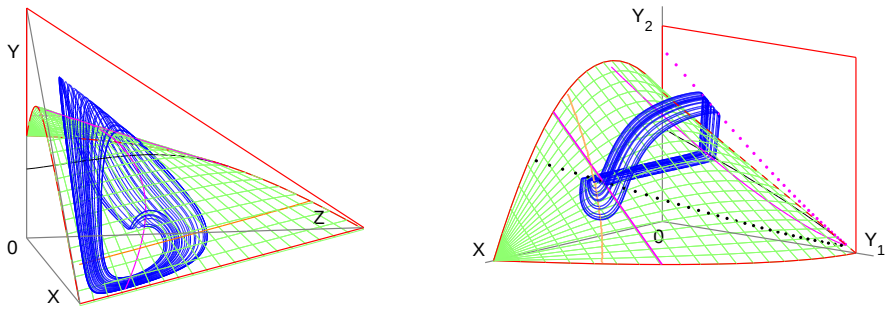


Figure 1.11: Left: Parameter for Eq.(1.15): $a_1 = 0.08, a_2 = 0.23, a_3 = 0.24, m_1 = 10, m_2 = 4, m_3 = 3.5$ and $b = b_1 = b_2 = 1$. Right: Parameter for Eq.(1.16): $\beta_1 = 1.1, \beta_2 = 0.33, \delta_1 = 0.37, \delta_2 = 0.25, \mu_1 = \mu_2 = 0, \sigma_1 = 0.2, \sigma_2 = 0.67, \kappa = 3.3, \zeta = 0.12, \epsilon = 0.01$.

is the key makeup of the attractor. A turning K-fold point is also present in the attractor along the top ridge of the surface, creating an additional place to fold the flows.

Fig.1.11 is for two different types of ecological models. The first is the standard chemostat model for which a fixed amount of nutrient or resource N is present for a food chain. The food chain consists of a prey or producer X, a predator or consumer Y, and a top-predator or top-consumer Z, measured in the same unit as the resource. $S = N - X - Y - Z$ is the amount available for the producer X to synthesize. The resource N is kept at a steady amount for the system and the food chain product is continuously brought out by a constant washout rate w. All species are subject to Holling's Type II rate assimilation constraints. The dimensional model takes the following form:

$$\begin{cases} N = S + X + Y + Z \\ \dot{X} = \dfrac{baS}{1 + ahS} X - wX - \dfrac{a_1 X}{1 + a_1 h_1 X} Y \\ \dot{Y} = \dfrac{b_1 a_1 X}{1 + a_1 h_1 X} - wY - \dfrac{u_2 Y}{1 + u_2 v_2 Y} Z \\ \dot{Z} = \dfrac{b_2 u_2 Y}{1 + u_2 v_2 Y} Z - wZ. \end{cases}$$

for which $b, b_1, b_2 \leq 1$ are the product-to-consumption ratios limited inside the range $(0, 1]$. The washout rate is significant enough to carry out all dead or non-

reactant elements of all causes. Thus no death or elimination rates of any kind are explicitly needed. By a change of variables we can transform the dimensional model to this dimensionless one:

$$\begin{cases} \dot{x} = x\left(\dfrac{bm_1(1-x-y-z)}{a_1+(1-x-y-z)} - 1 - \dfrac{m_2}{a_2+x}y\right) \\ \dot{y} = y\left(\dfrac{b_1m_2x}{a_2+x} - 1 - \dfrac{m_3}{a_3+y}z\right) \\ \dot{z} = z\left(\dfrac{b_2m_3y}{a_3+y} - 1\right). \end{cases} \tag{1.15}$$

The effective phase space is inside the simplex $x + y + z \leq 1$ in the first octant. Although this system is not in the strict singular-perturbation form with explicit small parameters, the multitime property is inherent of the system as the nutrient element N works its way up the chain its assimilation at a higher level becomes slower. Thus, the X-nullcline can be taken to be a default slow manifold. Again if the next consumer is inefficient a capacity fold will develope on the surface. Like the other cases it too becomes the prerequisit for chaotic dynamics. As it is shown by the left portrait of Fig.1.11, it also has a junction K-fold point for the chaotic makeup of the attractor.

The last model to show is the simplest stoichiometric food web model consisting of a prey or producer X and two competing predators or consumers Y_1 and Y_2 of which the second is subject to a stoichiometric constraint. That is, all species depend on more than one nutrients to grow but the prey and first predator always have the right amount balance but the second predator can be limited by the Law of Minimum. More specifically, assume they are measured by the unit of carbon in their population and they also depend on phosphorus for growth of which the two predators must maintain a constant phosphorus to carbon ratio s_1 and s_2 respectively. The available phosphorus amount for the prey is $P - s_1Y_1 - s_2Y_2$. The possible transferred phosphorus ratio from the prey to the second predator is $(P - s_1Y_1 - s_2Y_2)/X$. If it is below the preferred ratio s_2 the second predator becomes less productive. If it is above the preferred ratio, its reproduction depends only on the availability of the carbon. Under these assumptions the dimensional stoichiometry model for Holling Type II consumers becomes

$$\begin{cases} \dfrac{dX}{dt} = rX\left(1 - \dfrac{X}{K}\right) - \dfrac{c_1X}{a_1+X}Y_1 - \dfrac{c_2X}{a_2+X}Y_2 \\ \dfrac{dY_1}{dt} = e_1\dfrac{c_1X}{a_1+X}Y_1 - d_1Y_1 - m_1Y_1^2 \\ \dfrac{dY_2}{dt} = e_2\min\left\{1, \dfrac{(P-s_1Y_1-s_2Y_2)/X}{s_2}\right\}\dfrac{c_2X}{a_2+X}Y_2 - d_2Y_2 - m_2Y_2^2. \end{cases}$$

Here e_1, e_2 are the birth-to-consumption ratios. With a proper change of variables and parameters the model can be cast in the following dimensionless form

$$\begin{cases} \epsilon \dfrac{dx}{dt} = x\left(1 - x - \dfrac{y_1}{\beta_1 + x} - \dfrac{y_2}{\beta_2 + x}\right) \\ \dfrac{dy_1}{dt} = \zeta y_1 \left(\dfrac{x}{\beta_1 + x} - \delta_1 - \mu_1 y_1\right) \\ \dfrac{dy_2}{dt} = y_2 \left(\min\left\{1, \dfrac{(1 - \sigma_1 y_1 - \sigma_2 y_2)}{\kappa x}\right\} \dfrac{x}{\beta_2 + x} - \delta_2 - \mu_2 y_2\right) \end{cases} \tag{1.16}$$

Multitime scales dominate when parameter ϵ and ζ become small.

The attractor on the right of Fig.1.11 is from this model. Again a crashing fold develops on the ϵ-slow manifold, which is the capacity branch of the X-nullcline if one of the predators or both are inefficient. In the case shown it is the second predator. A junction K-fold point is also present on the rebounding path of the fast X-flows. It is the result of the second predator being subject the quality of its food. One can see that to the left of the rebounding junction K-fold point, the high concentration in element X is not favorable for predator Y_2 but indifferent for predator Y_1. Only after the high concentration is drawn down by predator Y_1 to a level does it become beneficial for predator Y_2 and prompt a rebound of its population. Then the crash in the prey ensues above the ridge of the X-nullcline, and the chaotic cycle continues. In any case, chaos is cooked up by a similar recipe.

The recipe is apparently not favored by selection because it asks for ineffective and inefficient species. The multitime scale analysis of the models reveals another unfavorable factor for ecological chaoses. It shows fast transitions are always associated with population collapses and all attractors are invariably tracking along the extinction zones of the crashing populations. It is hard for nature to maintain and sustain such ravages, and it should be improbable for us to observe. In fact, there are no confirmed ecological chaoses in the wild.

Chapter 2
Dynamical Systems

The French mathematician Jules Henri Poincaré (1854-1912) pioneered the dynamical systems approach for mathematical models. It concerns the long time behaviors of solutions of differential equations. The methodology is geometrical, more qualitative than quantitative. Because they last indefinitely, equilibrium solutions, periodic solutions, and chaotic attractors are the primary objects of interest. Of which we are particularly interested in those which attract all nearby solutions. This is because such so-called asymptotically stable steady states are what we suppose to observe for the physical processes the mathematical models describe.

2.1 Hyperbolic Decomposition of Linear Systems

Take the linear system of differential equations as an example, $x' = Ax$. The immediate standout steady state is the trivial equilibrium solution $x = 0$. If all the eigenvalues of the coefficient matrix A have negative real parts, then all solutions converge to the trivial equilibrium point, and the convergence is at an exponential decay rate. The equilibrium point is asymptotically stable. If all the eigenvalues have positive real parts, then every solution except for the equilibrium solution diverges without bound, at an exponentially divergent rate. The equilibrium point is asymptotically unstable. If some eigenvalues have positive real parts and all others have negative real parts, then orbital structure of the phase space is a mixed version of the last two types. We can organize the phase space in the following way and refer to it as a hyperbolic decomposition. In particular, we can find a change of the variable so that in the new coordinate denoted by $x = (x_s, x_u)$ the coefficient matrix is of two

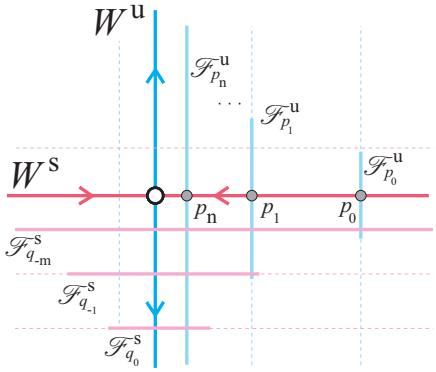


Figure 2.1: Hyperbolic decomposition at an equilibrium point. The solution of the equations maps stable foliations to stable foliations and unstable foliations to unstable foliations.

diagonal blocks $A = \mathrm{diag}(A_s, A_u)$ with the property that all eigenvalues of A_s have negative real parts and all eigenvalues of A_u have positive real parts. Corresponding to this eigenvalue splitting is the hyperbolic decomposition of the phase space $\mathbb{R}^n = E^s \oplus E^u$ with E^s corresponding to the eigenvector space of the matrix A having all negative real parts of its eigenvalues and E^u the eigenvector space having all positive real parts of A's eigenvalues. The E^s subspace is called the stable subspace or stable manifold when generalized to nonlinear systems. Similarly, the E^u subspace is called the unstable subspace or unstable manifold. This description is for good reasons. In fact, every solution from the stable subspace always stays in the subspace and converges to zero exponentially, every solution from the unstable subspace stays in the unstable subspace and diverges without bound. For every initial point not of either types, the full solution $x(t) = (x_s(t), x_u(t))$ always diverges without bound because of having a nonzero component in the unstable subspace. In this sense the trivial equilibrium point is unstable.

The space off the invariant stable and unstable subspaces can be further organized geometrically. First, for every point p from the stable manifold E^s there is a manifold (i.e. hypersurface) of the same dimension as the unstable subspace E^u having these properties. Denote the manifold by $\mathcal{F}^u_p$. Then it forms a family, $\mathcal{F}^u = \{\mathcal{F}^u_p : p \in E^s\}$, referred to as an unstable foliation. Each member of the family is identified by the point the manifold intersects the stable manifold E^s. The most essential property that defines the unstable

foliation is the invariance property: solutions starting from one member of the family stay in another member of the family at every moment of time. That is, $(\mathcal{F}^u_p)_t \subset \mathcal{F}^u_{p_t}$. As a result, the invariant foliation is set-convergent to the unstable manifold forward in time, $\lim_{t\to\infty}(\mathcal{F}^u_p)_t = E^u$. Similarly, there is an invariant stable foliation $\mathcal{F}^s = \{\mathcal{F}^s_q : q \in E^u\}$ so that it is set-convergent to the stable manifold backward in time, $\lim_{t\to-\infty}(\mathcal{F}^s_q)_t = E^s$. Fig.2.1 gives an schematic illustration of this hyperbolic decomposition of the phase space at the equilibrium point.

Generalization of this geometric description about orbits of dynamical systems can proceed in four directions: to nonlinear system of equations, to non-hyperbolic equilibrium points, to maps, to all above but for periodic orbits and chaotic attractors. We will describe these generations in various degree of details. All are about qualitative behaviors of orbits and solutions, and all are modeled after the linear system template above, none is about exactly form of particular solutions. The descriptions are all geometrical and qualitative.

2.2 Hyperbolic Decomposition of Nonlinear Systems

Suppose we have a nonlinear system of equations $x' = f(x)$ and suppose it has an equilibrium point q, i.e. $f(q) = 0$. We can consider at the same time or as the first step of investigation the linear approximation of the system at the equilibrium point, $x' = Df(q)(x - q)$. Here $Df(q)$ is the first derivative of the nonlinear righthand of the full equations. For the first generalization of the linear hyperbolicity we assume the trivial equilibrium for the linearized equations $u' = Df(q)u$ with the translation of variables $u = x - q$ is hyperbolic, i.e. the coefficient matrix has no eigenvalues whose real parts are zeros. With this primary assumption the objects of stable manifold, unstable manifold, stable foliation, unstable foliation, can be all defined and proven to exist with sufficient smoothness.

More specifically, for the stable manifold, its notation and definition are given as

$$W^s(q) = \{x : \lim_{t\to+\infty} x_t = q\},$$

i.e., the set of all initial points x so that their solutions x_t converge to the equilibrium point q. The unstable manifold $W^u(q)$ is defined similarly but with the solutions of its points converging backward to q:

$$W^u(q) = \{x : \lim_{t\to-\infty} x_t = q\}.$$

These manifolds exit, are unique and as many time differentiable as the vector field f itself. They further satisfy these properties. the stable manifold $W^s(q)$ is of the same dimension as the stable subspace E^s of the linearization and is tangent to it at the equilibrium point q; and similarly, the unstable manifold $W^u(q)$ is of the same dimension as the unstable subspace E^u of the linearization and is tangent to it at the equilibrium point.

Similar to the linear case, there exist for the nonlinear system an invariant stable foliation $\mathcal{F}^s = \{\mathcal{F}_p^s\}$ and an invariant unstable foliation $\mathcal{F}^u = \{\mathcal{F}_p^u\}$ satisfying the same invariance properties. The only difference is that the existence of these objects is only local, guaranteed to exist in a small neighborhood of the equilibrium point. The stable manifold is a special member of the stable foliation through the equilibrium point and the likewise the unstable manifold is a special member of the unstable foliation through the equilibrium point. Also, if the unstable spectrum, i.e. the set of eigenvalues of positive real parts, is empty, then full local space near the equilibrium point is the stable manifold. Similarly, if the stable spectrum, defined analogously, is empty, the full local space near the equilibrium point is the unstable manifold. If both spectra are nonempty, then the stable and unstable foliations split is the hyperbolic decomposition of the equilibrium point. The same depiction by Fig.2.1 applies.

The structural decomposition can get finer. Take the unstable manifold for example. Assume the unstable eigenvalues have different negative real parts. Then we call the eigenvalues having the largest negative real part (i.e. closest to the imaginary axis in the complex plane) the principal eigenvalues and all the rest unstable eigenvalues the strong unstable eigenvalues or the non-principal unstable eigenvalues. The linear unstable eigenspace can be further split into the direct sum of the principal unstable eigenvector subspace and the strong unstable eigenvector subspace like this $E^u = E^{pu} \oplus E^{uu}$. For the linear system, solutions from the strong unstable manifold E^{uu} diverge without bound at a greater exponential rate than solutions from the principal unstable manifold E^{ps}. For a perturbed nonlinear system this split mostly persists. More specifically, the strong unstable manifold W^{uu} persists in the sense that it is invariant, tangent to the strong unstable eigenvector subspace E^{uu}, and all orbits from the submanifold are characterized by a divergent rate higher than the principal unstable eigenvalues.

2.3 Nonhyperbolic Decomposition

The generalization to the case for which the linear system $x' = Ax$ or the linearization $x' = Df(q)(x - q)$ of a nonlinear system at an equilibrium point has eigenvalues of zero real parts is both straightforward in form and subtle in structure. For the linear system the phase space can be decomposed into $\mathbb{R}^n = E^s \oplus E^c \oplus E^u$ with E^s, E^u meaning the same and E^c being the eigenvector space of all eigenvalues of zero real parts. The subtle part of the generalization even for the linear system case is that for point from the center subspace or center manifold E^c the solution can converge to the trivial equilibrium point, or diverge away from it, or simply stay put being another equilibrium point, or go around in cycle.

For a nonlinear system of equations at a nonhyperbolic equilibrium point, the same definition and result hold for the stable and unstable manifolds W^s, W^u. In addition, there exists an invariant manifold W^c, referred to as a center manifold of the equilibrium point satisfying these properties. It is of the same dimension as the center subspace E^c of the linearized system at the equilibrium. It is tangent to linear subspace E^c at the equilibrium. It contains all solutions that remain in a small neighborhood of the equilibrium, i.e. any equilibrium point, any periodic solution, any solution that stays in both forward and backward time. Unlike the linear systems, center stable manifolds is not unique for nonlinear systems, but unique up to a change of variables. That means for any two different center manifolds there is a change of variables from one to the other so that any solution from the first manifold is changed to a solution from the second manifold and vice versa.

Jointly, there exist the so-called center-stable manifold W^{cs} and the center-unstable manifold W^{cu}. Each is invariant, i.e. a solution which starts from the manifold stays in the manifold. Each is tangent to the linear subspace $E^{cs} = E^s \oplus E^c$ and $E^{cu} = E^c \oplus E^u$, respectively. And the center-stable manifold contains the stable manifold as a subspace and the center-unstable manifold contains the unstable manifold as a subspace. Moreover, the intersection of both is automatically a center manifold, i.e. $W^c = W^{cs} \cap W^{cu}$.

The stable and unstable foliations also exist, $\mathcal{F}^s = \{\mathcal{F}^s_p : p \in W^{cu}\}$ and $\mathcal{F}^u = \{\mathcal{F}^u_p : p \in W^{cs}\}$. For the former, the membership is defined by the leaf's intersection with the center-stable manifold and for the latter, the intersection with the center-unstable manifold.

2.4 Connecting Orbit

Before we continue to generalize the hyperbolic and nonhyperbolic decompositions for periodic orbits and for maps, let's give a brief description for some interesting results that cannot be described otherwise without featuring the stable and unstable manifolds of equilibrium points.

Equilibrium points, limit cycles are all closed orbits. The next in line of closed orbits in terms of complexity are homoclinic and heteroclinic orbits. A homoclinic orbit of an equilibrium point q lies in the intersection of the point's stable and unstable manifolds. That is, if a point is from the intersection set, $x \in W^s(q) \cap W^u(q) \neq \emptyset$, then the solution x_t must converges to the equilibrium point both in forward and in backward direction, that is,

$$\lim_{t\to+\infty} x_t = q \text{ and } \lim_{t\to-\infty} x_t = q.$$

A heteroclinic orbit is one that lies in the stable manifold of one equilibrium point and the unstable manifold of another equilibrium, i.e. the orbit converges to different equilibrium points in the opposite directions:

$$\lim_{t\to+\infty} x_t = q_1 \text{ and } \lim_{t\to-\infty} x_t = q_2.$$

The existence of such orbits in a system almost always raises the interest level multiple folds. For a homoclinic orbit a small perturbation to the nonlinear system usually creates a periodic orbit. When the conditions are right, they can bring along chaotic orbits, including infinitely many periodic orbits. Heteroclinic cycles can bring out similar complex dynamics as well.

Two-dimensional examples are fairly straightforward. If a Hamiltonian $H(x, y)$ has a saddle critical point which lies on closed level curve $H(x, y) = c$, then the critical point is a saddle equilibrium point of the Hamiltonian system $x' = H_y(x, y),\ y' = -H_x(x, y)$ and the closed level curve is a homoclinic orbit. Also, if we construct a mockup system by Rössler's singular perturbation method for which an equilibrium is on a singular limit cycle, then a homoclinic orbit can be maintained by a perturbed system. Homoclinic and heteroclinic orbits as various intersecting configurations of stable and unstable manifolds are much more interesting in three dimensional systems. Let us take a look at some examples, mostly for their geometric appeals.

2.4.1 Twisted Homoclinic Orbit Doubled

To motivate we first take a look at a mockup system by Rössler's singular perturbation construction. The system of Fig.2.2 has a homoclinic orbit for which

Twist Like A Möbius Band:

$$\begin{cases} \dot{x} = (2-z)a(x-2) + (z+2) \times \big[\alpha(x-\bar{x}) - \beta(y-\bar{y})\big] \\ \dot{y} = (2-z)\big[c(b-a)(x-2)/4 + by\big] + (z+2)\big[\beta(x-\bar{x}) + \alpha(y-\bar{y})\big] \\ \epsilon\dot{z} = (4-z^2)\big[z+2-m(x+2)\big] - dz \end{cases}$$

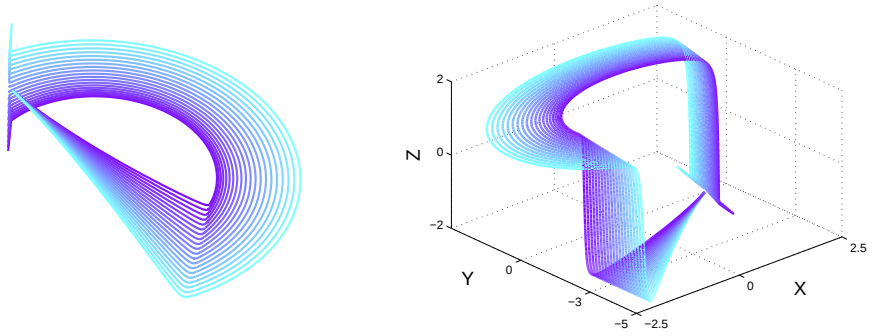


Figure 2.2: A twisted homoclinic orbit in a projected 2-dimensional view and a 3-dimensional view. Parameter values: $a = 1, b = 1.5, c = -3.5, d = 0.02, m = 1.1845, \alpha = 0.01, \beta = -5, \bar{x} = -0.1, \bar{y} = -2, \epsilon = 0.01$.

the unstable manifold of the equilibrium point in question is 2-dimensional and it has a half-twist like a Möbius band around homoclinic orbit before converging to the strong unstable manifold W^{uu}. The same type of z-switch as before is used. On the bottom branch $z = -2$ of the z-switch with $d = 0$, the slow equations are

$$\begin{aligned} \dot{x} &= 4a(x-2) \\ \dot{y} &= 4\big[c(b-a)(x-2)/4 + by\big]. \end{aligned}$$

It is a linear system, and as such everything that needs to be computed can be computed. In particular, the equilibrium point is at $(2, 0)$. The coefficient matrix has two eigenvalues, both are positive, $0 < \lambda_1 < \lambda_2$. The eigenvector for the principal unstable eigenvalue λ_1 is on the line from the equilibrium point to the point $(-2, c)$. The eigenvector for the strong unstable eigenvalue λ_2 is on the line through the point and parallel to the y-axis. The triangle band that fans out from the equilibrium point contains the principal eigenvector solution and a few slow orbits around it. The vantage point of the band for the plane view is from the negative z-axis and with the coordinate rotated 180°. The

band hits the bottom turning point edge, jumps up to meet the top branch of the z-switch. The slow equations on the top branch is a linear spiral around the equilibrium point $(\bar{x}, \bar{y})$. It is positioned so that the band makes a half turn and then hits the top turning edge. It falls down exactly on top of the equilibrium point on the bottom branch of the slow manifold, with the junction points line up along the strong unstable direction of the point. The twisted band contains a homoclinic orbit, and the band is the unstable manifold of the equilibrium point. The fast orbit whose junction point is the equilibrium point is the slow manifold of the equilibrium point. The two manifolds intersect to create the homoclinic orbit. When the unstable manifold goes around with full twists, the homoclinic orbit is nontwisted. It happens for the system when the parameter c changes from -3.5 to -0.2. Somewhere in between the homoclinic orbit hits a K-fold junction point on the top branch of the z-switch. Chaotic dynamics can arise from perturbed systems near such junction-fold homoclinic orbit.

So twisted homoclinic orbit can happen, and it can happen in the simplest setting by singular perturbation. For such an orbit the following theorem holds. Simply put, a twisted homoclinic orbit can double itself if it is to a resonant equilibrium point. Hereon we will always assume the right hand of any differential equations has derivatives of all orders.

Theorem 1 *Assume a system of equations in $\mathbb{R}^3$ has a twisted homoclinic orbit to a hyperbolic equilibrium point. Assume also the principal stable eigenvalue is a simple real number $\lambda^s < 0$ and the principal unstable eigenvalue is a simple real number $\lambda^u > 0$. Assume the resonance condition holds that the magnitude of the principal eigenvalues are equal in magnitude, i.e. $|\lambda^s| = \lambda^u$. Then one can always find a perturbation to the system so that the perturbed system has one homoclinic orbit winding around the tube twice.*

In fact, much more can be concluded than stated in the theorem. A lengthy version of the theorem is to say when the conditions are met, one of the following holds in a tubular neighborhood of the homoclinic orbit: (i) the perturbed system does not have a homoclinic orbit nor a periodic orbit; (ii) the perturbed system has only one periodic orbit; (iii) the perturbed system has one unstable periodic orbit and one stable periodic orbit winding around the tube twice; (iv) the perturbed system has one unstable periodic orbit and one homoclinic orbit winding around the tube twice. And furthermore this result can be generalized to higher than three dimensional spaces.

We will not even attempt an outline of proof for this theorem nor for any theorem stated in this chapter. Doing so will take us too far offtrack. Suffice it

to say all proofs require careful tracking of most if not all the invariant manifolds introduced in the previous sections.

2.4.2 Spiral At Itself

Leonid Pavlovich Shilnikov (1934-2011) was a mathematician from the former Soviet Union. He was generally considered to be the pioneer in the study of homoclinic and heteroclinic orbits of differential equations in higher dimensions. He discovered the first criterion for chaos generation that involves homoclinic orbit to equilibrium point of differential equations, being influenced by and shortly after Smale's discovery of the horseshoe map which we will discuss in the last chapter of this book. His criterion stands to this day to be the simplest kind and the most common mechanism for chaos generation. His work is well regarded in the West but he had a considerable hard time to gain the recognition he deserved at home. Here below is his best known theorem, which says chaos will surely happen if a system has a homoclinic orbit to an equilibrium whose unstable manifold solutions spiral out slower than whose stable manifold solutions close in. Such an equilibrium point is referred to as a saddle-focus equilibrium.

Theorem 2 *(Shilnikov's Saddle-Focus Homoclinic Chaos) Assume a system of equations in* $\mathbb{R}^3$ *has a homoclinic orbit to a saddle-focus equilibrium point at which the principal stable eigenvalue is a simple real number* $\lambda^s < 0$ *and the principal unstable eigenvalue is a pair of complex numbers* $\lambda^u + i\mu,\ \mu > 0$ *so that the contraction is stronger than the expansion at the equilibrium point along the homoclinic orbit, i.e.* $|\lambda^s| > \lambda^u > 0$. *Then in a tubular neighborhood of the orbit and the equilibrium point there is a set of orbits inside the neighborhood in both forward and backward time so that the set contains infinitely many periodic orbits, uncountably many aperiodic orbits, and infinitely many dense orbits of the set. Furthermore, this result can be generalized to higher than three dimensions.*

Actually, if it is not for our consideration for presentation to postpone the introduction of Bernoulli's shift dynamics to the last chapter, the more general way to state his result is to say the solution set is equivalent to a Bernoulli shift on infinitely many symbols.

Fig.2.3 shows a number of Shilnikov's orbits. The top left plot is for the mockup system Eq.(1.8) which was used in the previous chapter to simulate the Rössler attractor. The construction of the Shilnikov's orbit is by design. The

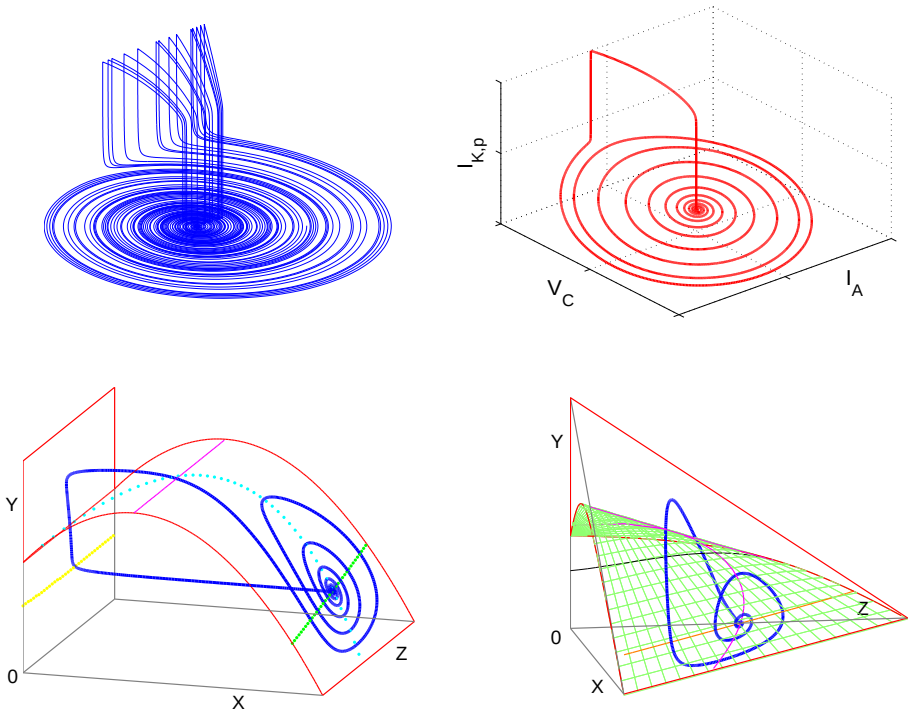


Figure 2.3: Top left: $a = 2, b = -2, c = 1, m = 0.923, r = 2, R = 10^{10}, \alpha = 0.3829, \beta = 5, d = 0.08, \zeta = 4, \epsilon = 0.001$, for Eq.(1.8). Top right: $E_{\text{Na}} = 0.5, E_{\text{K}} = -0.7, r_1 = 10, r_2 = -12.2, i_1 = 0.24, i_2 = 0.43, G_1 = 15, G_2 = -0.37, V_1 = -1, V_2 = -0.02, \lambda = 0.055, C = 2.2, R = 0.18, \epsilon = 0.001$, and zeroes for other parameters, for Eq.(1.1). Bottom left: $\beta_1 = 0.26, \beta_2 = 0.5, \delta_1 = 0.2, \delta_2 = 0.19, \mu_1 = 0, \mu_2 = 0.1, \zeta = 1, \varepsilon = 0.02$, for Eq.(1.4). Bottom right: $a_1 = 0.08, a_2 = 0.23, a_3 = 0.185, m_1 = 15, m_2 = 7, m_3 = 2.5$, and $b = b_1 = b_2 = 1$, for Eq.(1.15).

top right plot is for the neuronal circuit model Eq.(1.1) for which the type of Shilnikov's chaos occurs when the membrane capacitance becomes large. The bottom left plot is for the dimensionless food chain model Eq.(1.4) for which chaos arises when the predators are not proficient in catching their preys with small half saturation parameters β_1 and β_2. The bottom right plot is for the chemostat model Eq.(1.15). Again, chaos can only occur when the consumers are not proficient with small a_1 and a_2. The last three model attractors would

spread out to fill more space like the first attractor had we followed up the solutions over a longer period of time like we did for the first one. The beauty of the theorem lies in the simplicity the condition entails. All one has to establish for a system of differential equations to be chaotic is to prove the existence of a Shilnikov's saddle-focus homoclinic orbit. For fast-slow systems this strategy has proved to be extremely effective.

2.4.3 One Twisted Loop And Infinite Connections

Next, let us take a look at a specific model with a very interesting feature around a pair of heteroclinic orbits to two equilibrium points that forms a loop if one follows the time evolution of the orbits. That is, the model has one heteroclinic orbit from equilibrium one to equilibrium two and another heteroclinic orbit from equilibrium two back to equilibrium one. The interesting feature is the center-stable manifolds of the equilibria that are twisted respect to the connecting loop.

The model system in question is the FitzHuge-Nagumo reaction-diffusion equations

$$v_t = v_{xx} + f(v) - w, \;\; w_t = \epsilon(v - \gamma w), \tag{2.1}$$

with the function f being this cubic polynomial $f(v) = v(1-v)(a-v)$. The system is used as a model for electrical impulse transmission along an infinitely long cable. It is often cited in the literature as a model for nerve axon, which is more of a tenuous application. Nevertheless in the setting of the nerve axon, the spatial variable represents the axial dimension of the axon. The variable w with ϵ small is often referred to as the recovery variable which from our modeling of the neurons corresponds to the ion pump dynamics. The function f is used to model the simplest type of nonlinearity for cross-membrane passive currets in parallel channels.

Nerve impulses are thought to maintain their spatial profiles as they travel along the axon, and the model is used to capture this so-called traveling wave phenomenon. In terms of solutions to the reaction-diffusion equations, they are represented by the traveling wave solutions of the form

$$(v, w)(x, t) = (v, w)(x + ct),$$

where $c > 0$ is the wave propagation speed. Casting the equations in the traveling coordinate frame $z = x+ct$ transforms the reaction-diffusion equations Eq.(2.1) into a system of ordinary differential equations as follows

$$c\dot{v} = \ddot{v} + f(v) - w \text{ and } c\dot{w} = \epsilon(v - \gamma w),$$

where the dot represents taking derivative in the traveling wave coordinate z, i.e. $\dot{} = \frac{d}{dz}$. By introducing the velocity variable of the v-variable, $u = \dot{v}$, the higher order system of equations is transformed into a first order system:

$$\dot{u} = cu - f(v) + w, \quad \dot{v} = u, \quad \dot{w} = \epsilon(v - \gamma w)/c \tag{2.2}$$

Since ϵ is a small variable, representing the slow pump dynamics, this system acquires the attractive feature of being multiple time scaled. As a fast-slow system, variables u, v are the fast variables and w the slow variable.

The slow manifold consists of the equilibrium points of the fast subsystem: $u = 0, w = f(v)$, i.e., the cubic curve on the vw-plane. It shapes like the letter N upsidedown in the vw-plane view. Fig.2.4 shows only the left and right branches of the upsidedown N. Depending on the size of γ, there are one, two, or three equilibrium solutions. Fig.2.4 depicts an interesting case of having three equilibrium points, the origin 0, the equilibrium point p, respectively on the left and right branches of the N slow manifold. The figure depicts a parameter case in which the unstable manifolds of both 0 and p land precisely on the slow manifold at the singular value $\epsilon = 0$. The unstable manifold of 0 lies on the fast uv-plane with $w = 0$, the unstable manifold of p lies on a plane parallel with the uv-plane as $\dot{w} = 0$ at the singular limit $\epsilon = 0$. The slow dynamics moves down above 0 on the slow manifold's left branch because the w-rate of change is negative, $w' = (v - \gamma w)/c < 0$. Below p on the right branch the derivative is positive, $w' > 0$, thus the slow dynamics moves up. As a result, the fast and slow flow concatenations result in a singular heteroclinic orbit from 0 to p and a symmetric one from p to 0.

The figure gives a qualitatively correct depiction of the invariant manifolds at the singular value $\epsilon = 0$. In particular, the center manifold of 0 is exactly the left-branch of the slow manifold, and similarly for the center-manifold W^c_p of p. The most interesting feature is the fact that the singular heteroclinic orbit from 0 to p connects opposite sides of the center-stable manifold of 0. Similarly, the heteroclinic orbit from p to 0 arises from the back side of the center-stable manifold W^{cs}_p of p and ends on the front side of it, which lines up along the strong stable manifold of 0 when its orbits flow backward in time.

Consequences to this twisted heteroclinic pair are very interesting. It is a proved fact that for each sufficiently small ϵ and each γ near the parameter point (c^*, γ^*) for the singular heteroclinic cycle, there are two decreasing sequences of c, $c_{0,n}$ and $c_{p,n}$, near c^* so that the reaction-diffusion equations have a heteroclinic orbit from 0 to p around the loop n times at the wave speed $c_{0,n}$ and another heteroclinic orbit from p to 0 around the loop n times at the wave

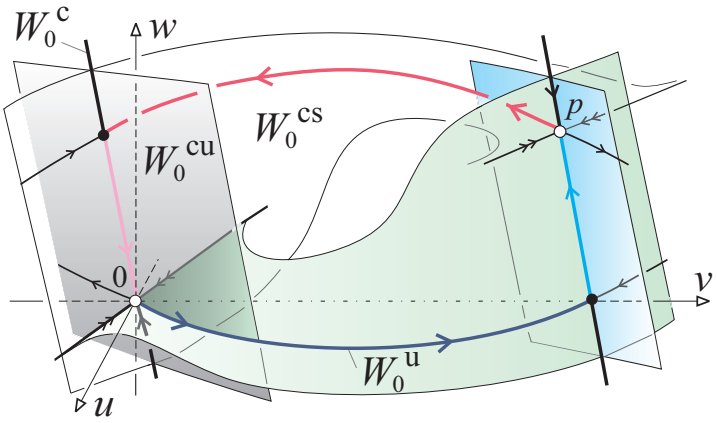


Figure 2.4: A twisted embrace of the center-stable manifolds of 0 and p.

speed $c_{p,n}$. Furthermore, the equations have a homoclinic orbit to 0 around the loop once at the wave speed $c_{0,\infty}$ which is the limit of $c_{0,n}$, and symmetrically another homoclinic orbit to p around the loop once at the wave speed $c_{p,\infty} = \lim_{n\to\infty} c_{p,n}$. In addition, for all c sufficiently near but below $c_{0,\infty}$ or $c_{p,\infty}$ the equations have a periodic orbit around the loop once.

The parameter point (c^*, γ^*) from which all the n-loop heteroclinic orbits arise can be explicitly computed. Let's file it away in the statement below.

Theorem 3 *At $\epsilon = 0$ and for $0 < a < 1/2$, the parameter point (c^*, γ^*) is given by $c^* = \frac{1-2a}{\sqrt{2}}$, $\gamma^* = \frac{9}{(1-2a)(2-a)}$. Also the corresponding equilibrium point $p = (0, v^*, w^*)$ is given by $v^* = \frac{2}{3}(1+a)$, $w^* = f(v^*)$.*

Although these orbits are close to each others as solutions to the ordinary differential equations (2.2), as solutions to the partial differential equations they are far apart in the spatial profile dimension. The question for the PDE, which is an infinitely dimensional system, is are these infinitely many traveling waves stable? The last, but not the least, interesting property is the fact that they are indeed all asymptotically stable, and the stability is determined by the way the center-unstable manifold of 0 (resp. p) and the center-stable manifold of p (resp. 0) intersect at an acute angle in the direction of w on a plane perpendicular to the fast unstable orbit W_0^u (resp. W_p^u). Without the twist all the infinitely many heteroclinic pulse waves are not possible.

Double Shilnikov Attractor:

$$\begin{cases} \dot{x} = z(z+r)\zeta - z(z-r)\zeta + (r^2 - z^2)[\alpha x - \beta y \\ \dot{y} = z(z+r)[b + \frac{c-b}{-3-a}(x-a) - y] \\ \qquad\qquad +z(z-r)[-b + \frac{-c+b}{3+a}(x+a) - y] + (r^2 - z^2)[\beta x + \alpha y] \\ \epsilon\dot{z} = z(r^2 - z^2)\big[z - m(x+3)\big]\big[z - m(x-3)\big] - \epsilon d(z+x) \end{cases} \tag{2.3}$$

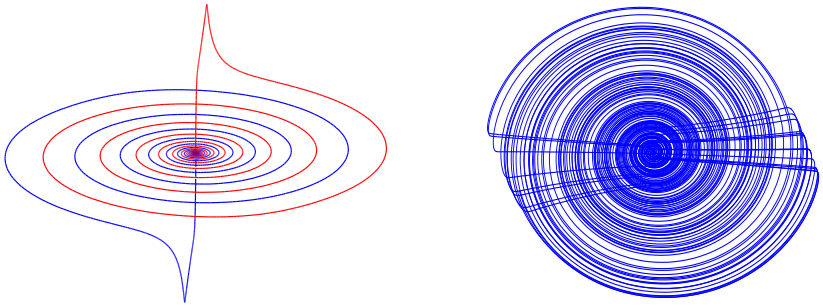


Figure 2.5: Parameter values: $a = 0, b = 0, c = 0, d = 1, m = 0.8993, r = 2, \beta = 5, \zeta = 47, \epsilon = 0.1$, and $\alpha = \frac{\beta \ln(1+\Phi)}{2\pi} = 0.3829$. Left: the up and the down homoclinic orbit. Right: the attractor.

2.4.4 Up Or Down: Random At The Core

Up to now we have relied on eye witness account to slap a chaos label on anything that looks complicated. We still have to wait until the last chapter to give one precise definition of chaos. In this example, however we can be more convincing about the chaos than about all other chaoses, even though we still rely on pictures. It is to apply one working definition of chaos to make the determination. It is the property of sensitive dependence on initial conditions Lorenz had emphasized with his butterfly attractor discovery. Instead of theorizing it, let us present another mockup system whose unstable manifold of an equilibrium is fantastically pulled, twisted, torn in two different directions.

Fig.2.5 shows the equations and the attractors. Instead of a single z-switch the switch slow manifold consists of three horizontal branches $z = -2, z = 0, z = 2$ for bottom, middle, top branches, respectively, and two slanted planes. Two adjacent branches are part of a z-switch. Together they form a double

z-switch, one Z at the top and one Z at the bottom, both share the middle horizontal plane. The slow orbits on the top and bottom branches are linear flows, both pulling toward their respective turning edge. An unstable spiral equilibrium point is placed at the center $(0, 0, 0)$ in the middle branch of the double z-switch of the slow manifold, so that it is the junction target of the two turning edges. There are two K-fold turning points in the middle branch, one switches the slow flows up to the top branch of the double z-switch and the other switches the slow flows down to the bottom branch. So depending on where a point is located in the middle branch, from one region the orbit is sent to the top and from another region the orbit is sent to the bottom. And the two types of regions are tightly tangled up at the equilibrium point.

By turning the parameter values as nobs one can simultaneously construct two homoclinic orbits to the center equilibrium point. Each orbit is a Shilnikov's saddle-focus homoclinic orbit. Because of the double z-switch, we get two. In this case, both shares the common equilibrium point. In terms of the unstable manifold of the spiral center $(0, 0, 0)$, one piece has to be pulled up to meet the top branch of its stable manifold and another piece has to be pulled down to meet the bottom branch of the stable manifold. While doing that they have to rotate infinitely many times near the unstable manifold. Like inside a shredder, infinitely many pieces are torn away to be sent up and infinitely many pieces are torn away to be sent down, again and again, and so on. Yet, inside this double chaotic funnels things must move on orderly – these infinitely many threaded pieces do not bump on each others because if they do they would create additional homoclinic orbits but the only orbits that are allowed to hit the bullseye are the two homoclinic orbits.

The sensitive dependence on initial conditions is never so obvious. No mater how close two points are near the equilibrium point, one can be sent out spiraling to go up and the other can be sent out spiraling to fall down. Any random perturbation or fluctuation to an initial point near the core can send the orbit onto a completely different path. This random switching goes on indefinitely at the center. Without any formal verification of any version of chaos definition, this double whammy Shilnikov casts much lesser doubt on its chaosness.

2.5 Decompositions For Map

The idea of invariant manifolds decomposition near equilibrium points can be easily extended to periodic orbits. One can do this by generalizing the various definitions. But Poincaré had a better idea. Instead of working with the solution

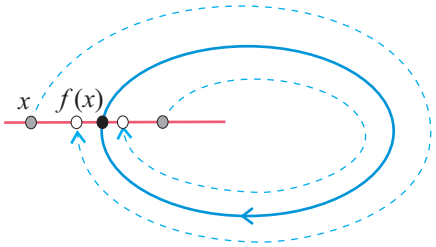


Figure 2.6: Return map on a cross section to a periodic orbit.

operator x_t of the equations with initial point x and the continuous time t, Poincaré introduced the idea to work with the first return map defined on a cross section of the periodic orbit, see Fig.2.6 for an illustration. Denote the return map by $f : S \to S$ with S being a cross section of the phase space. Then, the periodic orbit becomes a fixed point of f. It is always locally defined, and in many cases it can be defined globally as shown in Fig.2.6 for some examples.

Poincaré pioneered the modern version of hyperbolic decomposition of maps which was extended for equilibrium points of differential equations. Because a map can be thought to be an iterative, discrete process, we can use the natural number to denote the discrete time progression. So we denote x_n the current state of a process and $x_{n+1} = f(x_n)$ be the next state. If f is a linear map, then we can write $x_{n+1} = Ax_n$ after relocating any fixed point to the origin. The change of coordinate, the hyperbolic decomposition of the phase space, $\mathbb{R}^d = E^s \oplus E^c \oplus E^u$, are also done by eigenspace of the matrix A. The only difference between flows and maps with regard to the eigenvalues is that the unstable eigenvalues of E^u for maps are those lying outside the unit circle in the complex plane, and those of E^s inside the unit circle, and those of E^c on the unit circle. If $x \in E^u$, then the iterative orbit $x_n = f(x_{n-1}) = A^n x_0$ with $x_0 = x$ must diverge without bound at a geometrical rate. If $x \in E^s$, the orbit $x_n = A^n x$ must converge to 0 geometrically. For $x \in E^c$, the orbit x_n can behave in many different ways. It may remain bounded, or converge to the fixed point, or diverge without bound, but in the latter two cases they cannot do so at geometric rates, forms like λ^n with either $|\lambda| < 1$ or $|\lambda| > 1$, the discrete equivalence to exponential rate for flows when n is replaced by a continuous time.

For nonlinear maps with fixed points, the generalization for the various manifolds and foliations proceed in the ways as for differential equations, using the linear subspace decomposition $E^s \oplus E^c \oplus E^u$ as various tangent space combi-

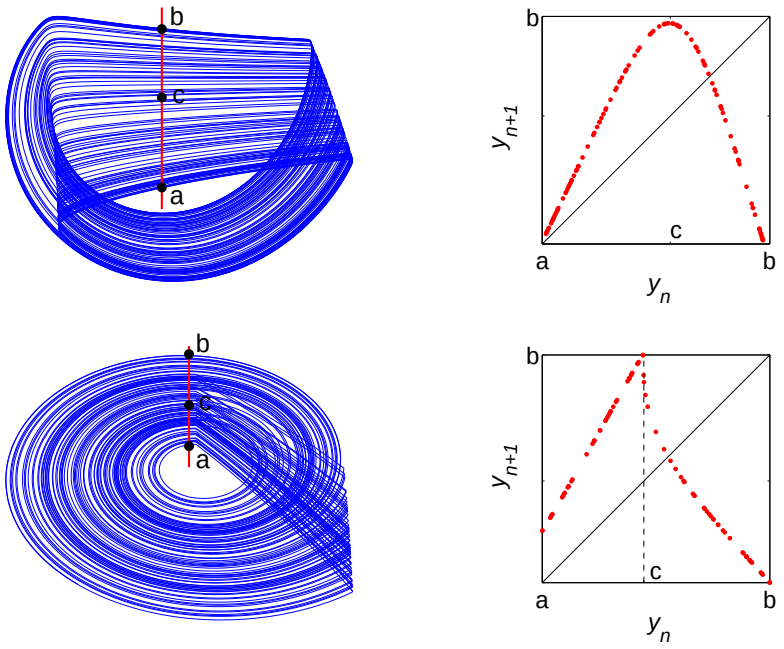


Figure 2.7: K-fold attractors and their return maps. Top: The same junction K-fold attractor as Fig.1.4. The return map is a unimodal map in the interval [a,b] with c being the maximal point. Bottom: Turning K-fold attractor of Eq.(1.8) with parameters: $a = 2, b = -10, c = -8.5, m = 0.92, r = 2, R = 10^{10}, \alpha = 0.5, \beta = 5, d = 0.08, \zeta = 8, \epsilon = 0.001$

nations at the fixed point. The same picture (2.1) can be used as an geometric illustration of the local hyperbolic decomposition for both flows and maps, one at equilibrium points and the other at an fixed points. If we use W^u_{loc} to denote the local unstable manifold that contains the fixed point and is tangent to the linear subspace E^u at the fixed point, then the global unstable manifold W^u is also the dynamic spread of the local one, i.e. $W^u = \cup_{n=0}^{\infty} f^n(W^u_{loc})$ for map and $W^u = \cup_{0\le t<\infty}^{\infty} \phi_t(W^u_{loc})$ for flow. Similarly, we have $W^s = \cup_{n=0}^{\infty} f^{-n}(W^u_{loc})$, assuming f is a diffeomorphism.

2.5.1 Tangled Chaos

Homoclinic and heteroclinic orbits are defined similarly as for the flows to be the intersections of the global stable and unstable manifolds of the same or different

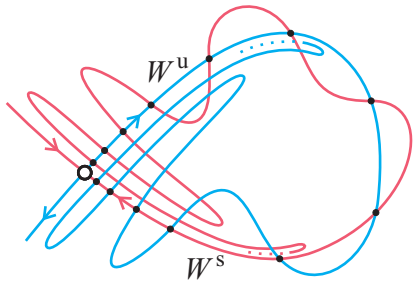


Figure 2.8: The stable and unstable manifolds of a map can intersect transversally as well as tangentially, forming an extremely complex tangle.

fixed points. Unlike differential equations in 2-dimension, two-dimensional maps can be much more interesting and complex. This is perhaps not a fair comparison because a 2-dimensional map is inherently a 3-dimensional flow on a cross section. Nonetheless, when the stable and unstable manifolds of a fixed point of a map intersect, the typically do so with the manifolds go through each other at an angle, referred to as transversal intersection. If x_0 is such a homoclinic point, then both forward and backward iterates x_n converge to the fixed point as the discrete time goes to infinity, $n \to \pm\infty$. Since a piece of the unstable manifold goes through the homoclinic point x_0, that piece is replicated infinitely many time along the forward orbit $x_n, n > 1$, piling up closer and closer to the local unstable manifold W^u_{loc} first and then being stretched and spread along the global unstable manifold W^u. The same picture is true for a piece of the stable manifold through x_0 that is carried backward in time. This creates an orderly but tangled mess that even befuddled Poincaré to the point he had to leave a three-body problem unsettled which earned him his first official accolade not because he did not solve the problem completely but because he had made so much progress on an extremely hard problem. With the Smale horseshoe and chaos to be defined precisely in the last chapter later we can state the following simplest chaos scenario for maps.

Theorem 4 *If a differentiable map has a transversal homoclinic point then the map has an invariant set near the point that contains infinitely many periodic orbits, infinitely many homoclinic orbits of periodic orbits, infinitely many heteroclinic orbits of periodic orbits, uncountably many non-periodic orbits, infinitely many dense orbits, and all things of chaos.*

In its flow equivalent setting, the fixed point of the return map corresponds to a periodic orbit of the flow, the transversal homoclinic point of the map corresponds to a transversal intersection of the stable and unstable manifolds of the periodic orbit. Thus, this theorem gives a simple way, of course, simple in a relative sense, that a system of differential equations in three dimension or higher can become chaotic.

Back to the map, the transversal intersection of the stable and the unstable manifolds are not necessary. Actually there is nothing in theory to prevent them from intersecting tangentially. When they do, there will be infinitely many such tangential homoclinic points referred to as sinks. One can show that you cannot destroy one tangency without creating another somewhere else. We will discuss this later in a different setting.

2.5.2 Markov Chain: A Deterministic Random Process

Let us close this chapter with a nonhyperbolic map which is used to model a lot of discrete stochastic processes. Such processes can be cast in the following setting. First there are a finite number of states, state 1, state 2, and etc. all up to the last state n. Second the process at a particularly discrete time, called it the stage, is at a state. This process's state at stage t is denoted by X_t and it is equal to state j, the process's state at the stage. Hence we can denote it by $X_t = j$. Third the process is not deterministic but probabilistic. That is, the state at the next stage is not determined by a function but by a probability. It can be any number of the states. Finally if we assume the probability to determine the transition over one stage from a state j to a state i is independent of other states nor any prior stages, the resulting probability distribution, denoted by p_{ij}, completes the probabilistic model for such a process. A process satisfying these conditions is called a Markov chain, the changing state X_t is referred to as the random variable taking values in the states and the probability p_{ij} from state j to state i is referred to as the transition probability. In notation, we have, $p_{ij} = \Pr\{X_{t+1} = i | X_t = j\}$, which reads for the righthand side that the (conditional) probability of the process at state i given the process at state j at the previous stage.

Because the state j must transit to one of the n states, the possibility to itself also included if that probability is not zero, the sum over the destination states must be 1 for each given state j, that is $\sum_{i=1}^{n} p_{ij} = 1$. List them all in an array we obtain the so-called transition matrix $P = [p_{ij}]$, of which each column is the transition probabilities from the column j state to other states, which are

the row numbers. If we stop motivating further about such Markov processes, the transition matrix formally has already defined a linear map dynamical system $x_{k+i} = Px_k$. What we like to illustrate below is the fact all pertinent questions about the Markov process have something to do with the matrix as a dynamical system. The most interesting feature of all is the property that the invariant spectral space split of the map at the trivial fixed point $x = 0$ is always nonhyperbolic. In fact, the first natural number 1 is always an eigenvalue of the matrix and the center stable manifold W_c is exactly the linear eigenspace E^c of the eigenvalue and it is exactly one dimensional. That is, there are no other center eigenvalues for the map. Furthermore, all other eigenvalues are stable eigenvalues, strictly inside the unit circles. That is, the stable manifold W^s is the linear spectral space E^s, filling up the rest complementary space for the phase space $\mathbb{R}^n$. These results and more are based on one hard theorem, referred to as the Perron-Frobenius Theorem, which we will state shortly and then illustrate how the theorem leads to the nonhyperbolic decomposition of the map alluded here above.

Back to the Markov process. The transition probability matrix is only the means not the questions about the process. The first question one asks usually starts with an initial probability distribution about the process. Say, at the initial stage the process is likely to be found at a given state i with probability $x^{(i)}$. Here $x^{(i)}$ is the ith row component of the probability vector $x = [x^{(1)}, x^{(2)}, \ldots, x^{(n)}]'$ with $'$ denoting the transpose operation from a row to a column vector and from a column to a row vector. This is the so-called marginal probability for the random variable X_t. Use the transition probability as the conditional probability, then we can find the marginal probability for the random variable X_{t+1} at the next stage. Denote the probability by $x_1 = [x_1^{(1)}, x_1^{(2)}, \ldots, x_1^{(n)}]'$. Then we have by Bayes' rule, $x_1^{(i)} = \sum_{j=1}^n p_{ij} x^{(j)}$. That is, the probability to find the process at state i is distributed over all possible paths into state i: from state 1 to state i, from state 2 to state i, and so on. The application of Baye's formula gives the first usage of the transition matrix as one can check trivially $x_1 = Px_0$ with $x_0 = x$. Use x_1 as a new initial state's marginal distribution and apply the same computation as above, we then have $x_2 = Px_1 = P^2x$ for the marginal distribution for the process at the new stage 1, i.e. the original stage 2. In general, carry this iteration forward, we have $x_k = Px_{k-1} = P^kx$ for the stage k's marginal distribution x_k for the process states. If we let $p_{ij}^{(k)}$ denote the components of the kth iterate matrix P^k, then it represents exactly the (conditional) transitional probability of being at state i given that the process is at state j but k many stages prior,

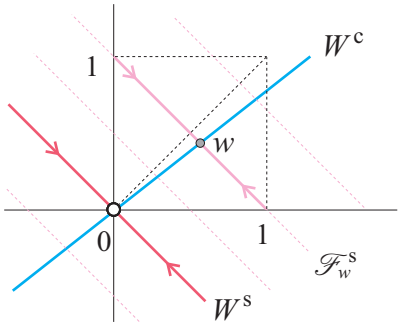


Figure 2.9: The dynamics of the transition matrix of a Markov process at the trivial fixed point and the steady state distribution.

that is $p_{ij}^{(k)} = \Pr\{X_{t+k} = i | X_t = j\}$ with $p_{ij}^{(1)} = p_{ij}$.

The next question obviously is to ask what happens in a long run, or if the asymptotic marginal distribution $x_\infty = \lim_{k\to\infty} x_k$ exists? If it does, does the limit depend on the initial distribution x? If not, the interpretation for the limiting distribution x_∞ is this steady state interpretation for the Markov process. Which means no matter what initial marginal distribution for the process states, in a long run the probability to find the process at state i is $x_\infty^{(i)}$. If such a steady state exists, we use w to denote the asymptotic limit, $w = x_\infty$, for the property that it is independent of the initial state distribution x. The vector w is a column vector, and a probability vector with $w_i \geq 0$ and $\sum_{i=1}^n w_i = 1$. This leads to the following statement.

Theorem 5 *(Perron-Frobenius Theorem) If the transition matrix P satisfies this transitive property that after a finite iteration k, all the entries of P^k is positive, i.e. every state is probable to become any other state after k many stages, then the limit $\lim_{k\to\infty} P^k$ exists and the limit is consisting of a unique steady state distribution probability vector w repeated for every column of the limit, i.e. $\lim_{k\to\infty} P^k = [w, w, \ldots, w]$. In components, $\lim_{k\to\infty} p^{(k)}ij = w_i$ for all i and all j.*

This result immediately implies that as a discrete dynamical system the matrix P has the first natural number 1 as an eigenvalue and the steady state distribution vector w is an eigenvector. In fact, we have this rewriting of the limit,

$$P[w, w, \ldots, w] = P \lim_{k\to\infty} P^k = \lim_{k\to\infty} P^{k+1} = [w, w, \ldots, w]$$

implying $Pw = w$, when equating the left and the right one column a time. Since each column of P^k sums to 1, so is the limit and the w vector is a probability vector and as such it is a nonzero vector. So $\lambda = 1$ is an eigenvalue and w is its eigenvector, and the center manifold W^c is at least of the linear span by w. That is, W^c must contain the line through the origin and the point w in the phase space R^n, see Fig.2.9. In fact, a few lines of simple manipulation will show the eigenvalue 1's spectral line is all of the center stable manifold W^c.

Here is how. As far as the eigenvalue 1 goes, we don't need the big theorem of Perron-Frobenius. We can check easily that if we denote $\mathbf{1}$ for the column vector consisting of 1s for all its entries, then it is an eigenvector for the transpose of the transition matrix, i.e. $P'\mathbf{1} = \mathbf{1}$. The calculation is trivial because $\sum_{i=1}^n p_{ij} = 1$ for all js. State it differently, the row vector $\mathbf{1}'$ is a left eigenvector of the transition matrix as $\mathbf{1}'P = \mathbf{1}'$ whose eigenvalue is 1. So $\lambda = 1$ is an eigenvalue of the matrix.

The dynamics on the center manifold W^c is trivial. It consists of all fixed points of the matrix as $P(rw) = rw$ for all coefficients r. Next, we find the stable manifold of the matrix at the trivial fixed point $x = 0$ or at every fixed point on the center stable manifold W^c. Surprisingly it is the hyperplane through the origin that is perpendicular to the transpose's eigenvector $\mathbf{1}$. That is, for every x from W^s, the dot product with the normal vector $\mathbf{1}$ is zero for orthogonality, $x \cdot \mathbf{1} = 0$, or $W^s = \{x = [x_1, x_2, \ldots, x_n]' : \sum_{i=1}^n x_i = 0\}$. The line of argument is almost trivial because for each x from the set W^s, its iterate converges to zero: $\lim_{k\to\infty} P^k x = [w, w, \ldots, w]x = (\sum_{i=1}^n x_i)w = 0$. Because P is a matrix, the convergence must be geometrical (i.e. exponential) and x must be from the eigenspace of a stable eigenvalue. Because the dimension of W^s is $n-1$-dimensional, there must be $n-1$ many stable eigenvalues and therefore 1 is the only center eigenvalue of the matrix.

The spectral decomposition of the phase space of the matrix at the trivial fixed point is exactly depicted by Fig.2.9. Notice that every plane parallel to the stable manifold W^s is a leaf of the stable foliation. The one through the steady state distribution point w is the one containing the probability simplex, $\sum x_i = 1, x_i \geq 0$, pertinent to the Markov process. Herein we can recover the asymptotic steady state property of the process satisfying the transitive condition of the Perron-Frobenius Theorem. For any initial marginal distribution x, the long term marginal distribution is the constant steady state distribution: $\lim_{k\to\infty} P^k x = [w, w, \ldots, w]x = (\sum_{i=1}^n x^{(i)})w = w$. If x is from any other stable foliation leaf, the limit converges to the fixed point $(\sum_{i=1}^n x^{(i)})w$ on the center manifold W^c.

Chapter 3

Universal Number 1

In the 1970s the American physicist Mitchell Feigenbaum and, independently, two French physicists Pierre Coullet and Charles Tresser made a fantastic discovery on the simplest possible nonlinear map, the logistic map $f_r(x) = rx(1-x)$. Their discovery is now referred to as the Feigenbaum-Coullet-Tresser Universality because their result not only applies to the logistic map but also to all dynamical systems which undergo the cascade of period-doubling bifurcations to chaos.

3.1 Period-Doubling Universality

The logistic map maps the unit interval $[0, 1]$ into itself for $0 \leq r \leq 4$. The origin $x = 0$ is always a fixed point $\bar{x} = f_r(\bar{x})$ for all r. It is globally attracting only for $0 < r < r_0 = 1$ but for r above r_0, it becomes unstable. For $r > r_0$ a nontrivial fixed point is born inside $(0, 1)$. Between $r_0 < r < r_1 = 3$ the new fixed point is stable and then becomes unstable for $r > r_1$. The parameter value r_1 starts the first period-doubling bifurcation point in the cascade: for $r > r_1$ the map has a period-two periodic orbit which attracts all orbits except for the two unstable fixed points. This attraction holds for parameter values below the next bifurcation point $r_2 = 3.449490$ in the cascade. From r_2 onward the period-two orbit losses its stability and a period-four orbit is born for all $r > r_2$. It stays stable only below another parameter point $r_3 > r_2$ and losses its stability beyond r_3 from which a period-2^3 is born, and so on. That is, the cascade consists of an increasing sequence of period-doubling bifurcation points, $r_1 < r_2 < \cdots < r_n < \cdots$, so that for $r_n < r < r_{n+1}$ the map has a stable period-2^n orbit which losses its stability as r across r_{n+1} from below at

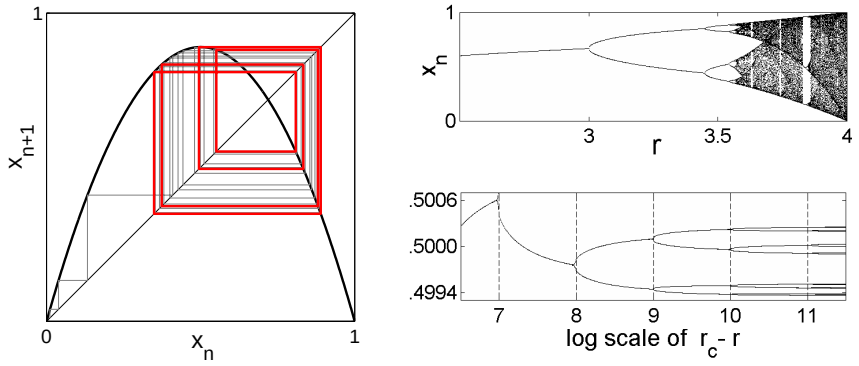


Figure 3.1: Left: the so-called cobweb plot of orbit for 1-dimensional maps. Let $\{x_0, x_1, x_2, \dots\}$ with $x_{n+1} = f(x_n)$ for a 1-d map f in a normalized interval $[0, 1]$ so that $0 \le x_n \le 1$. One starts at point (x_0, x_0), connects it to (x_0, x_1) by a line, draws another line to (x_1, x_1) on the diagonal, and then repeats the process to the last point in a finite sequence of the orbit from the initial point x_0. The plot shows an orbit of the logistic map at parameter value $r = 3.56$ that is converging to the attracting period-8 orbit (red). Right top: the bifurcation diagram for $2 < r < 4$, showing the attractors, chaotic or periodic, in the unit interval for a discretized sequence of the parameter. Bottom: a room-in window for the bifurcation diagram, not against the parameter $r \sim r_c - \alpha\delta^{-n}$ but against this logarithmic transformed parameter, $n = -[\log \frac{r_c - r}{\alpha}]/\log \delta$ with $\alpha = 2.5029$. It shows the period-2^n bifurcation points approximately fall on the integers $n = 7, 8, 9, 10, 11$, numerically demonstrating the convergence of r_n to r_c is exponential.

which a stable period-2^{n+1} is born, see Fig.3.1.

3.1.1 The First Universal Number

This sequence has a limiting value $r_c = 3.569945672...$ at which the map has an unstable periodic orbit of period 2^n for all nonnegative integers $n = 0, 1, 2, \dots,$ the onset of chaos for the logistic family. These three physicists also found that the convergence of the cascade to the limit point is exponential: $r_c - r_n \sim \delta^{-n}$ with

$$\delta = 4.6692016091029....$$

They also observed that for different families of maps, such as $f_r(x) = r\sin(\pi x)$, the period-doubling cascade may take on different bifurcation values r_n and different limit r_c but the exponential convergence rate $r_c - r_n \sim \delta^{-n}$ remains the same. In an equivalent fashion for exponential decay, the quotient sequence of adjacent differences satisfies

$$\lim_{n\to\infty} \frac{r_n - r_{n-1}}{r_{n+1} - r_n} = \delta.$$

That is, the number δ is universal for the class of dynamical systems undergoing the period-doubling bifurcation cascade, independent of any particular family in the class. For this reason the number δ is referred to as the Feigenbaum-Coullet-Tresser universal constant.

3.1.2 Renormalization: Return Map Of Map

What they did next was truly fascinating — they constructed another dynamical system on top of the dynamical systems of all logistic-like maps to explain their universal constant δ. The new dynamical system is referred to as the renormalization map R. It is not a 1-dimensional mapping or a n-dimensional mapping, but a mapping on a set of all logistic-like mappings on the unit interval. The set, denoted by $\mathcal{U}$, is called the renormalizable space, consisting of mappings like the logistic map. The operation R does nothing but essentially find a return map of any logistic-like map from $\mathcal{U}$! See Fig.3.2 for an illustration.

Here is how R and $\mathcal{U}$ are defined. Notice that a point f from $\mathcal{U}$ is actually a 1-dimensional mapping of the unit interval into itself. The first requirement of f is to have two subintervals I_0, I_1, not sharing any interior point so that f maps I_0 into I_1 and I_1 into I_0: $I_0 \xrightarrow{f} I_1 \xrightarrow{f} I_0$. This combinatoric condition is referred to as renormalizable. Because of this recurring relationship, the return map of f in I_0, $f^2 : I_0 \to I_0$, is well-defined: the first application of f takes I_0 to I_1 and the second takes I_1 back to I_0 to complete the return. A mapping having a unique maximum in the interior of the unit interval and no interior minimums is referred to as unimodal. The last requirement for f in $\mathcal{U}$ is for it to be unimodal. Once these two conditions are met, the renormalization $R(f)$ of f needs at most two more steps to complete. The first is a must: scale the domain-image interval I_0 of the return map f^2 to the unit interval $[0, 1]$.

Normally this completes the definition of $R(f)$. For the case these researchers were interested in, the unimodality was essential. In fact, they were dealing with the case that f is unimodal in I_0 but monotone decreasing in I_1. As a result the return map f^2 on I_0 has a unique interior minimum and no interior maximums

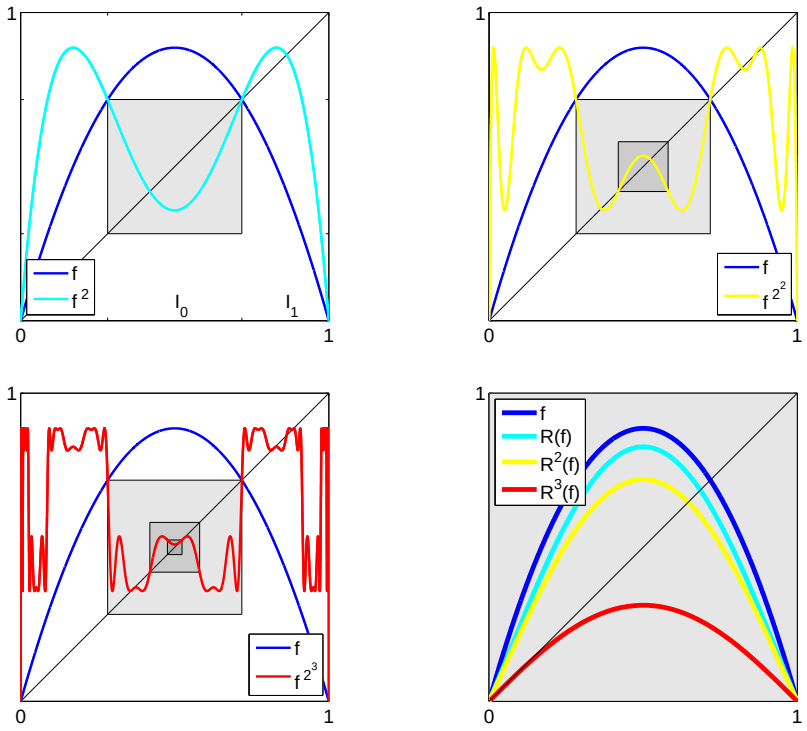


Figure 3.2: Top left: the restriction of the double iterate f^2 to the middle interval in shade is the Poincaré return map of f in the subinterval. Flipping the shaded box left right, top side down and then stretching everything in it, especially the graph of f^2, to fill the unit box complete the definition of $R(f)$ (bottom right). The two flipping operations are the same in result as the rotation of the center box by 180°. In the case shown $R(f)$ is again renormalizable and $R(R(f))$ is essentially defined by the return map of f, i.e. $f^4 = (f^2)^2$, restricted to the center most invariant interval (top right). Rescale the center box to the unit box again to complete $R(R(f))$ (bottom right). $R^2(f)$ is again renormalizable (bottom left) but $R^3(f)$ is not, escaping the space $\mathcal{U}$ because it no longer admits the required interval combinatory: $I_0 \to I_1 \to I_0$. This illustration is for the logistic map f_r at the period-8 bifurcation point r_3.

in I_0. Thus, to make it unimodal again we only need to flip the interval I_0 left right and the image of f^2 upside down to make the return map unimodal before scaling the definition interval to the unit interval. The effect of these reflective

flips can also be achieved by rotating the I_0-by-I_0 box 180° before the scaling. With this change of orientation of I_0, the definition of $R(f)$ is at last completed.

The image $R(f)$ may or may not be in $\mathcal{U}$, depending on only one thing. Because of the reflective flips when f is unimodal in I_0, $R(f)$ is automatically unimodal. Thus whether or not the renormalized image is in $\mathcal{U}$ is whether or not the image $R(f)$ is renormalizable, i.e. having its own $I_0 \to I_1 \to I_0$ recurring split. If so, the second iterate $R^2(f) = R(R(f))$ is also defined and so on. Fig.3.2 shows the case for which $R(f), R^2(f)$ are renomalizable but $R^3(f)$ is not. Notice that the iterative sequence of R is essentially a family of return maps of f on nested subintervals.

3.1.3 Hyperbolic Decomposition

In a true dynamical systems fashion, Feigenbaum, Coullet, and Tresser suggested the following picture, which was rigorously proved later by others, see Fig.3.3. The mapping R has a unique fixed point g in $\mathcal{U}$, $g = R(g)$. This fixed point admits a hyperbolic decomposition: $W^u \oplus W^s$, of which the unstable manifold W^u is 1-dimensional and the stable manifold W^s is of $\mathcal{U}$'s dimension but one, namely W^s is a co-dimension-one manifold. Both intersect each other at g transversely. Moreover, the one-parameter family of the logistic maps f_r forms a smooth curve in $\mathcal{U}$ and it intersects the stable manifold W^s transversely at a point precisely corresponding to the onset of chaos at the end of the period-doubling cascade, f_{r_c}. Carried forward by R, the family of curves $R^n(f_r)$ converges to the unstable manifold W^u. The image family can be a part of continuous unstable foliation of the fixed point g.

Fig.3.3 also shows a part of a stable foliation $\mathcal{F}^s_{r_k}$. Here and anywhere else, a foliation is nothing more than a class or classification of objects or points. The class of $\mathcal{F}^s_{r_1}$ consists of all unimodal maps that undergo a period-doubling bifurcation from a fixed point to a period-2 orbit, at which the derivative of the unimodal map f at the fixed point $\bar{x}$ is -1, i.e. $f'(\bar{x}) = -1$ and $f(\bar{x}) = \bar{x}$. Similarly, the class of $\mathcal{F}^s_{r_2}$ consists of unimodal maps which undergo a period-doubling bifurcation from a period-two orbit to a period-4 orbit, and so on. For this foliation, the logistic mapping f_r at, say, the period-8$(= 2^3)$ bifurcation point $r = r_3$ of the cascade, is on the leaf $\mathcal{F}^s_{r_3}$. This means that the stable period-4$(= 2^2)$ orbit of f_r is undergoing a period-doubling bifurcation at $r = r_3$ to become a period-8 orbit. Because the renormalized mapping $R(f_r)$ is essentially the same mapping f_r, i.e. a return map of f_r on a shorter interval, counting every two iterates of f_r as one iterate of $R(f_r)$, and every four (2^2) iterates of

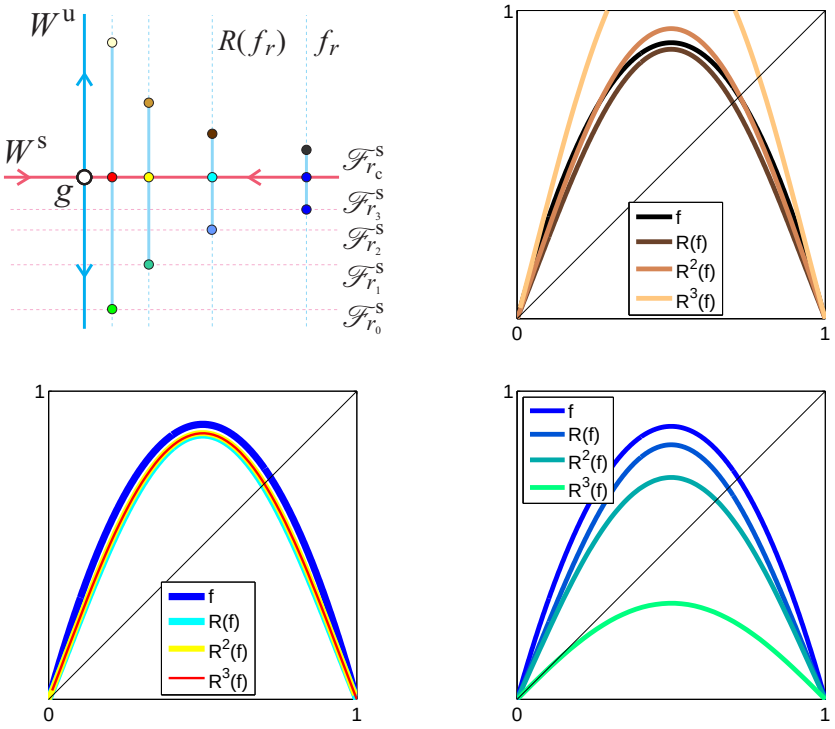


Figure 3.3: Top left: the dynamical systems view of the map R. The rest plots show how some typical points behave in the functional space $\mathcal{U}$ under the action of R. Top right: three iterates of f_r by R for $r = 3.583$, corresponding a point lying above the co-dimension 1 stable manifold W^s. $R^3(f_r)$ is outside $\mathcal{U}$. Bottom left: by the third iteration, $R^3(f_{r_c})$ already looks fixed, an approximation of the fixed point g. It is a good approximation of g because the convergence to g on W^s is exponential. Bottom right: the finite orbit of f_{r_3} by R. The third iterate is outside R's domain of definition $\mathcal{U}$.

f_r as one iterate of $R^2(f_r)$, and every eight (2^3) iterates of f_r as one iterate of $R^3(f_r)$, and so on. Thus, the parameter value r_3 for f_r is for a corresponding period-2 orbit undergoing period-doubling to become a period-4 orbit for $R(f_r)$, and for a fixed point undergoing period-doubling to become a period-2 orbit for $R^2(f_r)$, and so on. Therefore, $R(f_{r_3})$ and f_{r_2} belong to the same leaf $\mathcal{F}^s_{r_2}$, all undergoing period-doubling from period-2 to period-4 bifurcation. Similarly, $R^2(f_{r_3})$ and f_{r_1} belong to leaf $\mathcal{F}^s_{r_1}$, and so on. This explanation shows the

foliation family $\mathcal{F}^s_{r_k}$ is invariant under the renormalization, i.e. $R(\mathcal{F}^s_{r_k}) \subset \mathcal{F}^s_{r_{k-1}}$. Since we know the period-doubling bifurcation at a fixed point $\bar{x}$ of a unimodal map is characterized by the condition $\bar{x} = f(\bar{x})$ and $f'(\bar{x}) = -1$, Fig.3.3 shows it for $R^2(f_{r_3})$ for which its derivative at the interior fixed point is -1. However, rolling forward two more iterations, this interior fixed point of f^{2^2} becomes the boundary fixed point of the return map f^{2^3} or the trivial fixed point $\bar{x} = 0$ for the third iterate of the renormalization $R^3(f_{r_3})$, at which its derivative becomes $+1$ instead. At this point it is no longer renormalizable.

3.1.4 Universal Number Is An Expanding Eigenvalue

As the limiting map f_{r_c} is on the stable manifold W^s, it is indefinitely renormalizable and $R^n(f_{r_c})$ converges to the fixed point g. For every other parameter value $r \neq r_c$, f_r is only finitely renormalizable, diverging away from g along the W^u direction. The universality about the number δ lies in the fact that orbits on the unstable manifold W^u are repelled away from g at the exponential rate $\delta = 4.6692....$ As the stable leaves $\mathcal{F}^s_{r_n}$ intersect the unstable stem W^u transversely and they are invariant under the renormalization R, the distance between neighboring leaves shrink at the reciprocal rate $1/\delta$ with increasing n. This adjacent distance is approximated by the parameter difference $r_n - r_{n-1}$. Thus as an exponentially decreasing sequence it is captured exactly by the aforementioned ratio test:

$$\lim_{n\to\infty} \frac{r_n - r_{n-1}}{r_{n+1} - r_n} = \delta.$$

The Feigenbaum-Coullet-Tresser Universality is about the fact that the logistic family f_r is just one family of dynamical systems going through the period-doubling cascade to chaos, and any sufficient nice family of dynamical systems, mappings or differential equations, going through the same type of cascade permits exactly the same description above, the limit behavior in particular. In the exact technical term, δ is the unique expanding eigenvalue of the derivative $DR(g)$ of the operator R at the fixed point g. The essence of δ's universality.

3.2 Return Map of Bursting Spikes

Let us switch topic to a different type of bifurcations. It is about cross-membrane voltage spikes from the class of neuronal circuit models introduced in a previous

chapter. We recall one of the models here below for conveniences,

$$\begin{cases} CV_{\rm C}{}' = -[I_{\rm K} + f_{\rm Na}(V_{\rm C} - \bar{E}_{\rm Na}) + I_{\rm pump} - I_{\rm ext}] \\ I_{\rm pump}{}' = \lambda I_{\rm S}[V_{\rm C} - \gamma I_{\rm pump}] \\ I_{\rm S}{}' = \lambda I_{\rm pump}[V_{\rm C} - \gamma I_{\rm pump}] \\ \epsilon I_{\rm K}{}' = V_{\rm C} - \bar{E}_{\rm K} - h_{\rm K}(I_{\rm K}). \end{cases} \tag{3.1}$$

We have learned that this type of model is capable of generating spikes on transient with the absolute total pump current $I_{\rm S}$ changing so slowly that it can be treated like a parameter. For all intents and purposes we will drop the $I_{\rm S}$ equation and treat $I_{\rm S}$ as a parameter.

For the resulting 3-dimensional system, the $I_{\rm K}$ is the fastest variable because its equation is to simulate the idea Z-hysteresis $V_{\rm C} = \bar{E}_{\rm K} + h_{\rm K}(I_{\rm K})$. The cross-membrane voltage $V_{\rm C}$ equation is the fast variable because the capacitance C is usually small to generate healthy spikes. The net pump current $I_{\rm pump}$-equation is the slow, i.e. the slowest equation.

These conditions provide an idea set-up for a singular perturbation analysis. The $I_{\rm K}$-hysteresis is an ideal Z-switch. For the parameter region we are interested in but without the need to specify, the reduced 2-dimensional fast-slow flow on the top branch of the switch is pulled to the top turning edge of the switch. On the bottom branch of the z-switch, the reduced flow is again a fast-slow system with $C\lambda I_{\rm S}$ as the combined slow parameter. We will lump them together to call it $\mu = C\lambda I_{\rm S}$. The reduced fast-slow flow is determined by the nullclines $V_{\rm C}{}' = 0$ and $I_{\rm pump}{}' = 0$ with the former forming the slow manifold of the 2-d system. It shapes like a letter V on its side. The horizontal part of the side-way V is the attracting branch of the slow manifold. The tip of the V is a turning point for the reduced μ-singular system. The slant part of the V intersects the bottom turning edge of the Z-switch at a K-fold turning point K_1. The $I_{\rm pump}$-nullcline $I_{\rm pump}{}' = 0$ is simple. It cuts a line on the bottom branch of the Z-switch, above it trajectories move rightward in the increasing $I_{\rm pump}$ direction and below it trajectories move leftward. Last, if the $I_{\rm pump}$-nullcline $I_{\rm pump}{}' = 0$ is not caught between the junction lines and the turning edges of the Z-switch, the prototypical configuration for spike generation is finalized.

Fig.3.4 gives a numerical simulation for the description given above. The top-left figure shows a train of four spikes follows a refractory phase, which resets the membrane for another burst of four spikes. The number of spikes in each burst is determined by the slow parameter μ: the smaller the μ the more spikes in a burst. We note that this phenomenon of spike-burst is not unique to neuron models. The first food chain model from the previous chapter is capable of exhibiting it too.

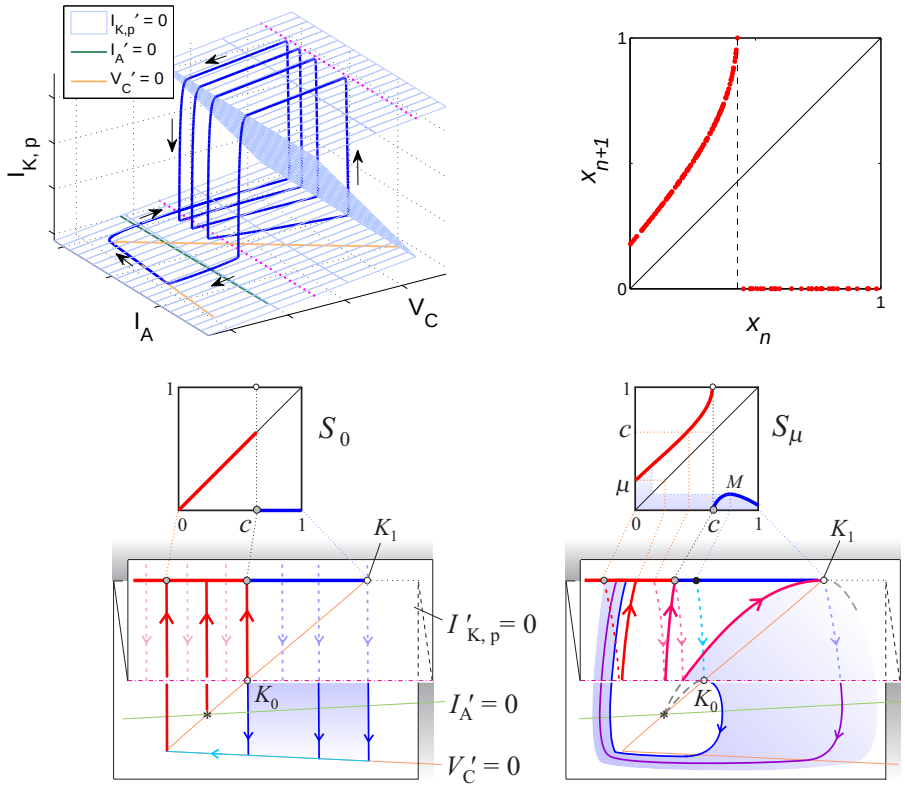


Figure 3.4: Top-left: bursts of spikes determined by the slow manifold ${I_\mathrm{K}}' = 0$, the slower manifold ${V_\mathrm{C}}' = 0$, and the nullcline ${I_\mathrm{pump}}' = 0$. Arrows represent the direction of the multi-time scale vector field. Top-right: a return map of the model with x_n representing the pump current variable I_pump. Bottom-left: the return map at the limiting value $\mu = 0$. Bottom-right: the return map for small $\mu > 0$. Both looking down at the I_K-hysteresis from the I_K-axis.

3.2.1 Fast-Slow Deconstruction Of Spike

For further simplification, we take up the idea of Poincaré return map. One can take any two-dimensional cross-section through which the vector field of the model intersects transversely. Because the model is naturally multi-scale in time with the I_K variable switching up and down by its hysteresis, the system is essentially a concatenation of the reduced μ-singular systems on the top and bottom branches of the hysteresis switch, pieced together by the fast jumps

between the switch's turning edges and their junction lines. As a result any cross-section can be effectively a line or a curve transversal to the reduced μ-singular system, or a line through the plane between a turning edge of the switch and its junction line. The corresponding return map at the singular limit $\epsilon = 0$ is 1-dimensional, a map from a line segment to itself. For a non-zero ϵ but extremely small, we can still use the limiting 1-d return map as a very good approximation.

A simulated return map for the model is shown in Fig.3.4 that demonstrates this 1-dimensionality property of the full return map. The map is neatly divided into two intervals, or two phases, by a discontinuity critical point, which we denote by c throughout. From the left interval, the map is monotone increasing above the diagonal line, representing the active phase of spiking of the model. In the right interval, the graph of the map is nearly flat, representing the silent phase of the action potential of the model. It resets the membrane back to the active phase for a new burst of spikes. The presence of the discontinuity point between the active and silent phases is due to the existence of the K-fold turning point at which the $V_{\rm C}$-nullcline intersects the bottom turning edge of the Z-switch. At this point, the reduced fast-slow trajectory on the bottom branch of the switch is tangent to the turning edge.

As mentioned above the return map not at the ideal singular value $\epsilon = 0$ can be approximated by the one at the limit. What makes the singular perturbed model even more appealing is the fact the limiting case can be made precise by a geometrical analysis, which is illustrated in the bottom figures of Fig.3.4. All it takes for the geometric analysis to work is to piece together the μ-fast-slow flows on the top and bottom branches of the Z-switch — how do the flows hit the turning edges of the Z-switch and where do they land on the junction branches of the switch, and so on.

For the illustration, all dashed flow lines with arrows are trajectories of the reduced μ-fast-slow system on the top branch of Z-switch, and all solid flow lines with arrows are the said trajectories on the bottom branch. The top flow lines have one simple motion — all move toward the top turning edge to be switched down. The bottom flow lines can move in two general ways. The flow line that separates others into these two groups is the trajectory connecting the K-fold turning point K_1. Call it Γ. If a flow line starts on the left side of Γ it moves up toward the bottom turning edge, and jumps up to the top branch following the fastest $I_{\rm K}$-dynamics. If it starts right of Γ it will eventually move down toward the horizontal side of the the $V_{\rm C}$-nullcline, i.e. the slow-manifold of the reduced μ-singular system. This branch of the side-way V-shaped slow

manifold represents the silent phase of the spike dynamics. The flow line then moves slowly leftward toward the turning point of the V_C-nullcline, the knee point of the V. Goes around the point and moves up like any flow line left of the flow line Γ. The V_C-turning point represents the start of a new burst of spikes. The flow lines right of Γ are further divided up by the K-fold junction point K_0 at which the unstable part of the V_C-nullcline intersects the junction line of the top turning edge of the Z-switch. At this K-fold junction point K_0 the flow line is tangential to the junction line. To the left side of K_0 the flow line first goes up and then down. To the right side of K_0 the flow line heads straight down toward the slow manifold of the reduced μ-system.

As we have seen in the previous chapter, if a return map's defining flow lines include a K-fold turning point, such as K_1, the map will have a discontinuity point. In addition, if its defining flow lines include a K-fold point, such as K_0, the return map will have a fold point which is either a local maximum or a local minimum of the map. If these descriptions and the figures are enough to convince you that a return map can be constructed, you can skip right to the return map families S_μ, s_μ, ψ_μ, and then to the next section. Otherwise, read on to find out how return maps are pieced together.

3.2.2 Geometric Portrait Of Return Map

In fact, we can make this general description a little bit more precise. We can do so by looking at the limiting case with $\mu = 0$, which is the simplest, and separately a typical case with $\mu > 0$. To start, we can take any line segment as a cross-section as long as the line is transversal to the singular vector field at $\epsilon = 0$. For example, we will take the cross-section right on the bottom turning edge, but left the K-fold turning point K_1. We cannot include K_1 because the singular flow fails to be transversal to the edge at this point.

For the case of $\mu = 0$, the bottom-left figure of Fig.3.4 gives an illustration for the return map. The cross-section interval is divided by a discontinuity point c, which corresponds exactly to the K-fold junction point K_0. To the left of c, every point from the cross-section return exactly to itself. The map maps the interval left of c onto the diagonal. That is, when the pump current I_{pump} is fixed at a constant left of the K_0 value, the circuit is in a perpetual state of periodic firing. On the other hand, for any point right of K_0 (i.e. c for the return map), the flow first goes up to the Z-switch, then down on the top branch of the Z, and down off the branch's turning edge. Instead of heading up for points from the left side of c, it moves down on the bottom branch to the

steady state $V_C{}' = 0$. That point is then concatenated to the turning point of the V_C-nullcline by the μ-slow orbit along the nullcline μ-slow manifold. The turning point brings the concatenation back to the cross-section to complete the return of all points right of c. As a result, the graph of the return map right of c is a flat line. Since this return map is about capturing the spikes of the model, we use S_μ to denote the return map, with the cross-section normalized to be the unit interval $[0, 1]$. The description above is thus for the graph of S_0.

The case with a typical non-vanishing value $\mu > 0$ is illustrated by the bottom-right figure of Fig.3.4. It is of course a bit more involved than the simpler limiting case of $\mu = 0$. But the complexity is only about keeping track of the two K-fold points K_0, K_1, on how the fast-slow flow on the bottom branch of the Z-switch is affected by their existence. For example, the return map will still have a discontinuity point c. Instead of directly lining up with the K_0 point, it is related to the K_1 point in concatenation. First, the ϵ-fast flow brings it to the top branch of the Z-switch. A μ-fast-slow flow brings to the top turning edge and then down to a junction point on the bottom branch of the Z-switch. That junction point is connected to the K-fold turning point K_1. Now, depending which side of the critical point c, the subsequent flow line can proceed in different directions. If one takes the left side of c, then the point is considered to have made the required return to the cross-section. If one takes the right side of c, then the point is considered to miss the cross-section and continue on the bottom branch of the switch, first across the V_C-nullcline, around its turning point, and at last returning to the cross-section at its lowest end which is designated as 0 for the return map on the unit interval.

To the left of the critical point c, every point from the cross-section will go around the Z-switch, shift rightward, and return to the cross-section, generating one spike in the process. The reason that it progresses rightward is because the fast and slow orbits in the concatenation lie above the I_{pump}-nullcline in which $I_{\text{pump}}{}' > 0$. Thus, the graph of S_μ is monotone increasing, above the diagonal line. It represents the active phase of the model in which the train of spikes moves rightward.

For point right of c, it goes around the Z-switch but does not return right away like points from the other side. Instead, it falls to the right side of the slow trajectory Γ that defines the point c. As a result, it will moves around the turning point of the V_C-nullcline before returning to the cross-section near its lower end. One point stands out in this interval. It is the point which connects to the K-fold junction point K_0 on the bottom branch of the switch. Such a point folds the flow around it, creating a local maximum in the right interval.

The local maximum value is denoted by M in the illustration. The equilibrium point at which all three nullclines intersect plays an indirect role. From which the slow trajectories connecting the two K-fold points arise. Since it stays in one side of the turning edges and their junction lines of the switch, its effect does not show up directly in the return map S_μ.

3.2.3 Quantitative Fit Of Return Map

The description above alone explains qualitatively how spikes are generated. But to understand how the number of spikes changes from n to, say, $n+1$, we need to incorporate quantitative information into the return map S_μ. We will take a short cut to do it because it will take too much space to derive it from the model even though the model is already simple enough. We will do so phenomenologically.

First, the functional form of S_μ in the silent phase interval right of c is easier to fit. We know it has the minimum vale at the left end point c and a local maximum between c and 1. So we will first fit it with this function: $(x-c)\exp(-(x-c)/(b\mu)$. This function is 0 at $x=c$ and is maximal at $x=c+b\mu$ with the maximal value $b\mu\exp(-1)$. Here b is a parameter. We also know that the maximum value of the return map, i.e. amplitude is exponentially small against the slow parameter μ as the lower branch of the $V_{\rm C}$-nullcline attracts the silent phase points exponentially in time. We will complete the fit by multiplying the function with an exponentially small amplitude. The final fit in the right interval is $S_\mu = a\exp(-d/\mu)(x-c)\exp(-(x-c)/(b\mu)$ for $c < x \le 1$, with $a\exp(-d/\mu)$ being the control on the amplitude.

The fit of S_μ in the active interval takes some work. We know it is the diagonal x when $\mu = 0$, and it is more or less parallel the diagonal and above it when $\mu > 0$. We also know it takes the fixed maximum value 1 at the left limit of c. We also know it has a vertical tangency at the left limit of c for $\mu > 0$ because the slow flow on the bottom branch of the Z-switch has a horizontal tangency at the K-fold turning point K_1. Here is the functional form we use to fit S_μ in the left interval

$$S_\mu(x) = \mu + x + [1-(\mu+c)]\frac{w\mu^p}{w\mu^p + (c-x)^{1-q\mu}} \qquad \text{for } 0 \le x \le c.$$

We impose the constraint $1-\mu-c > 0$ so that the last term is a non-negative addition to the line $\mu + x$ parallel the diagonal. We require $p > 1$ so that the addition is a high order correction term to the line $\mu + x$. We need the power $1-q\mu$ to be less than 1 because we want the derivative of $(c-x)^{1-q\mu}$ at $x = c$

to be infinity. The use of the quotient form is motivated by the Holling Type II functional form $x/(h+x)$ which saturates at 1 for large x with h being the half saturation point. All parameters are non-negative. It is straightforward to check $S_\mu(c) = 1,\ S'_\mu(c) = \infty$ and $S_0(x) = x$, and that S_μ is approximately $\mu + x$ but with a rapid ascension only in a μ-neighborhood of the critical point.

To summarize, the phenomenological return map is

$$S_\mu(x) = \begin{cases} \mu + x + [1-(\mu+c)]\frac{w\mu^p}{w\mu^p + (c-x)^{1-q\mu}}, & 0 \le x \le c \\ a\exp\left(-\frac{d}{\mu}\right)(x-c)\exp\left(-\frac{x-c}{b\mu}\right), & c < x < 1 \end{cases} \tag{3.2}$$

for which all parameters are non-negative with $0 < c < 1$, $p > 1$, and $\mu < 1 - c$.

Of course, this return map is defined on the maximal cross-section possible, all the way to the K-fold turning point K_1 at which the vector field fails to be transversal to the cross-section. One can choose a shorter cross-section to define a return map. For example, one can take the interval left of the critical point c to be the cross-section. One can either directly derive a phenomenological return map or use S_μ to find a return map of it on the left interval. We will pursue the latter idea in greater details later. For now, let us use the direct fit to derive such a short return map.

We will use Fig.3.4 as a prop. From the $\mu > 0$ illustration, we designate c to be the new 1, the right end of the cross-section. But we must use the old c's immediate backward concatenated point on the cross-section as the new critical point c. Instead of staying in the interior of the interval $[0, 1]$ like the old c, this new critical point depends on μ, i.e. $c = c(\mu)$. In particular, at $\mu = 0$, $c(0) = 1$ as all the spikes are frozen as fixed points. If we denote this short return map by s_μ, then $s_0(x) = x$ for $0 \le x < 1$, and $s_\mu(x) = \mu + x$ for $0 \le x < c(\mu)$. For points from the right interval $(c(\mu), 1]$, they will first hit the turning edge between the old c and K_1, not returning to s_μ's cross-section, and then hit the silent phase space before rounding around the turning point of the V_C-nullcline and returning to the new cross-section. We can lump all the actions together by a similar bump function of the same form of S_μ on its right interval. A phenomenological fit to the short return map is thus given as below,

$$s_\mu(x) = \begin{cases} \mu + x, & 0 \le x \le 1-\mu \\ a\exp\left(-\frac{d}{\mu}\right)(x-1+\mu)\exp\left(-\frac{x-1+\mu}{b\mu}\right), & 1-\mu < x \le 1 \end{cases} \tag{3.3}$$

with c being substituted by $1 - \mu$ for simplicity.

Two key differences stand out between the full return map S_μ and the short one s_μ: the discontinuous critical point for S_μ is more or less fixed and is in the interior of $[0, 1]$ but that for s_μ is not, moving to 1 at $\mu = 0$; and S_μ has a

vertical tangency at the left of c but s_μ does not. Although s_μ should have a vertical tangency at the right end point 1 of the interval, we will ignore it since its amplitude is dominated by the exponential decay of the slow trajectories journeying through the silent phase. Both are good approximations of return maps of the differential equations at the singular limit $\epsilon = 0$, the short return map s_μ is simpler than the long one S_μ.

Because of its simplicity, one more special subfamily of the short return map s_μ is considered in subsequent sections. It is given a name, the ψ_μ family. It is s_μ with the silent phase amplitude parameter a setting to zero. That is,

$$\psi_\mu(x) = \begin{cases} \mu + x, & 0 \le x \le 1 - \mu \\ 0, & 1 - \mu < x \le 1. \end{cases} \tag{3.4}$$

Obviously, of all maps ψ_μ is the simplest to approximate the spike generation of the neuronal models.

3.3 Isospiking Bifurcation

The number of spikes in a burst after a neuronal model starts in its silent phase has a direct correspondence with the number of iterates in the active phase interval of a return map of the model when the iterate starts in the silent phase interval. For the ψ-family of return map, this spike dynamics can be determined completely.

In fact, start any point $x_0 \in [1 - \mu, 1]$ from the right interval, $x_1 = \psi(x_0)$ is always zero because $\psi(x) = 0$ for $1 - \mu \le x \le 1$, with the interpretation that $x_1 = 0 \in [0, c)$ is counted for 1 spike. Here $c = 1 - \mu$ is the discontinuity point separating the active and the silent intervals. Because $\psi(x) = \mu + x$ for $x < c$, the 2nd iterate is $x_2 = \psi(x_1) = \mu + x_1 = \mu$, the 3rd $x_3 = \mu + x_2 = 2\mu$, etc. The $(k+1)$st iterate is explicitly

$$x_{k+1} = k\mu$$

as long as the kth iterate x_k is in the active interval $x_k < c$. If the $(k+1)$st iterate falls into the silent phase interval, the burst stops, having exactly k spikes. Because of the form of this family, every initial point from the silent interval goes through the exact spike iterates and has the exact number of spikes in one burst.

The condition for the k-spikes burst is

$$x_k < c \le x_{k+1}, \text{ i.e. } (k-1)\mu < 1 - \mu \le k\mu.$$

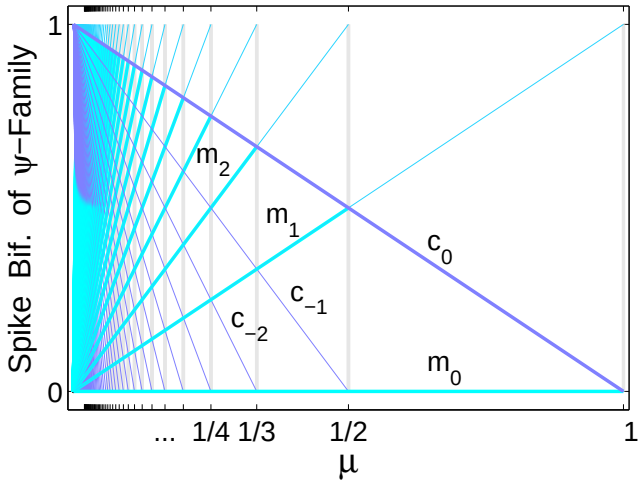


Figure 3.5: Spike bifurcation diagram for the ψ-family.

Solving the inequalities to obtain $\frac{1}{k+1} \leq \mu < \frac{1}{k}$. Now let

$$\mu_k = \frac{1}{k+1}.$$

Then the map ψ_μ has exact k spikes for each burst if $\mu_k < \mu < \mu_{k-1}$. If the parameter μ slips down through μ_k, the map will gain at least one more spike to have at least $k+1$ spikes. If the parameter climes up through μ_{k-1}, the map will lose at least one spike to have at most $k-1$ spikes.

3.3.1 Harmonic Sequence Progression

This simple result gives the most enduring feature of all spike return maps. That is, as the singular parameter μ decreases to its natural limit $\mu = 0$, the spike number progresses like the natural numbers, from 1 to 2, to 3, and so on. In addition, the bifurcation parameter μ_k at which the spike number progresses from k to $k+1$ forms a decreasing sequence converging to the singular limit and most importantly the bifurcation sequence is the harmonic sequence:

$$\mu_k = \frac{1}{1}, \ \frac{1}{2}, \ \frac{1}{3}, \ \frac{1}{4}, \cdots \to 0.$$

As an illustration, Fig.3.5 is the spike bifurcation diagram for the ψ-family. In it, the c_0 line is the critical point: $c_0 = 1 - \mu$. The c_{-1} line is the pre-image

of c_0 under the map, i.e. $\psi_\mu(c_{-1}) = c_0$. It is $c_{-1} = 1 - 2\mu$. The c_{-k} line is the pre-image line of c_{-k-1}, and so on, and it is exactly $c_{-k} = 1 - (k+1)\mu$. In the opposite direction, m_0 is the forward iterate of the critical line c_0 with $m_0 = \psi_\mu(c_0) = 0$, and m_k is the kth iterate of m_0. They are given exactly as $m_k = k\mu$. The intersections of the m-lines with the critical line c_0 are the spike bifurcation point: μ_k is the μ-value of the intersection of m_{k-1} with c_0. Above the c_0 line is the silent phase region. For any $\mu \in (0, 1)$, one only needs to count how many m-lines lie below c_0 to determine the number of spike per burst. The map has k spikes per burst if the diagram has k many m-lines, including the m_0 line always, under c_0. The slope of the m-family lines progresses up on the positive integers, and the slope of the c-family lines progress down on the negative integers. Wherever they intersect the per-burst spike number bifurcates, and they only intersect along the harmonic sequence.

3.3.2 What Isospiking Is

Next, we take a look at how many of the spike bifurcation properties of the ψ-family are preserved for the short return map family s_μ, which is a closer approximation to the neuronal models than the ψ-family is. The key difference is that the ψ-family has a zero maximum in the silent phase interval but the s-family has a non-zero maximum instead $M = \max\{s_\mu(x)) : c < x < 1\} = a\exp(-d/\mu)b\mu\exp(-1)$. This difference makes it possible that not all silent phase initial points will have the same number of active phase iterates. That is, the per-burst spikes may vary. This leads to the following definition:

Definition: *The dynamics of the system (or the parameter that represents the system) is said to be isospiking if the per-burst spike number is fixed for every initial point from the silent phase.*

For all the spike return maps, finding the conditions for isospiking parameters is surprisingly straightforward. It is all determined by keeping track of the minimal value, m, and the maximal value M of the map in the silent phase interval $[c, 1]$. The minimum m is always normalized to 0 but the maximum M is not, which is an offset above the baseline $m = 0$. Denoting their forward iterates by $m_0, m_1, m_2, \ldots$ and $M_0, M_1, M_2, \ldots$, with $m_0 = m$ and $M_0 = M$. Then we realize that the maxmin iterative pair forms what we call the orbital corridor by their cobweb plots, see Fig.3.6. That is, every initial point from the silent phase interval must move inside the maxmin corridor. The space outside the corridor is forbidden for all iterates as long as they start from the right interval. It is abundantly clear that if the critical point c_0 is inside the corridor,

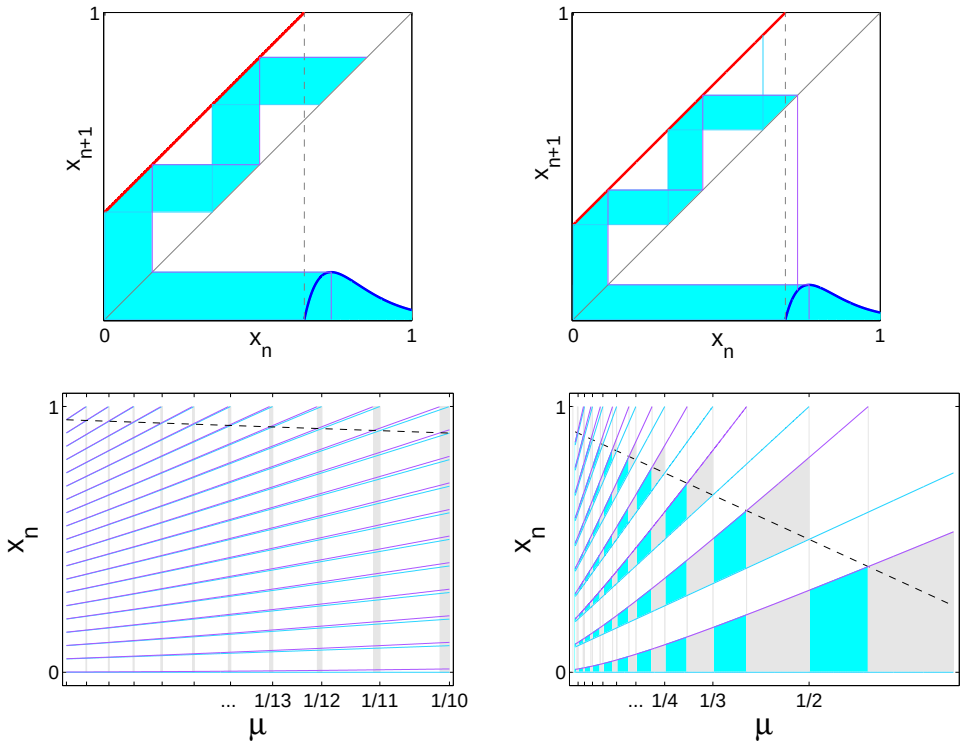


Figure 3.6: Top: the maxmin orbital corridor and the configurations for isospiking and non-isospiking state. Bottom: the spike bifurcation diagram for the s-family of return maps. The corridor is bounded by the m iterative curves (cyan) and the M iterative curves (light purple). When the critical curve c_0 (dashed) splits the corridor, the state is non-isospiking, with the region colored gray. If it does not, the state is isospiking, with the region colored cyan. The number of cyan patches below the c_0 curve is the isospike number.

i.e. splits the corridor, the parameter value or the state of the system is not isospiking. If it is outside the corridor, the map is isospiking. The isospiking condition is simply as follows:

$$M_{k-1} < c_0 < m_k.$$

And when the condition holds, the map is of isospiking of k spikes per burst for all bursts. If the critical point splits the corridor with

$$m_k < c_0 < M_k,$$

then some bursts will have $k+1$ spikes and some bursts will have k spikes.

The bifurcation values at which the corridor boundary and the critical point cross each other can be computed. We already have the value from the ψ-family above for the m-iterate corridor boundary crossing, $c_0 = m_k$, whose solution is $\mu = \mu_k = 1/(k+1)$. We denote this sequence value by $\mu_k^- = 1/(k+1)$. The superscript is used to connote the fact that it is the intersect point of the critical line c_0 with the lower corridor boundary m_k. Likewise, we denote by μ_k^+ the μ-parameter value when the upper corridor boundary $M_k = M + k\mu$ crosses the critical point c_0. For the non-isospiking interval $m_k < c_0 < M_k$, we solve the left and right inequalities simultaneously to get

$$\mu_k^+ < \mu < \mu_k^-.$$

It is a simple algebraic exercise to get $\mu_k^+ = (1-M)/(k+1)$ from the equation $c_0 = M_k$. Here because $M = ab\mu e^{-d/\mu - 1}$ is exponentially small as $\mu \to 0$ and $\mu_k^- = 1/(k+1)$, we reach the simple but important conclusion that the non-isospiking interval is exponentially narrow! That is, if the non-isospiking state of neurons is undesirable for neurological reasons, this result simply implies that the bad set is negligible. We seem to have a mathematical basis here to posit that evolution is responsible for this favorable outcome.

Once we have found the non-isospiking intervals $\mu_k^+ < \mu < \mu_k^-$, the isospiking intervals are easy to understand. As μ moves below the plus-point μ_k^+, the system enters the $(k+1)$-isospiking state. As it moves above the minus-point μ_k^-, the system enters the k-isospiking state. Hence, as μ decreases to the singular limit $\mu = 0$, the minus-point μ_k^- is the bifurcation point at which the system leaves behind the k-isospiking state, and the plus-point μ_k^+ is the bifurcation point at which the system enters into the $(k+1)$-isospiking state. In between lies the $k(k+1)$-non-isospiking state. So if we take the decreasing direction of μ as the preferred reference direction, we can refer to the plus-point μ_k^+ as the start of the $(k+1)$-isospiking state, and to the minus-point μ_k^- as the end of the k-isospiking state. Since the start of something must the be end of something else, we can also say, for example, that the minus-point μ_k^- is the start of the $k(k+1)$-non-isospiking state.

3.3.3 Notation For The Start And End States Of Isospiking

The notation, description used above are still not descriptive enough. It becomes more confusing as the bifurcation points are going down rather than up, which is the most natural direction for most people. We will give a try to simplify the

matter. Since α and ω are the first and the last letter in the Greek alphabet, we will denote by $\alpha_k = \mu_{k-1}^+$ for the start of the k-isospiking parameters and $\omega_k = \mu_k^-$ for the end of the k-isospiking parameters. Thus the isospiking intervals are

$$\ldots, (\omega_k, \alpha_k), \ldots (\omega_3, \alpha_3), (\omega_2, \alpha_2), (\omega_1, \alpha_1),$$

i.e., the interval sequence $\{(\omega_k, \alpha_k)\}$ is decreasing in order as μ is toward the singular limit $\mu = 0$. For the s-family, the end of isospiking is the harmonic sequence $\{\omega_k\} = \{1/(k+1)\}$ and the non-isospiking interval width $|\omega_k - \alpha_{k+1}| \sim \exp(-1/\mu) \sim \exp(-k)$ is exponential small as μ is approximated by $1/k$ in the interval.

We also gained a valuable insight from the work-out above. That is as long as spike bifurcation is concerned only the minimal and the maximal point from the silent interval matter, nothing else. The minimum takes place to the right side of the critical point c and the maximum takes place in the interior of the interval $(c, 1)$. The right end point 1 plays no role nor any point else except for the two extrema.

3.3.4 More Accurate Model But Similar Result

We are now ready to take a look at the full return map family S_μ. We will first formalize a notation we used informally above. Two points x and y are said orbitally related by a map f or simply related within the same context if $x_k = f^k(x) = y$ or $y_k = f^k(y) = x$ for some iterate k. For $k = 0$ the iterate f^0 is just the identity map. We have for this trivial case that $x_0 = x$. That was the reason we used c and c_0 interchangeably above. With negative integer subscript, we say c_{-k} is related to c_0 if $f^k(c_{-k}) = c_0$ with $f = S_\mu, s_\mu$. Because the spike maps are monotone increasing the first backward iterate c_{-1} of c_0 always exists if the minimum of f in the left interval is lower than the critical point, $f(0) = m < c_0$. With this notation in place, we can now consider the return map of S_μ in the interval $[0, c_0]$, and denote it by rS_μ. As we have seen from the previous section, rS_μ is just return map of the singularly perturbed neuronal model on a shorter cross-section. That is, a return map of a return map of a flow is just a return map of the flow.

This return map rS_μ's definition interval $[0, c_0]$ can be divided into the left and the right intervals by the backward iterate c_{-1} of c_0: $[0, c_{-1}], (c_{-1}, c_0]$. Since S is monotone in $[0, c_0]$ it maps $[0, c_{-1}]$ to $[0, c_0]$, the automatic return. So we do nothing, and assign $rS_\mu(x) = S_\mu(x)$ in the left interval $[0, c_{-1}]$. For the right interval $(c_{-1}, c_0]$, the return map takes two applications of the original map S_μ

to define. The first application maps $(c_{-1}, c_0]$ out to $(c_0, 1]$, the silent phase interval of the old map, the second application returns the silent interval back to the definition interval $(0, c_0]$ of the return map. That is, $rS_\mu(x) = S_\mu^2(x)$ for x from $(c_{-1}, c_0]$. Equivalently, the right interval $(c_{-1}, c_0]$ is the silent phase interval for rS_μ.

The return map preserves the minimal value, m, and the maximal value, M, of the original map in their respective silent intervals. This is the only property we need and care about for the silent phase interval as far as the spike bifurcations are concerned. Because they are the same map, the bifurcations points α_k, ω_k are the same except that the subscript shifts by 1. This is because for the return map these bifurcation points are the intersections of m_k, M_k with c_{-1}, which are equivalently the intersections of m_{k+1}, M_{k+1} with c_0 for the original map. Therefore, it is only left to find out what c_{-1} behaves like.

First of all, because the graph of S_μ in the spiking interval $[0, c_0]$ is at least a μ-displacement above the diagonal line, c_{-1} is at least that much of a magnitude away from c_0: $c_0 - c_{-1} > \mu$. Therefore, every point x from the left interval of the return map, $0 \le x < c_{-1}$, is at least that much of a distance away from c_0: $c_0 - x > \mu$. Because of this estimation, the third term of S_μ is truly a higher μ-order term and therefore can be ignored: $w\mu^p/(w\mu^p + (c-x)^{1-q\mu}) \sim \mu^p/\mu^{1-q\mu} = \mu^n$ for $n = p + q\mu - 1 > 1$ if $p > 2$. That is, the sharp ascension of S_μ to the left of c takes place within a distance less than the order of μ. Since the behaviors of S_μ in this interval, regular or peculiar at the left side of c_0, never affect the minimal value or the maximal value of the right interval, they can be all ignored for the return map rS_μ.

We are left with this simple fact. For all intents and purposes, the return map rS_μ is approximated by $\mu + x$ in the left interval and the same functional form as S_μ in the right interval. If we scale its interval $[0, c_0]$ to the unit interval $[0, 1]$, and call it RS_μ, then for all intents and purposes we can replace RS_μ by a form similar to the s_μ-family. This is because scaling of the variable does not change the parameter nor the bifurcation points. The description above is for some people a proof and for everyone else a thorough outline for the following theorem.

Theorem 6 *For $p > 2$ the isospiking bifurcation points α_k, ω_k exist for the return map S_μ-family for large k with the properties that the end of the k-isospiking bifurcation point is approximately harmonic, $\omega_k = 1/k + o(1/k)$, and the non-isospiking intervals are exponentially narrow, $|\alpha_{k+1} - \omega_k| \sim \exp(-k)$. Here $o(1/k)$ means higher order of $1/k$.*

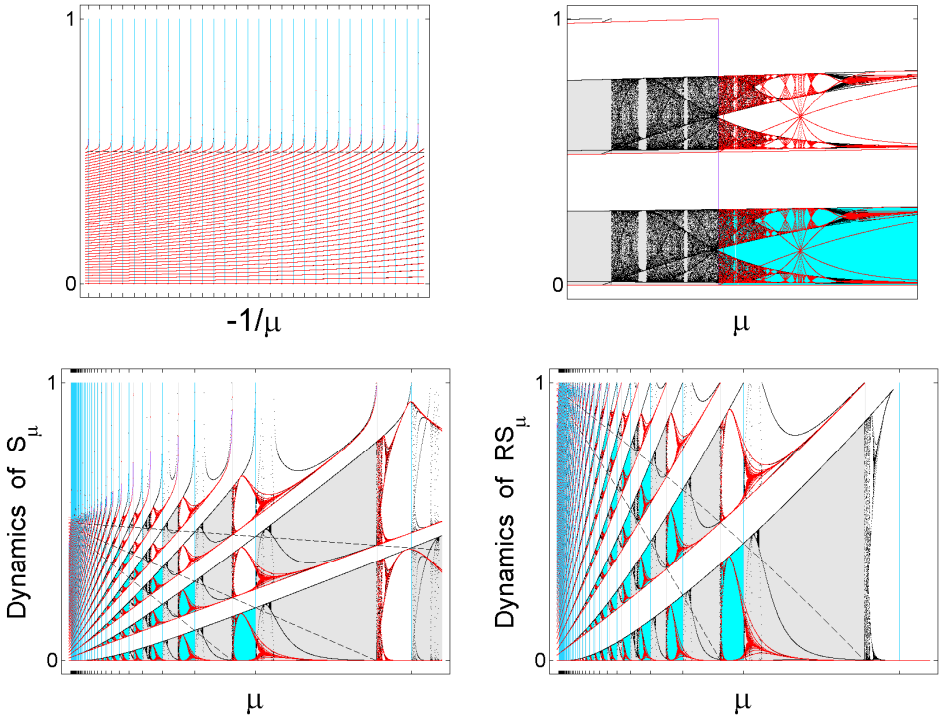


Figure 3.7: A numerical simulation for the S-family. Bottom left: the spike bifurcation diagram with the lower and upper corridor boundary iterates: m_k (black) and M_k (red). The corridor is colored cyan for isospiking and gray for non-isospiking. Dashed curves are c_0, c_{-1}, c_{-2}. Bottom right: The bifurcation diagram for the return map RS of the S-family. Top right: a zoom-in around ω_1 for RS and ω_2 for S. Top left: in the scale $-1/\mu$, $-1/\omega_k$ are spaced equally, a property defining the harmonic sequence ω_k. The non-isospiking intervals appear to be lines, i.e. $\alpha_{k+1} = \omega_k$.

Fig.3.7 gives a numerical simulation of the S-family consistent with the theorem. One more takeaway message from the simulation is that the immediate and transient property in bursting spikes is uncorrelated with the long-term dynamics of the system. The system can be in a chaotic but isospiking state or a chaotic non-isospiking state. This means if the immediate state of a neuron is what the neural code is about, the asymptotical state of the neuron, chaotic or quasi-periodic or periodic, is not important. All that matters is whether or not the spiking state is regular, isospiking always or misfiring occasionally.

As the last observation of this section that leads naturally to the next section, we note that a harmonic bifurcation sequence, such as $\omega_n = C/n$ for some near constant C, differs from all exponential bifurcation sequences, such as the period-doubling cascade, in a fundamental way. The so-call ratio test for convergence converges to 1 for the harmonic sequence and to the exponential ratio $\delta \neq 1$ for the latter:

$$\lim_{n\to\infty} \frac{\omega_n - \omega_{n-1}}{\omega_{n+1} - \omega_n} = 1.$$

In the Feigenbaum-Coullet-Tresser scenario the limit of the ratio test is their universal constant $\delta = 4.6692....$ We ask if the first natural number as the limit of the ratio test of a harmonic sequence is also a universal constant in some analogous setup.

3.4 Universal Number 1

It is the time to let the real player in who has been knocking the door for the last two sections. The renormalization operator. When Feigenbaum, Coullet, and Tresser used it, it was mysterious. We know now their renormalization is simply finding return maps of the logistic-like maps and then to scale them up to the unit interval to have a common framework. We too used a return map of the spike map family in the previous section to show that isospiking bifurcations scale along the harmonic sequence. Formalizing the process of finding return maps of return maps of neuronal models and scaling them to the unit interval results in our renormalization operator, R. The renormalizable space, $\mathcal{U}$, is a set of maps on the unit interval. Conceptually there should be little restriction on the set because after all one can define or find a return map on pretty much any map of the unit interval. To limit the scope we will restrict the set for now to maps similar to the return maps of neuronal models we introduced in the previous sections.

3.4.1 Spike Map Renormalization

More specifically, a map f is from $\mathcal{U}$ if these conditions hold. One, it has exactly one discontinuity point c in the unit interval minus the left end point 0. If it is continuous in the entire interval, then c is taken to be the right end point $c = 1$. Two, to the left side of c the function f is continuous, monotone increasing, not below the diagonal line. Three, the maximum of f on the interval right of c, $(c, 1]$, is no greater c. Four, the value of the function at the left end point is no

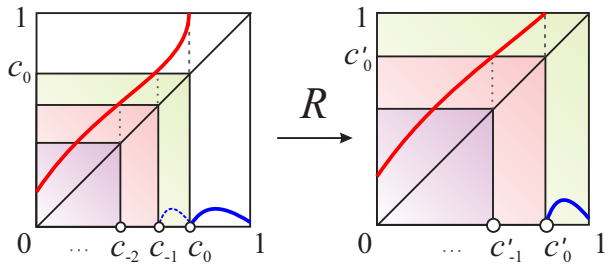


Figure 3.8: Renormalization of spike return map. Here c'_0 is the discontinuity point of the renormalized image. It is defined to be c_{-1}/c_0.

greater than c, $f(0) \leq c$. It is this condition that is key to f's renormalizibility. Last, it is not important in which interval f is defined at c. For convenience we can just assume c is included in the left interval.

The renomalization R is defined as follows. Since f from $\mathcal{U}$ is monotone to the left of its discontinuity c and its minimal value in the left interval is less or equal to c, the critical point c must have a preimage c_{-1} in $[0, c]$. This property permits a return map, rf, to be defined on $[0, c]$. Because f maps $[0, c_{-1}]$ inside $[0, c]$, we keep the map itself: $rf(x) = f(x)$ for $0 \leq x \leq c_{-1}$. For x from $(c_{-1}, c]$, f maps it out into $(c, 1]$, and then back into $[0, c]$. So we define $rf(x) = f(f(x)) = f^2(x)$ for $c_{-1} < x \leq c$. Finally we scale the return map's interval, $[0, c]$, to $[0, 1]$. This completes the renormalization of f:

$$R(f)(x) = \frac{1}{c} rf(cx), \quad \text{for } 0 \leq x \leq 1.$$

See Fig.3.8 for an illustration. The renormalized image $R(f)$ may or may be renormalizable depending on if $R(f)$ belongs to the renormalizable space $\mathcal{U}$. If it does, then the second backward iterate c_{-2} of the discontinuity point c must exist, and the return map of the return map of f is a return map of f. Without the scaling, it is the return map of f in the interval $[0, c_{-1}]$, of which it is the map f in the left interval $[0, c_{-2}]$ and it is the 3rd iterate f^3 in the right interval $(c_{-2}, c_{-1}]$. Scaling this return map by c_{-1} gives $R^2(f)$. This description is easily generalized for $R^k(f)$ if it exists. For which the return map before scaling is the $(k+1)$st iterate f^{k+1} in the right interval $(c_{-k}, c_{-k+1}]$ and just f in the left interval $[0, c_{-k}]$. Namely, if $R^{k-1}(f)$ is renormalizable, then

$$R^k(f)(x) = \begin{cases} \frac{1}{c_{-k+1}} f(c_{-k+1}x), & 0 \leq x \leq \frac{c_{-k}}{c_{-k+1}} \\ \frac{1}{c_{-k+1}} f^{k+1}(c_{-k+1}x), & \frac{c_{-k}}{c_{-k+1}} < x \leq 1, \end{cases} \tag{3.5}$$

where c_{-i} is related to c with $f^i(c_{-i}) = c$ for $i = 0, 1, 2, \ldots, k$. Again, without the scaling by c_{-k+1} the map is simply the return map of f in the interval $[0, c_{-k+1}]$.

3.4.2 Harmonic Invariance

A couple of simple exercises to warm up the idea. Take the simple ψ-family for example. At the limit, $\psi_0(x) = x$ is the identity map. It is in $\mathcal{U}$ and its 'discontinuity' point is the right end point 1. The entire interval $[0, 1]$ is the left interval and the right interval is empty set. So $R(\psi_0) = \psi_0$, a fixed point of R ! For positive μ, $c = 1 - \mu$ and $\psi_\mu(0) = \mu$. So ψ_μ is renormalizable if $\psi_\mu(0) = \mu \le c = 1 - \mu$, i.e. $\mu \le 1/2$. Since $\psi(x) = 0$ in the right interval, all its return maps are zero in their right intervals. Since ψ is a line in the left interval, any scalar scaling preserves the slope of the line, i.e. a scalar-scaled line is a line of the same slope. With the scaling by $1 - \mu$, the renormalization is

$$R(\psi_\mu) = \psi_{\mu/(1-\mu)}.$$

That is, R maps the ψ-family into itself, or the curve ψ_μ in $\mathcal{U}$ is invariant under R.

This is intriguing. One may become curious at this point to ask what about the isospiking bifurcation members ψ_{μ_k} of the family? They are onto themselves because for $\mu_k = 1/(k+1)$ we have $\mu_k/(1-\mu_k) = \mu_{k-1}$! That is, the harmonic sequence ψ_{μ_k} is an orbit of R, going in backward.

As the image $\psi_{\mu/(1-\mu)}$ seems to be 'farther away' from the fixed point ψ_0 than the preimage ψ_μ is because $\mu/(1-\mu) > \mu$, as a curve in the space $\mathcal{U}$ the ψ-family behaves like an unstable manifold of the fixed point ψ_0. The contracting backward orbit $\{\psi_{1/k}\}$ only deepens the impression. By most people's intuition the parameter values should be a measure of distance between the family members. If so the break-away rate from the fixed point along the unstable manifold should be gauged by the following ratio test,

$$r = \lim_{k\to\infty} \frac{\mu_k - \mu_{k-1}}{\mu_{k+1} - \mu_k}.$$

But the test limit

$$\lim_{k\to\infty} \frac{\mu_k - \mu_{k-1}}{\mu_{k+1} - \mu_k} = \lim_{k\to\infty} \frac{1/(k+1) - 1/k}{1/(k+2) - 1/(k+1)} = 1,$$

is not a number greater 1 as we normally expect. That is, this preliminary exploration suggests that the ψ-family is not the usually hyperbolic type of

unstable manifolds. Instead, it is a center manifold that is expanding, i.e. an unstable-center manifold.

3.4.3 Distance Between Spike Maps

Intuition is no substitution for proof. So the problem turns into a question about how to measure the distance between elements in the renormalizable space $\mathcal{U}$. As a starter we certainly would like the ψ-family to be a continuous family in the parameter μ with respect to whatever distance measure we will come up for $\mathcal{U}$. For f, g from $\mathcal{U}$, our immediate choice to try is the point-wise distance $|f(x) - g(x)|$. But this is not a good measure because for two arbitrarily close parameters μ_1 and μ_2 this point-wise distance between ψ_{μ_1} and ψ_{μ_2} is always 1 at one of the discontinuity points. Apparently this is the only deficiency of the point-wise distance we need to fix. We will instead use the average distance of the two in the unit interval. By elementary calculus the average distance is the area between the two graphs. We will denote it by

$$\|f - g\| = \int_0^1 |f(x) - g(x)|dx$$

and refer to it as the norm between f and g.

Let us try out a couple simple computations. First the distance of the ψ-family member ψ_μ to the fixed point ψ_0. The area between them consists of one parallelogram and one trapezoid. The parallelogram is over the left interval $[0, 1-\mu]$ with side length μ. So its area is $\mu(1-\mu)$. The trapezoid is over the right interval $[1-\mu, 1]$ between the diagonal line and the base. Its area is $\frac{1-\mu+1}{2}\mu$. Together we have

$$\|\psi_\mu - \psi_0\| = \mu\frac{4-3\mu}{2},$$

showing in fact the distance is a near constant factor of μ. We can do this type of simple calculations for any two members, $\psi_{\mu_1}, \psi_{\mu_2}$, of the family to have

$$\|\psi_{\mu_2} - \psi_{\mu_1}\| = |\mu_2 - \mu_1|\frac{4 - 3|\mu_2 - \mu_1| - 2\min\{\mu_1, \mu_2\}}{2}.$$

Again, it is proportional to the parameter difference $|\mu_2 - \mu_1|$ as we wanted. Applying the first formula to the backward harmonic orbit $\psi_{1/k}$ we have

$$\|\psi_{1/k} - \psi_0\| = \frac{1}{k}\frac{4-3(1/k)}{2}$$

going to zero monotonically in k. Applying the second formula to the same backward orbit we have the following ratio test limit

$$\lim_{k\to\infty}\frac{\|\psi_{1/k}-\psi_{1/(k-1)}\|}{\|\psi_{1/(k+1)}-\psi_{1/k}\|}=\lim_{k\to\infty}\frac{1/k-1/(k-1)}{1/(k+1)-1/k}=1.$$

These results can be generalized to any μ-harmonic orbit: $\{\psi_{\mu/(1+k\mu)}\}$ for any fixed value μ as $R(\psi_{\mu/(1+k\mu)})=\psi_{\mu/(1+(k-1)\mu)}$ since $\mu/(1+k\mu)/(1-\mu/(1+k\mu))=\mu/(1+(k-1)\mu)$. That is, the ψ-family is in fact a weakly expanding center-manifold.

3.4.4 1 Is An Eigenvalue

We will push our luck one more time. The following simple estimate proves that 1 is in fact an eigenvalue of R at the fixed point ψ_0:

$$\begin{aligned}\|R(\psi_\mu)-R(\psi_0)-1\cdot(\psi_\mu-\psi_0)\| &= \|\psi_{\mu/(1-\mu)}-\psi_\mu\|\\ &= \Big(\frac{\mu}{1-\mu}-\mu\Big)\frac{4-3(\mu/(1-\mu)-\mu)-2\mu}{2}\\ &\sim \mu^2\sim\|\psi_\mu-\psi_0\|^2,\end{aligned}$$

showing that $\psi_\mu-\psi_0$ is an eigenvector of the the eigenvalue 1. All it takes is the distance formula for any two elements of the ψ-family. This is embarrassingly fortunate comparing to the Feigenbaum-Coullet-Tresser universality. It took about twenty years to have a conceptual proof for their universal number δ to be an eigenvalue of their renormalization.

3.4.5 0 Is Also An Eigenvalue

There is no reason to stop pushing our luck. Let us take a look at the full spike return map family S_μ. At $\mu=0$, we have $S_0(x)=x$ for $0\le x\le c$ and $S_0(x)=0$ for $c<x\le 1$. Thus, $R(S_0)=\mathrm{id}$ the identity map, i.e. $R(S_0)=\psi_0$. That is, in one iterate S_0 is onto the fixed point ψ_0, showing 0 is an eigenvalue of R:

$$\|R(S_0)-R(\psi_0)-0\cdot(S_0-\psi_0)\|=0\sim\|S_0-\psi_0\|^2.$$

One can actually show that $R^n(S_\mu)$ converges to the unstable-center-manifold ψ_μ in the following manner: for any arbitrarily small number $\varepsilon>0$ and any member, ψ_μ, of the ψ-family, there is a sufficient large integer k and a member, $S_{\mu'}$, of the S-family so that kth renormalization $R^k(S_{\mu'})$ of $S_{\mu'}$ is within ε-distance of ψ_μ:

$$\|R^k(S_{\mu'})-\psi_\mu\|<\varepsilon.$$

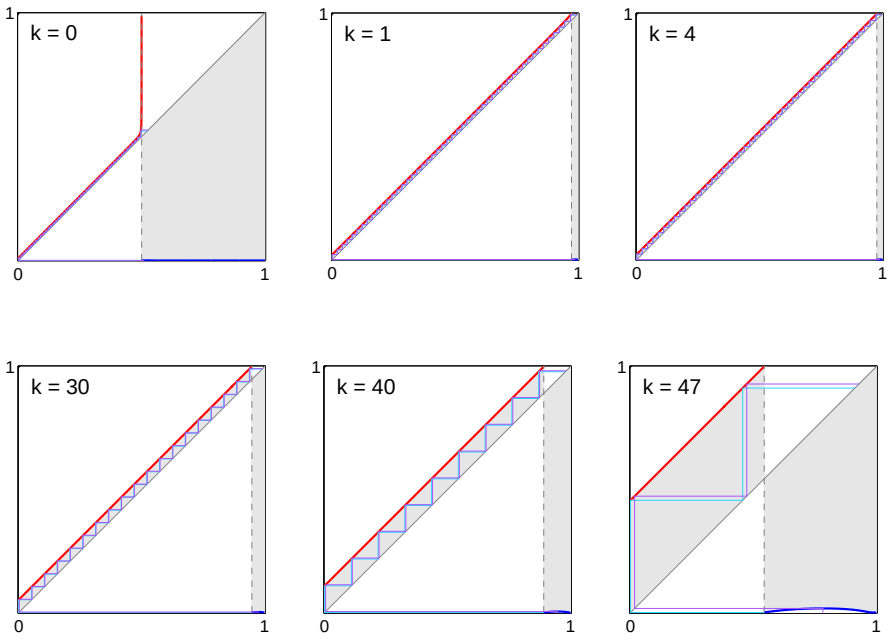


Figure 3.9: The shaded area is the distance between $R^k(S_\mu)$ and the fixed point ψ_0 for some μ which gives the map S_μ 49 isospikes because the 47th renormalization has 2 spikes per burst.

This result is the so-called λ-lemma. We will not dive into a proof for the S-family but to illustrate the idea for the s-family.

3.4.6 Weakly Hyperbolic Structure

The explanations start and end with the ψ-family. For each fixed μ of the ψ-family, we consider its μ-harmonic sequence $\mu' = \mu_k = \mu/(1+k\mu)$. Because it satisfies the μ-harmonic invariance $R(\psi_{\mu/(1+\mu)}) = \psi_\mu$ we have $R^k(\psi_{\mu_k}) = \psi_\mu$. Recall also that ψ_μ and s_μ are identical in the active interval $[0, 1-\mu]$ with exactly the shared back iterates $c_{-i} = 1-(i+1)\mu_k$ of the discontinuity point $c_0 = 1-\mu_k$ for $\mu = \mu_k$. Thus $R^k(s_{\mu_k})$ differs from $R^k(\psi_{\mu_k}) = \psi_\mu$ only in the silent phase interval $(1-\mu, 1]$ with its maximum the scaled maximum value $M(\mu_k)$ of s_{μ_k}. Because of the formula (3.5) for the kth iteration of R, the scaling factor is $c_{-k+1} = 1-k\mu_k = \frac{1}{1+k\mu}$ and the new maximal value is $M(\mu_k)/c_{-k+1} = (1+k\mu)M(\mu_k)$. Because $M(\mu_k)$ decays exponentially in k we can always find a

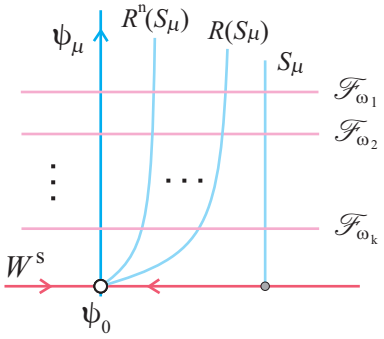


Figure 3.10: Invariant foliations near the non-hyperbolic fixed point ψ_0.

large k so that the new maximum is no greater than ε, $M(\mu_k)/c_{-k+1} < \varepsilon$. As a result $\|R^k(s_{\mu_k}) - \psi_\mu\| < \varepsilon$ as desired.

A proof for the S-family is very similar because the return map rS_μ is approximated by s_μ. However we will use a numerical simulation, Fig.3.9, instead for an illustration to spare the details. It shows that for any ψ_μ one can find μ_k so that $R^k(S_{\mu_k})$ is arbitrarily close to ψ_μ. It also shows the weakly hyperbolicity of the fixed point ψ_0 — the renormalization iterate $R^k(S_\mu)$ moves closer to the fixed point before it moves away from the fixed point.

What about foliation transversal to the unstable-center manifold ψ-family? Here is one. Let $\mathcal{F}_{\omega_k}$ denote the subset of renormalizable maps f so that $m_k = c_0$ for f's lower spike corridor boundary orbit m_k and f's discontinuity point c_0, i.e. the condition for the end of k-isospiking point for the ψ, s, and S families. Hence, for f from $\mathcal{F}_{\omega_k}$ if and only if $R(f)$ is in $\mathcal{F}_{\omega_{k-1}}$ because the application of R to its argument f simply takes out one spike count of f, showing

$$R(\mathcal{F}_{\omega_k}) \subset \mathcal{F}_{\omega_{k-1}}.$$

As for the ψ, s, S families they all intersect the $\mathcal{F}$-foliation transversely at the isospiking bifurcation points ω_k. See Fig.3.10

We list all the easy and fast goodies in the following theorem,

Theorem 7 *The first natural number 1 is a weakly expanding eigenvalue of the renormalization map R and the ψ-family is a corresponding invariant unstable-center-manifold. Also, the s-family and the S-family each converges to the ψ-family under the iteration of R. In addition, the family of sets $\{\mathcal{F}_{\omega_k}\}$ is an invariant foliation of R that is transversal to the ψ, s, S families.*

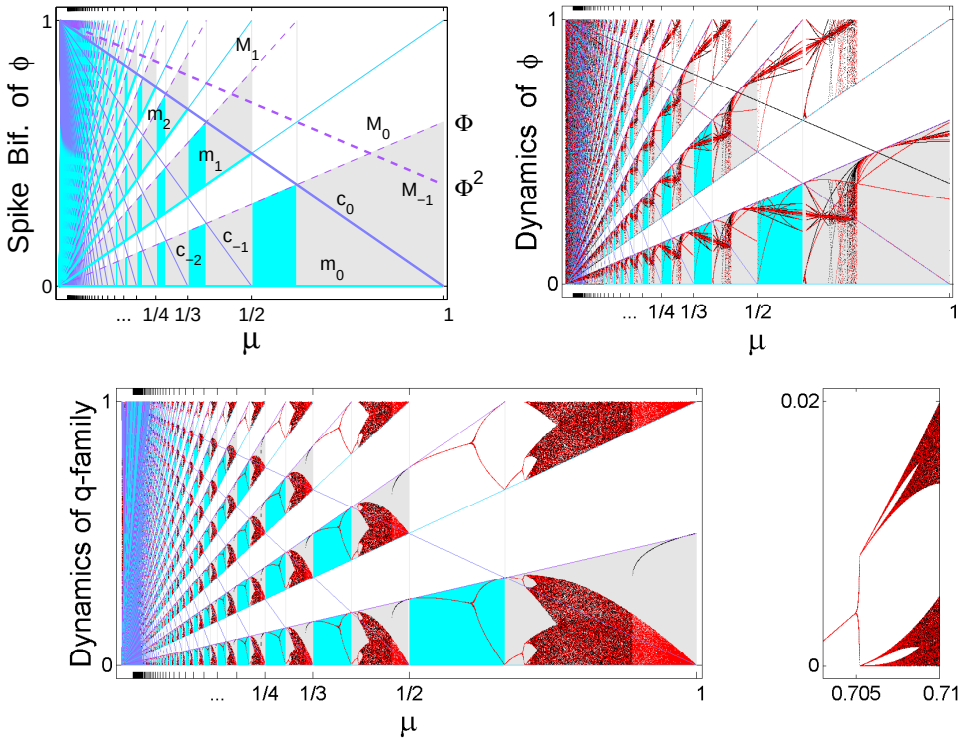


Figure 3.11: Top-left: The M_{-1} line splits the triangle above the c_0 line in golden ratio, so does the M_0 line for the triangle below the m_1 line. Other golden ratio objects are abundant in this this diagram. Top-right: The M-orbit (black) and the m-orbit (red). The diagrams are invariant under the renormalization R. If you the $\mu = 1/2$ line to $\mu = 1$ and the c_0 line to horizontal line $y = 1$ (i.e. the c_1 line), you get the exactly same diagrams. Bottom: The bifurcation diagram of the q-family. The zoom-in plot on the right shows a seemingly period-doubling cascade is interrupted prematurely.

3.4.7 More Harmonic Invariance Families

Before closing we want to point out that the ψ-family is not the only center-manifold of the fixed point. Here we give two more families. They are the specific cases of the general form

$$f_\mu(x) = \begin{cases} \mu + x, & 0 \le x < 1 - \mu \\ g_\mu(x), & 1 - \mu \le x \le 1, \end{cases}$$

for which the functional form in the left interval is exactly the same as the s-family and the functional form g in the right interval satisfies the following renormalization-invariance condition:

$$\frac{1}{c_0} f_\mu^2(c_0 x) = \frac{1}{c_0} g_\mu(\mu + c_0 x) = g_{\mu/c_0}(x)$$

with $c_0 = 1 - \mu$.

The first family is a special form of the s-family for which we set the parameter $d = 0$. Since the function form g_μ is $g_\mu(x) = a(x - c_0)\exp(-(x - c_0)/(b\mu))$ with $c_0 = 1 - \mu$, it is straightforward to check that $g_\mu(\mu + c_0 x)/c_0 = g_{\mu/c_0}(x)$. This holds for any choice in a and b. We will call it the ϕ-family for this parameter choice $b = \Phi^2$, $a = (\Phi + 1)\exp(1)$ with $\Phi = (\sqrt{5} - 1)/2$ being the golden ratio ($\Phi^2 = 1 - \Phi$),

$$\phi_\mu(x) = \begin{cases} \mu + x, & 0 \le x < 1 - \mu \\ (\Phi + 1)(x - 1 + \mu)e^{1-(x-1+\mu)/(\Phi^2\mu)}, & 1 - \mu \le x \le 1. \end{cases}$$

When $\mu = 1$, the maximum point $b = \Phi^2$ and the maximum value $abe^{-1} = (\Phi + 1)\Phi^2 = \Phi$ is a golden ratio split. The top-left figure of Fig.3.11 shows the isospiking bifurcation diagram of the family and the top-right figure shows the M-orbit and the m-orbit which define the spike corridor. These diagrams are invariant under R in the following way. For any vertical line μ find its harmonic predecessor $\mu/(1+\mu)$. The dynamical image of $R(\phi_{\mu/(1+\mu)})$ is exactly the dynamical image of ϕ_μ.

The second family of f_μ has the simplest nonlinear form for the right interval $g_\mu(x) = \lambda(x - c_0)(1 - x)/\mu$ with $c_0 = 1 - \mu$. We call it the q-family,

$$q_\mu(x) = \begin{cases} \mu + x, & 0 \le x < 1 - \mu \\ \frac{\lambda}{\mu}(x - (1 - \mu))(1 - x), & 1 - \mu \le x \le 1. \end{cases}$$

Again, it is straightforward to check the invariance condition: $g_\mu(\mu + c_0 x)/c_0 = g_{\mu/c_0}(x)$. The bottom figure of Fig.3.11 shows the asymptotic orbits for the upper and lower corridor boundary points for the parameter value $\lambda = 1/2$. For $\mu = 1$, q_1 is the quadratic map $x(1-x)/2$ for which the maximum point $x = 1/2$ is supper stable. It dynamics is far from chaos. But the dynamics of the spike family can be chaotic at every small scale of μ.

3.5 Universal Are All Numbers

The dynamical systems world would be extremely mysterious if the Feigenbaum-Coullet-Tresser number $\delta = 4.6692...$ were the only universal number. We now

know it has a companion. But the dynamical system world would still be very peculiar if δ and 1 are the only universal numbers. What about the number 2 or the number π? We will show in fact all positive numbers are universal by the same criteria as we have applied to the two universal numbers known so far. One criterion requires that a universal number be the limit of a ratio test on the bifurcation sequence of a dynamical system. The other criterion requires that the number is an eigenvalue of a renormalization map on the dynamical system.

3.5.1 Exponential Invariance Families

Let us first do the easier part of the two criteria because we only need to continue on with the same renormalization map R on the spike renormalizable space $\mathcal{U}$. For any real number $\lambda > 1$ we consider the following u-family of maps,

$$u_{\lambda,\mu}(x) = \begin{cases} \mu + \lambda x, & 0 \le x \le \frac{1-\mu}{\lambda} \\ 0, & \frac{1-\mu}{\lambda} < x \le 1, \end{cases} \tag{3.6}$$

for $0 \le \mu \le 1$. (When parameter λ is not part of an analysis we will drop it for a streamlined notation u_μ.) At the right end of the parameter μ, we have

$$u_0(x) = \lambda x \text{ for } 0 \le x \le \tfrac{1}{\lambda} \text{ and } u_0(x) = 0 \text{ for } \tfrac{1}{\lambda} < x \le 1.$$

It is the line through the origin with the slope $\lambda > 0$ in the left interval $[0, 1/\lambda]$ and the horizontal line in the right interval $(1/\lambda, 1]$. The critical point is $c = 1/\lambda$. Its backward iterates are $c_{-k} = 1/\lambda^{k+1}$ with $c_0 = c$. Since they form a decreasing infinite sequence, the map u_0 is indefinitely renormalizable. Since u_0 is zero in the right interval, its renormalized image is zero in the right interval. Since u_0 is a line, its renormalization is also a line in the left interval and with the same slope because scalar scaling preserves all lines through the origin. Therefore u_0 is a fixed point of R, $R(u_0) = u_0$, for any fixed parameter $\lambda \ge 1$.

In fact, for $\lambda = 1$, the u-family is the ψ-family we studied in the last section. As a result pretty much all analyses applied to ψ_μ apply to u-family as well. For example, for $\mu > 0$, u_μ is the u_0 lifted by μ and then clipped by the top of the unit box at the new critical point $c_0 = (1 - \mu)/\lambda$. By exactly the same reasons as above and as for the ψ-family, the u-family is also invariant with R:

$$R(u_\mu) = u_{\mu/c_0} = u_{\lambda\mu/(1-\mu)}.$$

Unlike the ψ-family which is a center manifold of the identity map ψ_0, the u-family is the unstable manifold of the fixed point u_0. This is backed up by the following computations.

3.5.2 Every Positive Number Is An Eigenvalue

First, we can find the distance $\|u_\mu - u_0\|$ of u_μ to the fixed point u_0. It consists of two areas between the graphs — the area of the parallelogram between u_μ and u_0 over the interval $[0, (1-\mu)/\lambda]$ and the area of the trapezoid between u_μ and u_0 over the interval $[(1-\mu)/\lambda, 1/\lambda]$. A simple tabulation of the areas gives

$$\|u_\mu - u_0\| = \frac{\mu}{\lambda}\frac{4-3\mu}{2},$$

the same distance between ψ_μ and ψ_0 scaled down by λ. It is an increasing function of μ for $0 \le \mu \le \frac{1}{3}$, i.e., u_μ moves away from u_0 in that parameter interval.

Second, we can show λ is likely an eigenvalue by calculating the rate of expansion along the expanding u-family. It is by the distance formula above and the identity $R(u_\mu) = u_{\lambda\mu/(1-\mu)}$ that we have

$$\lim_{\mu\to 0}\frac{\|R(u_\mu) - R(u_0)\|}{\|u_\mu - u_0\|} = \lim_{\mu\to 0}\frac{\frac{\mu}{1-\mu}}{\frac{\mu}{\lambda}}\frac{4-3\frac{\lambda\mu}{1-\mu}}{4-3\mu} = \lambda.$$

To show that λ in fact is an eigenvalue we need some not so trivial computations. It involves the so-called generalized function, namely the delta distribution function δ_a at a point $x = a$. It starts with finding the so-called linearization of R along the unstable manifold u_μ at the fixed point u_0. It is the limit of the following quotient

$$\lim_{\mu\to 0}\frac{R(u_\mu) - R(u_0)}{\|u_\mu - u_0\|} = \lim_{\mu\to 0}\frac{u_{\lambda\mu/(1-\mu)} - u_0}{\frac{\mu}{\lambda}\frac{4-3\mu}{2}}.$$

If the limit exists then it must be a number (a complex number in general) times a function whose norm or length is the unit 1. Without showing the detailed computations for the limit about, let Tu denote the following function

$$Tu(x) = -\frac{1}{2}\delta_{1/\lambda}(x) + \frac{1}{2}v_\lambda(x) \text{ with } v_\lambda(x) = \begin{cases}\lambda, & 0 \le x < \frac{1}{\lambda}\\ 0, & \frac{1}{\lambda} < x \le 1,\end{cases}$$

where $\delta_{1/\lambda}(x)$ is the δ-function with the norm $\|\delta_a\| = 1$ for any a and this formula $\|Tu\| = \|-\frac{1}{2}\delta_{1/\lambda}\| + \|v_\lambda\|$ to calculate Tu's norm. Then those omitted computations would have shown that

$$\lim_{\mu\to 0}\frac{R(u_\mu) - R(u_0)}{\|u_\mu - u_0\|} = \lambda Tu.$$

Because v_λ points in the same direction as $R(u_\mu) - R(u_0)$ and $\|Tu\| = 1$, it follows that λ is indeed the eigenvalue of the corresponding unit eigenvector Tu.

Thus, the only question remains is if the u-family is the return map family for a system of differential equations. The unimodal families can be the return maps of a lot of dynamical systems, e.g. the food-chain models from the previous chapter. The S-family are the return maps of the neuronal models for spike generation. What dynamical systems have the u-family as their return maps and what type of bifurcations will the family be about?

3.5.3 Graded Fast-Slow Systems

Up until this point we have used fast-slow systems constructed by singular perturbations to gain valuable insights into chaos generation in models. We used them as proxies. We will call them up one more time, but this time use it on their own right because we do not have physical models for the u-family. That is the key difference between the u-family and the other two families.

The system is Eq.(1.8) for the Shilnikov homoclinic orbit to an expanding spiral equilibrium point. It is a singularly perturbed system with one singular parameter ϵ,

$$\dot{x} = f,\ \dot{y} = g,\ \epsilon\dot{z} = h.$$

Here the zeroes of h define the Z-switch as usual, with the top and bottom branches approximated by the plane $z = \pm 2$. The xy-equations are the ϵ-slow system. On the bottom branch $z \sim -2$ of the switch, we will use one time scale for both x and y variables. In fact, we will use a linear, logarithmic spiral for the reduced slow system, which in the polar coordinate is simply,

$$\dot{r} = \alpha r,\quad \dot{\theta} = \beta.$$

For each θ_0, it takes $t = 2\pi/\beta$ amount of time to return to the radial line $\theta = \theta_0$. Its radial displacement on the return line from the center is $r = r_0 e^{\alpha t} = r_0 e^{2\pi\alpha/\beta}$. That is, with $\lambda = e^{2\pi\alpha/\beta}$ the spiral from the center expands outward λ-times away. For the system considered we will use $(x, y) = (2, 0)$ for the spiral center.

On the top branch $z \sim 2$, we will further split the time scale for the reduced slow subsystem. This is unlike all multi-scales systems we have constructed and encountered so far. We can either use a second singular parameter ζ but require, $0 < \epsilon \ll \zeta \ll 1$. That is, the second parameter is small but not smaller than the first in magnitude. Or we simply use $\zeta = \epsilon^p$ for some constant $0 < p < 1$. We will adopt the latter. Thus, up to a common scalar multiple, the ϵ-slow system takes the following ϵ^p-form on the top branch of the fast switch,

$$\dot{x} = \epsilon^p,\ \dot{y} = b + \frac{c-b}{-2-a}(x-a) - y.$$

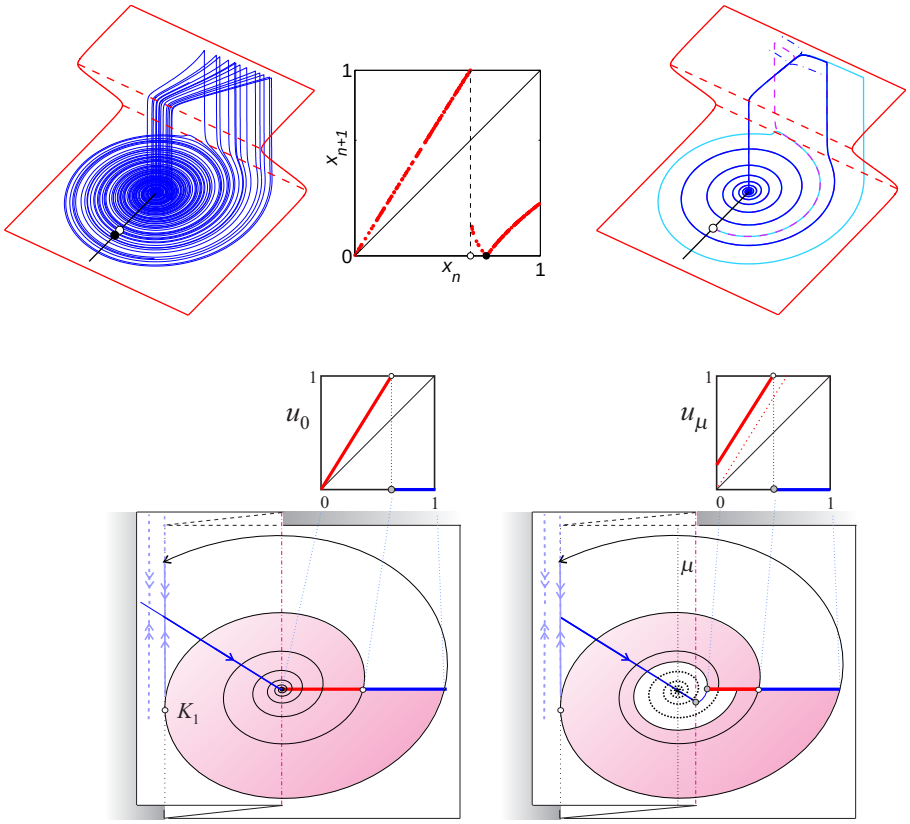


Figure 3.12: Top-left and center: Parameter values for the Shilnikov attractor of the system: $a = 2,\ b = -2,\ c = 1,\ m = 0.923,\ \alpha = 0.3829,\ \beta = 5,\ d = 0.08,\ \zeta = 4,\ p = 0,\ \epsilon = 0.001$. Top-right: The same parameters except for $\zeta = 1,\ p = 1/3,\ b = 0,\ m = 0.905$. The particular choice in α and β results a gold ratio spiral with $\lambda = 1 + \Phi$. Even though ϵ is at the singular value, all flows are channeled into one trickle on the top branch. Bottom: The return map u-family from the graded fast-slow system. The fall of the top flow misses the spiral center by an amount proportional to μ, the first return of the new left end point of the return map.

It is a linear but singularly perturbed system. Variable x is the ϵ^p-slow variable and variable y is the ϵ^p-fast variable. The corresponding ϵ^p-slow manifold is the

y-nullcline:

$$y = b + \frac{c-b}{-2-a}(x-a),$$

which goes through the points (a, b) and $(-2, c)$, completely controllable by parameters a, b, c. Thus, at the singular limit $\epsilon^p = 0$, all the flows on the top branch are squeezed onto the ϵ^p-slow manifold and then led rightward (because $\dot{x} > 0$) toward the turning edge of the switch. With the top turning edge running along the line $x = 2$, and the parameter choice $a = 2, b = 0$, all the ϵ^p-singular orbits are dumped down below onto the spiral center $(2, 0)$ on the bottom branch of the switch.

For convenience we recall the system below:

$$\begin{cases} \dot{x} = \zeta\epsilon^p(z+2) + (2-z)[\alpha(x-2)h - \beta y] \\ \dot{y} = (z+2)[b + \frac{c-b}{-2-a}(x-a) - y] + (2-z)[\beta(x-2) + \alpha y h] \\ \epsilon\dot{z} = (2^2 - z^2)[z + 2 - m(x+2)] - dz \end{cases}$$

and refer to it as a ϵ^p-graded fast-slow system. Fig.3.12 gives a numerical simulation of the system for a regular ζ value with $p = 0$ and for a ϵ-graded value with $p = 1/3$ and $\zeta = 1$.

Without the grading ($p = 0$ and ζ regular), the top slow flows fan out toward the top turning edge like a waterfall. The resulting dynamics is the Shilnikov chaos from a homoclinic orbit to a saddle-fucus equilibrium point. The presence of a K-fold turning fold point splits the reduced slow flow on the bottom branch in two directions, one going up to the top branch, the other staying on the bottom for another round of spiraling. It creates a discontinuity point for its return map on any radial line through the spiral center. In the right interval of the returning line ($\theta = 0$), the corresponding homoclinic point is a critical minimum, being the limit of the K-fold junction point if the waterfall misses the spiral center. With ϵ^p-graded, the top waterfall is pinched to one line. The corresponding return map on the right interval is the flat line zero.

3.5.4 Return Map Of Graded Fast-Slow Shilnikov Orbit

Fig.3.12 also gives a derivation of the return map on the line $x \geq 2$ (i.e. $\theta = 0$). To the left of the point corresponding to the K-fold turning point K_1, the map is exactly $\mu + \lambda x$ with $\lambda = e^{2\pi\alpha/\beta}$ being the logarithmic spiral ratio. The linearity in the return variable x is because the reduced slow system on the bottom branch is a linear spiral. To the right of the discontinuity point, the map is flat-lined at zero because the top ϵ^p-graded slow system reduces all flows to the 1-dimensional ϵ^p-slow manifold. The offset parameter μ is proportional

to the distance between the falling slow-manifold orbit and the spiral center the orbit missed. The return map is of the u-family.

Like the ψ-family, the u-family maps go through a sequence of bifurcations as μ decreases to zero. Unlike the ψ-family which undergoes the isospiking bifurcations the u-family undergoes changes in the number of spirals on the bottom branch of the switch. The number of spirals increase from 1 to 2, to 3, and so on as μ goes down to zero. The number reaches infinity when the orbit becomes the homoclinic orbit of the spiral center.

The bifurcation point, μ_k, at which the singular orbit changes from having k spirals to $k+1$ spirals is determined by the bifurcation equation

$$m_k = c_0,$$

the same as for the isospiking bifurcation point ω_k. Since $m_0 = 0$ and $m_k = u^k_{\lambda,\mu}(m_0) = \mu + \mu\lambda + \ldots \mu\lambda^{k-1} = \mu(\lambda^k - 1)/(\lambda - 1)$, we have from the equation $m_k = c_0 = (1-\mu)/\lambda$ the solution

$$\mu_k = \frac{\lambda - 1}{\lambda^{k+1} - 1},$$

approaching 0 at the exponential rate $1/\lambda$ as k goes to infinity. As a result the ratio test limit is

$$\lim_{k\to\infty} \frac{\mu_k - \mu_{k-1}}{\mu_{k+1} - \mu_k} = \lim_{k\to\infty} \frac{\lambda^{k+2} - 1}{\lambda^{k+1} - \lambda} = \lambda.$$

It is what we expect for any geometric sequence of bifurcations with an exponentially contracting rate.

To summarize we have the following theorem,

Theorem 8 *Every number $\lambda > 1$ is a universal number in the sense that it is the eigenvalue of the renormalization map R along a family of return maps for a graded fast-slow system. The family of return maps is the unstable manifold of the eigenvalue λ to a fixed point of R. The family of return maps undergoes an infinitely sequence of bifurcations in the number of spirals as the graded fast-slow system approaches a singular homoclinic orbit of an unstable saddle-focus equilibrium point.*

Curiously the Feigenbaum-Coullet-Tresser number δ is re-created by the graded singular homoclinic bifurcation when $\lambda = \delta$ but the natural number 1 cannot. This is because at the limit $\lambda = 1$ the return maps for the graded singular homoclinic orbit of the spiral center is always the singleton S_0 map, i.e. $u_{1,\mu} = S_0$ for all μ. The graded singular homoclinic bifurcation ceases to provide a

new mechanism by which the universal number 1 can be re-created by the u-family even though as λ goes to 1 the limit $\lim u_{\lambda,\mu} = \psi_\mu$ exists. One more curious feature. Formally as $\lambda \to 1$ the bifurcation point μ_k does converge to the bifurcation point μ_k of the ψ-family. This is because for each fixed k, $\mu_k = (\lambda-1)/(\lambda^{k+1}-1) = 1/(1+\lambda+\lambda^2+\cdots+\lambda^k)$ which converges to $1/(k+1)$ as λ goes to 1. Everything seems to fit except that the λ-limit $u_{1,\mu} = S_0$ is a family of only one member.

Chapter 4
Big Chaos

The American mathematician Stephen Smale had an ambitious plan in the 1950s. He wanted to have some definitive thing to say about all differential equations. If he could prove what he envisioned, the field of dynamical systems would be done, all taken, no more virgin land to plow. Fortunate for everybody, himself included, that idea of his did not pan out. The person who would thoroughly upset his plan was no one but himself. His thwarted effort directly led to the first distilled chaotic map and fueled much of the theoretical pursuit of the chaos revolution in the last century.

Smale started out by considering all sufficiently smooth differential equations on closed but bounded manifolds. Given our current understanding that our visible universe is only finite his class of dynamical systems had rounded up pretty much all the usual suspects of mathematical models. He also started out by considering only equilibrium solutions of his differential equations. It is true that if a differential equation has an equilibrium point either the point is already isolated or one can give it a little shake to make it so. In a more precise term every differential equation of his class has arbitrarily close neighbors which have only finitely many equilibrium points. In this way his class of differential equations can be neatly divided up into subsets each of which has only a finitely many equilibrium points give-and-take those equations which are completely surrounded by these stable and regular diffizens.

Inspired by this neat picture Smale took the next natural step. He conjectured that the same result held for periodic solutions of his differential equations. In his own words many years later his conjecture was the same to say "chaos doesn't exist!". That of course was when things got interesting. He was promptly informed by Norman Levinson, an American mathematician, that

a counterexample already existed in the work of two British mathematicians, Mary Cartwright and John E. Littlewood. It was an differential equation about radio waves that has infinite many periodic orbits which persist for all neighboring equations. Smale worked day and night while vacationing on the beaches of Rio de Janeiro to translate the analysis of the trios to his topological way of understanding. It turned out that the root cause of the infinite many periodic orbits was the existence of a homoclinit orbit to a periodic solution, the same tangled mess which confounded Poincaré. He succeeded in his translation. The horseshoe map was discovered, and out came the most common definition of chaos.

4.1 What Is Chaos

His horseshoe map lives up to its moniker, see Fig.4.1. Start with a rectangle. You compress it in the vertical direction and stretch it in the horizontal direction. The order of these operations is not important. You can first compress then stretch or vice versa or simultaneously. After the contraction and expansion you fold it at the middle into a horseshoe and then superimpose it to the original rectangle. You get Smale's configuration if the horseshoe horizontally lies on its side and its two ends and bend stick out from the rectangle as shown. Embed the rectangle, the horseshoe in the plane $\mathbb{R}^2$ and denote the rectangle by Q and the point to point correspondence by h, we obtain his horseshoe map $h : Q \to \mathbb{R}^2$.

A lot of iterates do not stay in the rectangle. Whenever an iterate is captured by the white vertical strips of the rectangle its next iterate under h escapes the rectangle. The set of those points which stay inside Q indefinitely in both forward and backward iterations is referred to as the non-wandering set and is denoted by Λ. Under some mild conditions on how much the rectangle contracts and expands Smale demonstrated that the non-wandering set Λ contains uncounterablly many points. Among many more properties Smale emphasized the following three which many researchers now take to be a working definition of chaos. The first property is what upset his grand plan the most.

Theorem 9 *Under some typical and mild conditions on h the following properties hold for the non-wandering set Λ:*

1. *The set of periodic orbits is dense in Λ.*
2. *The map has a sensitive dependence on initial conditions.*
3. *There is a dense orbit in Λ.*

A set A is dense in another set B if for every point x from B and any small

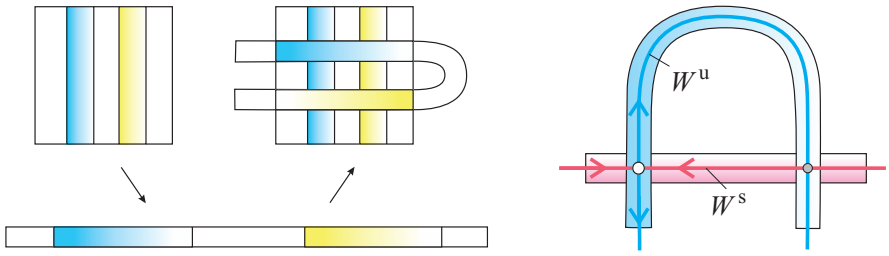


Figure 4.1: Left: A deconstruction of the horseshoe map into a sequence of four actions: contraction, expansion, fold, and return. The directionally shaded vertical bars are mapped to the directionally shaded horizontal strips. Right: Horseshoe arises from a homoclinit orbit to a fixed point. The stable and unstable manifolds of the fixed point provide with the contraction and expansion directions for the rectangle around the stable manifold. The intersection of the manifolds makes the compressed and elongated strip to fold and to return, resulting in a horseshoe map.

number $\epsilon > 0$ there is a point y from A whose distance to x is no greater than ϵ: $d(x, y) < \epsilon$. An orbit is dense in B if the set of iterates of the orbit $\gamma(x) = \{x_n : x_n = h^n(x), n = 0, 1, 2, 3, \dots\}$ is dense in B. A map f has a sensitive dependence on initial conditions in a set A (i.e. expansive) if there is a fixed number $\delta_0 > 0$ so that no matter how close two distinct points are to each other, $d(x, y) < \epsilon$, there is an integer $n > 0$ so that their nth iterates are at least δ_0-distance apart: $d(f^n(x), f^n(y)) > \delta_0$, a precise way to state the Lorenz butterfly effect.

The three properties of the theorem above consist of the working definition of chaos we will adopt for the remainder of the book. Truth be told or rather history be told, the word chaos was used first by Tien-Yien Li and James A. Yorke in the 1970s that caught on fire for the then emerging field and now burnt onto the doorway to the field. It came from the title of a paper they co-wrote. It proclaimed 'Period Three Implies Chaos'. Li was originally from Taiwan, and Yorke was Li's graduate advisor. The paper was the result of a small part of Li's graduate training. They did not think much of it and published it in a pedagogical journal for mathematicians. The paper is about the logistic type maps which says that if such a map has a period-three orbit then it must have orbits of all periods. Little did they know their result was published by the Ukraine mathematician Oleksandr Mikolaiovich Sharkovsky, a decade earlier,

under an impenetrable title, and in its complete form than their partial result. In fact, Sharkovsky rearranged the natural numbers in the following order now bearing this name:

$$3 \succ 5 \succ 7 \cdots \succ 2\cdot 3 \succ 2\cdot 5 \succ 2\cdot 7 \succ \cdots 2^k\cdot 3 \succ 2^k\cdot 5 \succ 2^k\cdot 7 \succ \cdots 2^3 \succ 2^2 \succ 2 \succ 1$$

and proved that if a map of the logistic family has a periodic orbit of period k then it must have a periodic orbit of period ℓ for every ℓ below k in his ordering. Li and Yorke proved the special $k = 3$ case. This piece of history just makes you wonder had they known Sharkovsky's work what would the field be called? Would it still be called chaos? Would a different name be able to stoke the same amount of creative fire, or capture the same level of public's imagination?

We will not give a proof of the horseshoe theorem above because there is an even simpler map and easier proof to use to illustrate the definition properties of chaos. Consistent with our singular perturbations approach, we can start the horseshoe map construction by compressing the rectangle vertically into a horizontal line segment. That is, we can imagine the vertical variable is a fast variable and the construction is done to the singular limit. In this simplified scenario, the line segment is stretched, folded, and returned to the line again. We can even further simplify the matter by assuming the reduced 1-dimensional map is piecewise linear, and when it is graphed out it looks like a tent depicted in Fig.4.2. Even the tent map is not the simplest. Its graph preserves the interval orientation of the left half interval but reverses the interval orientation of the right half interval, which can be a brain twister for the bookkeeping of its non-wandering set. So we will replace it by the simplest chaotic map, the baker's map, as depicted in the figure. It is the same as the tent map except that it preserves the orientation of the right half interval as well. The baker's map is not just a conjuring. In fact, the singular return map for the mockup Lorenz attractor introduced in a previous chapter is of the same type.

4.1.1 Itinerary Of Non-Wanderer

We will take the definition interval of the baker's map to be the unit interval $I = [0, 1]$. We will denote it by $B : I \to \mathbb{R}$:

$$B(x) = \begin{cases} 3x, & 0 \le x < 1/2 \\ 3x - 2, & 1/2 < x \le 1. \end{cases}$$

It is a line of slope 3 through the origin in the left half interval $[0, 1/2]$ and a line of slope 3 through the top-right corner point in the right half interval $[1/2, 1]$.

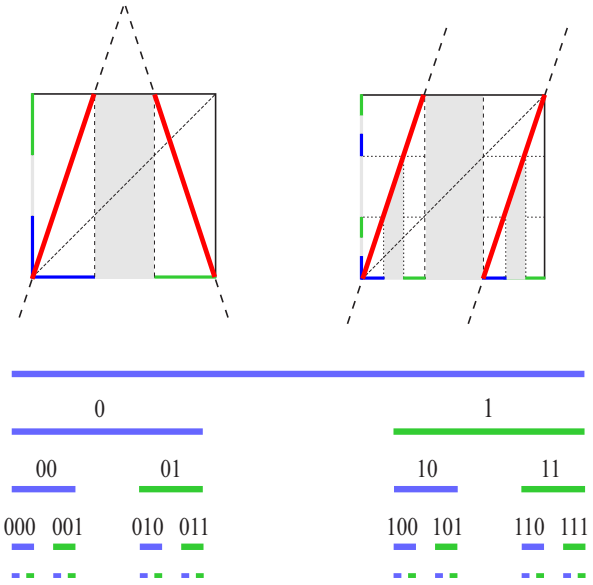


Figure 4.2: Top left: the tent map. Top right: the baker's map. Bottom: Cantor's middle third construction of the non-wandering set Λ of the baker's map in the unit interval $[0, 1]$: it is what remains after the open middle third interval is removed from every remaining interval ad infinitum.

We will work through some of the steps in the next section to see how simple or hard it is to demonstrate the simplest chaos.

To make the case of chaos for the baker's map it pretty much starts and ends with a good bookkeeping of points in the unit interval $I = [0, 1]$. Any point x of the interval is destined to one of two fates. Either one of its iterates $x_k = B^k(x)$ enters the open middle third interval $(1/3, 2/3)$ and then escape the unit square forever, or all iterates manage to avoid the mid-third trap: some or all fall inside the left third interval $I_0 = [0, 1/3]$ and some or all fall inside the right third interval $I_1 = [2/3, 1]$. Obviously the latter type points form the non-wandering set Λ and the former type form everything else but Λ, i.e. the compliment of Λ.

Of either type we can record the itinerary as the iterate travels in the interval. Whenever the iterate enters the escaping mid-third interval we stop recording the itinerary. If the iterate is in the left-third interval I_0 we record a 0 and if it is in the right-third interval I_1 we record a 1. Thus, the itinerary, denoted by $\rho(x)$, is either a finite sequence or an infinite sequence in itinerary symbol

0 and 1: $\rho(x) = s_0 s_1 \dots$ with $s_i = 0$ or 1. If a point is from the mid-third interval, its itinerary is an empty sequence $\rho(x) = \emptyset$. Otherwise, if its itinerary is a finite sequence $\rho(x) = s_0 s_1 \dots s_n$, then $x = x_0$ is from I_{s_0}, $x_1 = B(x)$ is in I_{s_1}, etc., and the nth iterate $x_n = B^n(x)$ is in I_{s_n} but the $(n+1)$th iterate $B^{n+1}(x)$ is in the mid-third interval at which time the itinerary recording stops. Such a point is from the compliment of the non-wandering set. So a point x is from the non-wandering set if and only if its itinerary is an infinite binary sequence $\rho(x) = s_0 s_1 \dots$. If we denote $\mathcal{B} = \{s : s = s_0 s_1 \dots,\ s_i = 0 \text{ or } 1\}$ the set of all infinite binary sequences then the itinerary map ρ is a map from the non-wandering set to the infinite binary sequence set $\mathcal{B}$. The questions are is the itinerary map ρ a one-to-one map, and is it a onto map?

The answers to both questions are yes. But to see the answers we need to get a better picture of the non-wandering set Λ. On the outset Λ is the fractal Cantor set which is obtained by first removing the open mid-third interval of the unit interval, then removing the open mid-third intervals of the left and right intervals I_0, I_1, and then removing the open mid-third interval of every remaining interval, and so on, recursively forever, see Fig.4.2. To see the itinerary map σ is one-to-one and onto, we go back to the itinerary idea once more.

4.1.2 Itinerary Of Set

This time around instead of individual point itineraries we define the itinerary of any set of the unit interval. The set itinerary $\rho(J)$ of a set J is again a binary sequence $\rho(J) = s_0 s_1 \dots s_n$ in 0s and 1s such that every point from J shares exactly the same destination itineraries from iteration 0 to iteration n and there must be at least two distinct points of J whose $(n+1)$th itineraries diverge. That is, the set itinerary is the maximally shared itinerary for all points of J. For example, the set itinerary $\rho(I)$ of the unit interval itself is the empty sequence because its points do not share a common itinerary. Clearly the left-third interval I_0 has its set itinerary to be at least one length long with the first itinerary $s_0 = 0$ because every point shares the same initial start. Because B expands I_0 to the full interval I, some points of I_0 are mapped to the left interval I_0, some to the right interval I_1, and so on. That is, its first iterates are everywhere, divergent. So $s_0 = 0$ is the only shared point itinerary of its points and $\rho(I_0) = s_0 = 0$, which is used for the interval's subscript, a zip code, so to speak, for the subset. The same explanation is for the right-third interval I_1 because of its set itinerary $\rho(I_1) = s_0 = 1$.

Let us now work out the set itineraries for the remaining subintervals at each

step of the Cantor mid-third construction. Take the I_0 interval for example. Its length is $1/3$. B expands I_0 three folds to the full interval $B(I_0) = I$. Because B is a straight line on I_0, the subinterval I_0 is divided into three equal parts, with the middle third interval mapped by B to the escaping mid-third of I. So it must be removed for not containing the non-wandering set. Two subintervals remaining, the left and the right third intervals of I_0. Because the left-third of I_0 is mapped by B onto I_0, all points receive one more shared itinerary 0. So we denote it by I_{00} with $B(I_{00}) = I_0$. In fact, $s_0s_1 = 00$ is the maximally shared point itinerary of I_{00} because the second iterate of the interval has diverging point itineraries as $B^2(I_{00}) = B(I_0) = I$. So $\rho(I_{00}) = 00$ as anticipated. Exactly the same arguments apply to the right-third interval of I_0, whose set itinerary is 01 and thus is denoted by I_{01}, all is because $B(I_{01}) = I_1$. The same applies to the subinterval I_1 whose left-third and right-third subintervals are I_{10}, I_{11} respectively, with their set itineraries being their subscript addresses. These four subintervals, each of length $1/3^2$, are the only subsets whose set itineraries take up all the binary sequences of length 2.

The same method of bookkeeping can be used for all intervals from the Cantor mid-third construction. More specifically, let us assume we have already constructed the 2^n many closed subintervals of length $1/3^n$: I_s with $s = s_0s_1 \dots s_{n-1}$ for all possible $s_i = 0, 1$. Each is nested inside the previous, $I_{s_0s_1\dots s_{n-1}} \subset I_{s_0s_1\dots s_{n-2}} \subset \cdots \subset I_{s_0s_1} \subset I_{s_0}$. And the action of the map on it is: $B^{n-1}(I_s) = I_{s_{n-1}}$ and $B^n(I_s) = I$. That is, the nth iterate B^n on I_s is the line of slope 3^n. Thus, the image trio intervals of I have three equally partitioned preimage subintervals of I_s, each is of length $1/3^{n+1}$. The mid-third is not given any more itinerary since it escapes at the nth iterate of B. The left-third is given one more set itinerary 0 because it is mapped by B^n to the left-third I_0 of I. And its set itinerary is exactly $s0$ because the point itineraries diverge by the $(n+1)$th iteration as $B^{n+1}(I_{s0}) = B(B^n(I_{s0})) = B(I_0) = I$. Hence the subscript for this interval is its set itinerary $s0$. The right-third subinterval of I_s receives exactly one more set itinerary 1 for similarly reasons. In this way, we obtain the 2^{n+1} many closed subintervals of length $1/3^{n+1}$ exhausting all possible set itineraries $s_0s_1 \dots s_n$ of length $n+1$ for their address subscripts.

4.1.3 Bernoulli Shift And Conjugacy

We are now ready to give a complete description of the non-wandering set Λ. For any point x from Λ with its point itinerary $\rho(x) = s_0s_1 \dots s_n \dots$, it must belong to the Cantor mid-third construction intervals $I_{s_0s_1\dots s_n}$ of the shared set

itinerary $s_0 s_1 \dots s_n$ for all iterates $n = 0, 1, 2, \dots$. Because the nested intervals contract to a unique point $\cap_{n=0}^{\infty} I_{s_0 s_1 \dots s_n}$ such a point x not only exists but also is unique with the said point itinerary $\rho(x)$ because the subinterval length $1/3^{n+1}$ converges to 0. This shows the itinerary map $\rho : \Lambda \to \mathcal{B}$ is one-to-one, i.e. no two points of Λ shares the same point itinerary. Conversely, for every infinite binary sequence $s = s_0 s_1 \dots s_n \dots$ from $\mathcal{B}$, the nested subintervals $I_{s_0 s_1 \dots s_n}$ exists for the shared set itineraries with s. The unique intersection point $\{x\} = \cap_{n=0}^{\infty} I_{s_0 s_1 \dots s_n}$ not only exits but also its point itinerary $\rho(x)$ is exactly the said binary sequence s. This shows the map ρ is onto, i.e. every infinite binary sequence is the itinerary of a unique non-wandering point.

This is not it. The itinerary has automatically encoded the dynamics of the baker's map. Because of its definition, we have for every point from the non-wandering set $x \in \Lambda$ with itinerary $\rho(x) = s_0 s_1 \dots s_n \dots$, the image stays in the set $B(x) \in \Lambda$ and its itinerary is $\rho(B(x)) = s_1 s_2 \dots s_n \dots$, the same as x's except that the itinerary is shifted one position to the right. Let $\sigma : \mathcal{B} \to \mathcal{B}$ denote the shift operation, $\sigma(s_0 s_1 \dots s_n \dots) = s_1 s_2 \dots s_n \dots$. Then we see that σ is an onto map, which is referred to as the Bernoulli shift on the symbolic space $\mathcal{B}$.

More importantly this identity holds: $\sigma(\rho(x)) = \rho(B(x))$ for all $x \in \Lambda$. In picture, the following diagram commutes when following the arrows:

$$\begin{array}{ccc} \Lambda & \xrightarrow{B} & \Lambda \\ {\scriptstyle\rho}\downarrow & & \downarrow{\scriptstyle\rho} \\ \mathcal{B} & \xrightarrow[\sigma]{} & \mathcal{B} \end{array}$$

This relationship gives rise to the definition that the backer's map B is conjugate to the Bernoulli shift σ. Because ρ is invertible, σ is also conjugate to B by definition. Hence B and σ is said to be equivalent. This means that the combinatoric property of B on Λ is identical to the combinatoric property of σ on $\mathcal{B}$.

4.1.4 Diagonalization For Uncountability Of Non-Wanderers

Here is a beautiful proof for the uncountability of the non-wandering set Λ, making use of the conjugacy. It is referred to as Cantor's diagonalization argument which was used for the uncountability of irrational numbers. The British mathematician Alan Turing used the same idea to demonstrate that it is algorithmically impossible to determine for all possible programs whether or not an

arbitrarily given program will terminate or run forever, one of the most celebrated results in the history of mathematics. Now about the non-wandering set Λ. Since it has a one-to-one correspondence with the set of all binary sequences $\mathcal{B}$, we only need to show that $\mathcal{B}$ is uncountable. Suppose on the contrary that $\mathcal{B}$ is countable. Then all elements of $\mathcal{B}$ can be arranged as a sequence, counted out by the natural number 1, 2, 3, ets. Since each element is a binary sequence, we can list all elements of $\mathcal{B}$ vertically like below

$$\begin{aligned} &s_{11}s_{12}s_{13}\cdots\\ &s_{21}s_{22}s_{23}\cdots\\ &s_{31}s_{32}s_{33}\cdots\\ &\vdots \end{aligned}$$

with each row a binary sequence of $\mathcal{B}$. We next construct a binary sequence $\bar{s}$ that is not on the list and hence the contradiction. All it takes is to flip the bit of each diagonal entry, and call the flipped diagonal sequence $\bar{s}$. That is, if $\bar{s} = \bar{s}_1\bar{s}_2\bar{s}_3\ldots$, then $\bar{s}_k = 1 - s_{kk}$ for all $k = 1, 2, 3, \ldots$. This sequence is in fact excluded by the list because it differs from every sequence of the list at least once at the diagonal position.

One may ask if the shift map also capture the metric property of the baker's map, namely the full dynamical property of B on its non-wandering set. The answer is yes. Suffice it to say all we need is to define a distance between any two itinerary sequences from the symbolic space. A full explanation is not very complicated but it will send us onto a sidetrack. The diversion is unnecessary as we are ready to demonstrate directly the baker's map's chaos in the next section.

4.1.5 Sensitivity To Initial Conditions

Here is how to show the baker's map has the property of sensitive dependence on initial conditions. For any two different points x, y from the non-wandering set, either they belong to the left-third and the right-third intervals I_0, I_1 respectively or both are in one of the two intervals. In the first case they are separated by the mid-third interval and thus they are at least $1/3$ distance apart: $|x - y| \geq \delta_0 = 1/3$. In the second case they must share a common set itinerary $s = s_0 s_1 \ldots s_n$ and diverge on the $(n+1)$th itineraries. This means of their $(n+1)$th iterates $B^{n+1}(x)$ and $B^{n+1}(y)$ one is in the left-third interval I_0 and the other is in the right-third interval I_0, separated apart at least by δ_0 amount, i.e. $|B^{n+1}(x) - B^{n+1}(y)| \geq \delta_0$. This completes the proof of the expansive property for chaos.

To show the set P of periodic points is dense in Λ, it is obvious that every periodic point p has a periodic itinerary. Now for any point x from the non-wandering set Λ and any arbitrarily small number $\epsilon > 0$, let $\rho(x) = s_0 s_1 \dots s_n \dots$ be its itinerary and let an integer n be sufficiently large so that $1/3^n \leq \epsilon$. Let p be the point whose itinerary repeats indefinitely of the first $n+1$ itineraries of x, i.e. $\rho(p) = \bar{s}$ with $s = s_0 s_1 \dots s_n$ and $\bar{s}$ means s repeats itself indefinitely. Then p is a periodic point. Because x and p share their first $n+1$ destinations both belong to the set I_s of the same shared itinerary. Since the length of I_s is $1/3^{n+1} < 1/3^n \leq \epsilon$ we have $|x - p| < \epsilon$. This shows the periodic set P is dense in Λ.

4.1.6 All For One And One For All: Dense Orbit

To show the last property that the non-wandering set has a dense orbit, we adopt the following all-for-one-and-one-for-all construction. We first list all finite binary sequences as

$$\alpha_1, \alpha_2, \dots, \alpha_n, \dots$$

so that the first two are the binary sequences of length one: 0, 1, the next four are the binary sequences of length two: 00, 01, 10, 11, the next eight are the binary sequences of length three, and so on. We then concatenate them all into one sequence $\alpha = \alpha_1 \alpha_2 \dots \alpha_n \dots$ to get an infinite itinerary. Let a be the corresponding point in Λ with the constructed itinerary α, $\rho(a) = \alpha$. Then the orbit $\gamma(a) = \{a_0, a_1, a_2, \dots\}$ must be dense in Λ. This is because for any non-wandering point x with itinerary $\rho(x) = s_0 s_1 \dots s_n \dots$ and any arbitrarily small number $\epsilon > 0$, we can find n so that $1/3^n \leq \epsilon$ and another integer $k \geq 1$ so that the kth iterate $B^k(a)$ of the point a shares with x the same first $n+1$ itineraries $s = s_0 s_1 \dots s_n$, which is either one of the α-sequence above or a subsequence of such a longer α-sequence. As a result both belong to the interval I_s of length $1/3^{n+1}$, i.e. $|x - B^k(a)| \leq 1/3^{n+1} < 1/3^n \leq \epsilon$, showing the orbit of a is dense in Λ.

Notice that such iterate k are infinitely many as longer and longer α-sequences will repeat any given shorter α-sequence. That is, the dense orbit will visit any neighborhood of any point infinitely often, with irregularly spaced iterates k. Notice also that dense orbits are infinitely many as there are different ways to order all binary sequences of any given length. A dense orbit is everywhere in the non-wandering set and forever present.

4.1.7 Small Chaos Is Hard

We also note that a proof of chaos for the tent map is similar. The only difference is that the location address of the daughter interval $I_{s_0 s_1 \ldots s_n}$ inside the mother interval $I_{s_0 s_1 \ldots s_{n-1}}$ does not follow the bookkeeping rule for the baker's map. Depending on whether the number of 1s in $s_0 s_1 \ldots s_{n-1}$ is odd or even the left daughter interval is given the additional itinerary 1 or 0 for its set itinerary and the opposing itinerary is given to the right daughter interval. A proof of chaos for the horseshoe map is also similar except that the itinerary of a point from the horseshoe's non-wandering set is a doubly infinitely sequence in the binary symbols, with the backward itinerary, $\ldots s_{-n} \ldots s_{-1}$, for the backward orbit of the point.

Proof of chaos for these systems may have a lot in common, but it is never simple even for the simplest system. It involves a lot of intricate combinatoric arguments and metrical estimations. Constructing a proof for chaos for a randomly chosen mathematical model is never a simple task. Currently, it has been done for the three-body problem Poincaré started but couldn't quite finish, the quadratic map family and a singular perturbed 2-d version. It has been done in a limited fashion for the Lorenz equations. It has been done for the food chain model of Eq.(1.2) at its singular limit and some perturbed state at which a Shilnikov's homoclinic orbit persists, and for a few system where a Shilnikov's orbit is demonstrated. And that is all we can count. None of these systems' dimensions or effective dimensions exceeds three. For all other mathematical models we resort to numerical simulations in order to allow us to say 'look there is a chaos'. Of course, for our mockup systems by singular perturbations most chaoses can be proved because their components are constructed to make it so.

Two postscripts on Smale's discovery of the horseshoe map. The first. Within two months of finding his horseshoe map he also succeeded in finding a solution to an outstanding Poincaré conjecture in the field of topology that earned him in 1966 the Fields Medal, considered to be the Nobel Prize equivalence in mathematics. Smale considered his horseshoe map discovery 'an omen of good luck'. The second. Smale revised his conjecture about all dynamical systems to this picture instead: every dynamical system can be perturbed to have the property that the non-wandering set has an exponentially contracting foliation and an exponentially expanding foliation. His program for all dynamical systems effectively reached an end when one of his students proved in the 1970s a result that says when the stable and unstable manifolds of a fixed point intersect tangentially at one point then they must intersect tangentially at in-

finitely many points and no small perturbations can make all these structurally unstable sinks to go away.

4.2 Big Chaos Is Easy

In terms of structures chaos is at the top echelon of dynamical systems. Equilibrium solutions of differential equations are zeroes of the equations' vector fields. Periodic solutions are fixed points of the equations' return maps. Chaos is an entirely different beast, many folds more complicated. Chaos of the simplest functional form, the logistic family, requires some of the most complex mathematics to prove. Even the smallest chaos is hard to prove, as a rule of thumb, unless of course we are talking about the spike renormalization. The renormalization dynamics is big, infinitely big in dimension. Yet contrary to intuition it is exceedingly simple to prove it to be chaotic.

4.2.1 The Stable Manifold Of Spike Renormalization

Recall from the previous chapter the renormalization map R. It is a map in the renormalizable space $\mathcal{U}$, of which any element is a map from the unit interval $[0,1]$ into itself with these properties. Property number one, each map f has a discontinuity point $0 < c_0 \leq 1$ of which f is increasing, not below the diagonal left of c_0. Property number two, f's maximum in the right interval $(c_0, 1]$ is no greater than f's minimum $f(0)$ in the left interval. Property number three, the critical point c_0 has the immediate backward iterate c_{-1}. For such a function the return map rf is defined on the left interval $[0, c_0]$ because f maps $[0, c_{-1}]$ into $[0, c_0]$ and it maps first $(c_{-1}, c_0]$ into $(c_0, 1]$ and then $(c_0, 1]$ into $[0, c_0]$. Scaling rf up into the unit interval $[0,1]$ completes the definition of $R(f)$ as below:

$$R(f)(x) = \begin{cases} \frac{1}{c_0} f(c_0 x), & 0 \leq x \leq \frac{c_{-1}}{c_0} \\ \frac{1}{c_0} f(f(c_0 x)), & \frac{c_{-1}}{c_0} < x \leq 1 \end{cases} \tag{4.1}$$

As we have demonstrated in the previous chapter, the identity map $\psi_0(x) = x$ is a fixed point and the ψ-family $\psi_\mu(x) = \mu + x$ for $0 \leq x \leq 1 - \mu$ and $\psi(x) = 0$ for $x \geq 1 - \mu$ is an invariant expanding-center-manifold of the identity fixed point. The orbit of any member ψ_μ under R is this generalized harmonic sequence $\{\psi_{\mu/(1+n\mu)}\}$ in backward iteration $R^{-n}(\psi_\mu)$. We consider in this section the stable manifold, denoted by W^s, of the fixed point ψ_0.

Actually we will relax the standard definition from asymptotic stability to include stability of non-asymptotic type as well. Specifically we will define

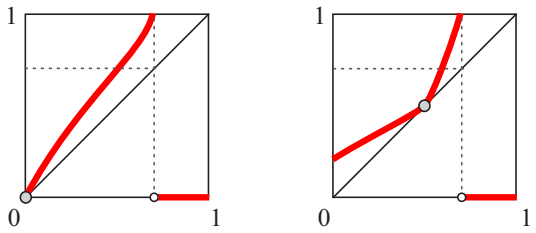


Figure 4.3: Partition of the stable manifold $W^s = X_0 \cup X_1$. Left: A member of X_0. Right: A member of X_1 of which the right-most fixed point is not the origin.

the stable manifold of the identity map to be the set of all maps which are indefinitely renormalizable. That implies first for any element f of W^s the graph of f must not strictly lies above the diagonal line in its left interval $[0, c_0]$. Otherwise f's backward iterates $\{c_{-k}\}$ of c_0 is a finite sequence and a finite iterate of f by R will escape the renormalizable space $\mathcal{U}$. This means, the backward orbit $\{c_{-k}\}$ must be infinite and non-increasing. As a result the limit $\lim c_{-k} = x^*$ exists and and the limit must be a fixed point $f(x^*) = x^*$ with $0 \le x^* \le 1$. This is only a necessary condition for an element f of $\mathcal{U}$ to be indefinitely renormalizable. The remaining sufficient condition is for the maximum of f in the right interval $(c_0, 1]$ to be no greater than the fixed point x^*. Therefore, the stable manifold is

$$W^s = \{f \in \mathcal{U} : \text{there is a } 0 \le x^* \le c_0 \text{ so that } f(x^*) = x^* \text{ and } \max_{(c_0,1]} f \le x^*.\}$$

For example, the u-family of maps at the limiting value $\mu = 0$ consists of a one-parameter family of fixed points, $u_{\lambda,0}$, of the renormalization. Strictly speaking this set is in the center manifold of the identity map ψ_0 because the renormalization has 1 to be its eigenvalue as this simple calculation shows:

$$\|R(u_{\lambda,0}) - R(\psi_0) - 1 \cdot (u_{\lambda,0} - \psi_0)\| = 0 \sim \|u_{\lambda,0} - \psi_0\|^2.$$

Unlike the ψ-family the $u_{\lambda,0}$-family is stable relative to the identity map ψ_0.

A further natural partition of the stable manifold W^s is self-evident: The subset X_0 of the manifold which consists of all elements for which the right-most fixed point x^* is the left end point 0, i.e. $X_0 = \{f \in W^s : x^* = 0\}$, and the subset X_1 for which the said right-most fixed point is not the zero, i.e. $X_1 = \{f \in W^s : 0 < x^* \le c_0\}$. See Fig.4.3. Both subsets are mutually complimentary for the stable manifold as $W^s = X_0 \cup X_1$.

4.2.2 Converging At Any Rate

The dynamical structure of R in the X_1-subset is simple. First, every orbit $R^n(f)$ converges to a fixed point in X_1 as n goes to infinity. One can see this easily in R's action which is this crop-enlarge operation on the box $[0, c_0] \times [0, c_0]$. At the limit $\lim R^n(f) = f^*$, f's right-most fixed point x^* becomes the right-end point $x^* = 1$. The identity map ψ_0 is one of infinitely such many points of the X_1-subset. Here is another family of fixed points $v_\lambda(x) = \lambda(x-1)+1$ with $0 < \lambda \leq 1$ which is the line of slope λ through the upper-right corner point $(1, 1)$.

Second, there is a peculiar property above every fixed point f^* in the X_1-subset: for every $0 < r < 1$ there is a point f_r so that $R^n(f_r)$ converges to f^* at the rate r. All it takes is to replace the function of f in the subinterval $[x^*, 1]$ by the line of slope $1/r$ from the diagonal point (x^*, x^*) to the top of the unit square and fill the right of the new critical point by zero. This is easy to check for the identity element ψ_0. More specifically, if we start with the S-family's S_0 map as such a function f, then f_r's graph in the right interval $[0, c_0]$ is the diagonal line plus the $1/r$-slope line with $c_0 = (1+r)/2$. By computing their distances to the fixed point one can check the following limit holds

$$\lim_{n\to\infty} \frac{\|R^n(f_r) - \psi_0\|}{\|R^{n-1}(f_r) - \psi_0\|} = r.$$

Here $\|f - g\| = \int_0^1 |f(x) - g(x)|dx$ is the average distance between the function values of f, g in the unit interval $[0, 1]$.

4.2.3 Infinity Chaos

What is unexpected is the property that the other piece of the stable manifold is chaotic under the renormalization operation, and even more so is the fact that it takes practically very little to prove.

Theorem 10 *The renormalization R is chaotic in the subset X_0 of the stable manifold of the fixed point ψ_0.*

Because the proof is simple we will give an illustration for the rest of the section. The most advanced piece of mathematics required is the property that the rational numbers are countable, i.e. having a one-to-one and onto correspondence with the natural numbers. This property will be needed to construct a dense orbit in the X_0-subset.

First, the property of sensitive dependence on initial conditions for R is illustrated by the picture of Fig.4.4. It shows for any element f of X_0 and any

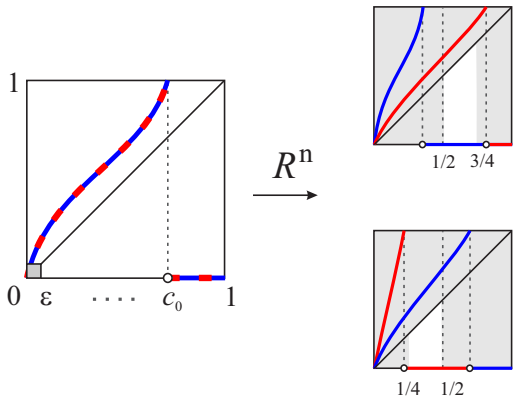


Figure 4.4: R's sensitivity to initial conditions. For any f (blue) of X_0 and any small $\epsilon > 0$, construct g (red) depending on two cases of f: $c_0 \leq 1/2$ (top-right) and $c_0 \geq 1/2$ (bottom-right) with c_0 being the discontinuity of $R^n(f)$. Both will be at least $\delta_0 = 3/32$ (the area of the small trapezoid in white) apart for some large iteration n.

arbitrarily small $\epsilon > 0$, we can always construct another element g to meet the requirement for sensitive dependence on initial conditions. As g is identical to f outside the small ϵ-box at the lower-left corner of the unit box they are ϵ-close to each other $\|f - g\| < \epsilon$. The ϵ-uncertainty is enough for us to construct anything we want out of g. If we take the integer n as such that $c_{-n-1} < \epsilon \leq c_{-n}$ with c_{-k} being the backward iterates of f's critical point c_0, then we know by the nth iteration R^n, the ϵ-box is enlarged to contain at least the unit box. What used to be covered up by the small ϵ-difference is now magnified to a full view. If the magnified new c_0 of f is left of the middle point $1/2$, we can then match up that with a discontinuity point of g which when under the R^n magnification is beyond the point $3/4$. The area difference between $R^n(f)$ and $R^n(g)$ is at least of the trapezoid under the diagonal and over the interval $[1/2, 3/4]$. If on the other hand the magnified new c_0 of f is right of $1/2$, we construct a g whose magnified new c_0 is below $1/4$ so that the difference between their nth iterates of R is at least of the trapezoid area under the diagonal and over the interval $[1/4, 1/2]$. In either case, the difference is at least $\delta_0 = 3/32$, the area of the latter trapezoid.

4.2.4 Concatenation

The only loose end we need to tie up is to explain how to attach an arbitrary piece to a given function f inside a prescribed lower-left corner box to get the function g, all from the X_0-subset. This idea is geometrical and straightforward. The essential steps are illustrated in Fig.4.5. Take any f from X_0 with its backward orbit c_{-k} of its discontinuity point c_0. Let us say another element h of X_0 is needed to be attached to f inside the box $[0, c_{-k}] \times [0, c_{-k}]$. We first cut away the piece of f inside the box. We then scale down h to fit it inside the box. Either the down-scaled h fits with the truncated f continuously or we only need to join them by a small line-segment. Denote the concatenated function by $f \vee h$ which is the function g in the proof for R's expansivity above. One can see easily that either $R^k(f \vee h) = h$ or $R^{k+1}(f \vee h) = h$. Let's refer to this glue operation the concatenation of f and g. Notice that the concatenation can be glued with another element, and so on, even indefinitely so. In such cases we denote them by $f_1 \vee f_2 \vee f_3 \vee \cdots$, etc. Such operations will be called up below for the construction of dense orbits.

The proof for the periodic set being dense in X_0 is straightforward. In fact, let f be any element of X_0, and let $\bar{f} = f \vee f \vee f \cdots$ be the concatenation of f with itself indefinitely and with a fixed choice c_{-k} related to its discontinuity point c_0 for all its concatenating copies throughout the construction. Obviously $\bar{f}$ is either periodic k or periodic $k+1$ of R. This shows for any element f from X_0 and any arbitrarily small $\epsilon > 0$ we can find a periodic orbit of R that is within the ϵ distance from f if we make $\epsilon < c_{-k}$.

4.2.5 Countable Rational Numbers And Dense Orbit

So we are left with the only argument for the existence of a dense orbit in X_0. It is harder to prove than the last two properties of chaos above, but not by a lot. We will still use the all-for-one-and-one-for-all type of argument as we did for the baker's map. The only short detour we have to take is about the fact that the set of all rational numbers is countable and dense in the unit interval $[0, 1]$. The set is countable because we can list all rational numbers in the array of the quotient form p/q with integer p increasing in row and q increasing in

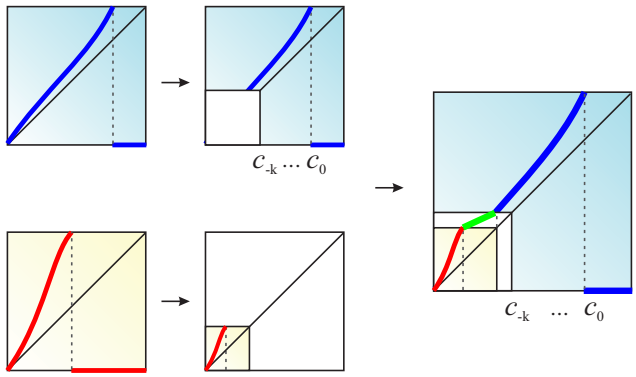


Figure 4.5: The concatenation of f (blue) and g (red) in the X_0-subset. This illustration is for the case for which $R^{k+1}(f \vee g) = g$.

column as below

$$\begin{array}{ccccc}
1/1 & 1/2 \to & 1/3 & 1/4 & \to \\
\downarrow \nearrow & & \swarrow \nearrow & & \cdots \\
2/1 & 2/2 & 2/3 & 2/4 & \cdots \\
\swarrow & & \nearrow & & \cdots \\
3/1 & 3/2 & 3/3 & 3/4 & \cdots \\
\downarrow \nearrow & & & & \cdots \\
4/1 & 4/2 & 4/3 & 4/4 & \cdots \\
\vdots & \vdots & \vdots & \vdots & \ddots
\end{array}$$

We then count them from the upper-left corner one forward-slash diagonal a time. This is the classical argument for countability of the rationals. Using decimal point expansion one can show that in every neighborhood of any real number there are infinitely many rational numbers inside the neighborhood, a standard argument for the denseness of the rationals. A set having a countable and dense subset is called a separable space. It is based on the separability of the real numbers that one can prove the set of all integrable functions in $[0, 1]$ is also separable. Here by definition integrable means the net area between the graph of a function f and the x-axis as defined by the definite integral $\int_0^1 |f(x)|dx$ exists. This set of functions is the so-called $L^1[0, 1]$ space. As an inherent property the X_0-subset of the renormalizable space with the average distance metric is also separable. This fact is the only nontrivial property we will assume without detailed justification for the last part of the chaos proof for R on X_0.

We first list the dense subset of X_0 as $g_1, g_2, g_3, \ldots$. For each g_i denote its discontinuity point by $c_{0,i}$ and its backward iterates by $c_{-k,i}$. For each k, we cut away its piece in the lower-left corner $[0, c_{-k,i}]$ for $k = 1, 2, 3, \ldots$, and

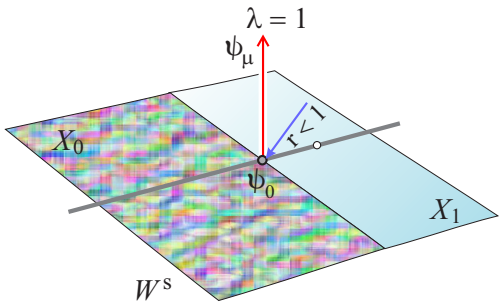


Figure 4.6: The non-hyperbolic split of the fixed point ψ_0 in the renormalizable space $\mathcal{U}$ with ψ_μ the expanding-center manifold and W^s the stable manifold.

denote the truncated piece by $g_{k,i}$. The set of all truncations of all gs is again countable because $g_{k,i}$ can be arranged in an array with k being the row numbers and i being the column numbers and count them as we has done the same for the rational numbers above. We list the result as a sequence and denote it by $h_1, h_2, h_3, \ldots$. Without further cut-aways the sequence concatenation

$$h = h_1 \vee h_2 \vee h_3 \vee \cdots = \vee_{n=1}^{\infty} h_n$$

is the needed dense orbit. The reason is that for any n there is a k so that $R^k(h) = h_n \vee h_{n+1} \vee h_{n+2} \cdots$, and h_n can be the needed piece that is within an arbitrarily small neighborhood of an arbitrary element f of X_0.

Fig.4.6 gives a graphical summary for the map R in the space $\mathcal{U}$. The line through ψ_0 in the X_0 subset represents the family of fixed points given by the u-family, i.e. $u_{\lambda,0}$ for $\lambda > 1$. The line segment in the X_1 subset represents the fixed point family v_λ for $\lambda < 1$. The rate of convergence to any fixed point in X_1 can be any number r between 0 and 1, the 0 included. The critical difference between our renormalization and the Feigenbaum-Coullet-Tresser renormalization is the fact that their fixed point is hyperbolic and ours is non-hyperbolic. A hyperbolic structure is a rigid structure of which the non-wandering set is the fixed point singleton. A non-hyperbolic structure can be flexible. In our case it is extremely flexible to the point that the non-wandering set is infinitely dimensional and is chaotic.

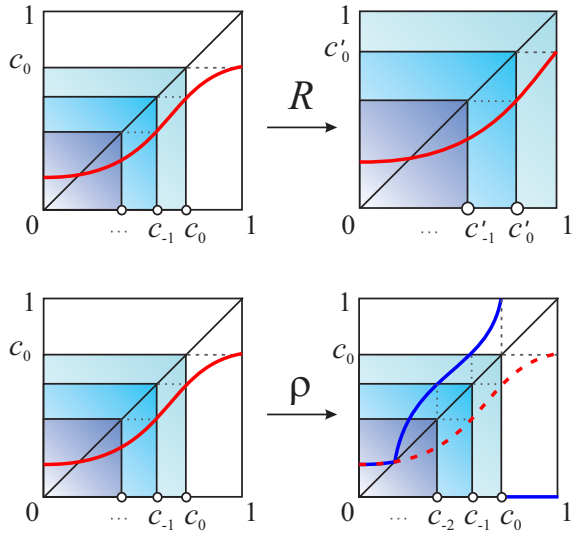


Figure 4.7: Top: The renormalization in $\mathcal{V}$ of monotone increasing functions in $[0,1]$. It is the same crop-enlarge operation as R in $\mathcal{U}$. Bottom: The conjugacy map ρ from $\mathcal{V}$ into W^s that commutes with the renormalizations in both $\mathcal{U}$ and $\mathcal{V}$.

4.2.6 Renormalization Without Spike

We end this section by asking the question that are there other chaotic and infinitely dimensional non-wandering sets? The answer is yes but there is only one that we can find and it is not that much different from our spike renormalization map R. In fact, the new system is conjugate to R in W^s but no bigger. For the new system let $\mathcal{V}$ be the set of monotone increasing functions on $[0,1]$ without discontinuity, see Fig.4.7. Such a function f can be the return map of a periodic orbit of a two-dimensional differential equation. Let formally $c_1 = 1$ and c_{-k} be the backward iterates of c_1. For such maps return maps can always be defined. For example, the return map of f can always be defined on the interval $[0, c_0]$. Scale up this return map by $1/c_0$ and we obtain the renormalized image $R(f)$ in $\mathcal{V}$. In this setting of return maps the Feigenbaum-Coullet-Tresser renormalization, the spike renormalization, and this new type renormalization are all of the same type. The only difference is in the type of renormalizable space.

Dynamically the new renormalizable space is only a subspace of the spike renormalizable space. In fact, we can construct a conjugate map ρ from $\mathcal{V}$ into

W^s, see Fig.4.7. It goes as follows. For any f from $\mathcal{V}$ we first modify it this way: wherever f is below the diagonal we replace the point by its mirror image of the diagonal. This will create a graph over the left interval $[0, c_0]$ and leave the right subinterval $(c_0, 1]$ empty. All we need to do is to fill the void zero. This defines the element $\rho(f)$ in W^s, and hence the map $\rho : \mathcal{V} \to W^s$. Clearly it is one-to-one. More importantly the backward iterates c_{-k} of the point c_0 are preserved by the mirror-reflection. As a consequence the renormalizations are preserved by the mirror-reflection as well. This means that the map ρ commutes with the renormalizations, $R(\rho(f)) = \rho(R(f))$ for every f from $\mathcal{V}$. That is, ρ is a conjugacy. If we let Y_0 be the preimage in $\mathcal{V}$ of X_0 by ρ, then ρ is both one-to-one and onto from Y_0 to X_0, and hence the renormalizations on them are equivalent. The dynamics of R in Y_0 is the same of R as in X_0. As a result R is chaotic on Y_0 as well. As of this time these two systems are the only known systems whose chaotic non-wandering sets are infinitely dimensional.

4.3 Defining God

On the issue of divine creation Newton and Darwin can be considered far apart and similar at the same time. To Newton the Sun, the Moon, the falling apple were all governed by the law of gravity and the universal law had to have a first cause. To him that first cause must be the existence of a supernatural being. To Darwin the immense order, diversity, and beauty of life could be explained by the evolutionary process of selections, but he did acknowledge occasionally the possibility of a first course by a creator who had set the process in motion from a primitive common ancestor for all life forms on Earth. To many scholars Newton was a theist but Darwin an atheist. Without changing their respective views or positions no one could prove them right or wrong. Because both are backed up by another great genius in the human history.

At the age of 25 Kurt Gödel published his incompleteness theorem in 1931. Many consider his work one of the three greatest intellectual achievements of the 20th century: the theory of relativity, the theory of quantum mechanics, and his incompleteness theorem. In a nutshell he demonstrated that for any formal mathematical system permitting the addition and multiplication of the natural numbers, if it is self-consistent, i.e. without the possibility to generate propositions which are both true and false at the same time, then there must be propositions which cannot be proven either true or false within the system. He further demonstrated that the consistency of such a system cannot be proved within the system. It is not much a stretch to say his mathematics has provided

the theoretical foundation for all religious systems.

4.3.1 The Mathematical Basis For All Religions

The only difference that separates the human race from other animals is the written language. There would be no reason for its existence if it were not used by us to capture, to express, and to reason with causality. Every writing system can be considered to be a symbolic system of causality and hence logic in extension. Mathematics happens to be the most formal of all. Without symbolism would there be religions? The answer is probably not just by taking a look at the animal kingdom. As far as we know no animal species practice any religion by our definition, which perhaps is one of many consequences to the written language difference between us and them. Without causality reasoning can any proselytism be effective? It is probably not a stretch to argue that without reasons or assumed logics there would be no religions. So if the most rigorous symbolic system, the mathematics, is incomplete, why would systems of lesser rigor be more complete? Would it be reasonable to argue that there are statements, religious statements perhaps in particular, that cannot be proved nor disapproved? Perhaps the proposition about the existence of God, defined or not, is such a proposition that no human symbolic systems is capable of demonstrating. Gödel, a theist, has given both theists and atheists the same theoretical basis for their respective beliefs.

Gödel had solved the most burning issue, but not all the issues. For examples, his theorem cannot tell if a given proposition is provable or not provable. Until it is proven one way or another the provability question remains perpetually open. So you would never know if your God proposition can be proved inside your system until you had. You may consider a shortcut to make your God supposition an axiom. But you must proceed with caution because you may assume too much to make your system inconsistent, proving something both true and false at the same time. Besides you will not be able to solve the problem that would allow you to brag because Gödel's theorem also foretells that you cannot prove your expanded system is consistent within the system. Gödel has given every belief an egalitarian basis to start, as well as an egalitarian basis to finish because none can be proved logically and therefore intellectually superior than the others.

4.3.2 Dense But Empty Or Dense And Full, That's The Difference

If mathematics has this much to say about the basis of all religions, can it have something more to say about God? If one has to give a mathematical definition of God, what will it be? The one mathematical object that is the closest parallel to the characteristics of an omnipotent being is the dense orbit. A dense orbit is not just any orbit but one that can visit every point of a set indefinitely frequent and arbitrarily close. It can be everywhere, at the least opportune time, forever stalking every point. In limit it creates every member of the set. It possesses this one-for-all and all-for-one quality. But not all dense orbits are God-like. Size matters. For the smallest chaos, the non-wandering set of the baker's map is puny, practically nothing even though it is uncountably numerous. It is the leftover after all the middle third intervals being removed, first the one middle third of length $1/3$, then the two middle thirds of length $1/3^2$ each, and so on. At the nth step, there are 2^{n-1} many middle third intervals of length $1/3^n$ each that are removed. When one tallies up all the removed intervals the total length is 1, the measure of the whole interval. When added up the complimentary non-wandering set occupies a set of the length 0. The set is also called the Cantor's dust but in size it is measured less than anything. So any orbit, dense or otherwise, is of no significance if it is only relevant to a set of zero measure.

Unless, of course, it is about the spike renormalization map R. We know it is chaotic in the X_0-subset of the stable manifold of the identity map ψ_0. We know it has infinitely many dense orbits. To make these dense orbits God-like the chaos set X_0 has to be big. It has to be stupendously big to contain all systems, deterministic or nondeterministic. It has to have the trajectory of every atom of everybody, the trajectory of every subatomic particle of the universe. We seem to ask too much. But the map has never disappointed, it has always delivered. And it will come through once more. In fact we can prove the following theorem.

Theorem 11 *For every finite dimensional mapping $F : \mathbb{R}^k \to \mathbb{R}^k$ there is a conjugacy $\rho : \mathbb{R}^k \to X_0$ so that $R(\rho(p)) = \rho(F(p))$ for every point p from the phase space $\mathbb{R}^k$. The same holds for the infinite dimension case with $k = \infty$.*

It takes very little to prove this theorem, less than proving the existential proof of dense orbit. It is so simple that we can get away with committing a cardinal sin in mathematics — proof by example. That is, pick any of your favorite system, deterministic or otherwise, we can show you how to construct the required conjugacy which embeds your system inside the X_0-set of the renor-

malization R so that the orbital structure of your system is perfectly preserved by the renormalization. Once your example is done you will see how to scale the argument up to get a formal proof for all systems. Let's say your favorite dynamical system is the Lorenz equations. Here below is what it takes to equate it with a piece of the renormalization.

4.3.3 Time Your System

The Lorenz system has three equations, each defines one variable referred to as phase variable and denoted by X, Y, Z respectively. Together these phase variables define the so-called phase space which is the three-dimensional Euclidean space $\mathbb{R}^3$. Let $p = (X, Y, Z)$ denote a point in the phase space with the corresponding phase variable coordinates, and p_t denote the solution of the equations at time t. As before p_0 and p are interchangeable. When plotted out in the phase space p_t defines what is called the orbit of the system with the initial point p_0. Although the ancient Greek philosophers believed the Euclidean space exist independent of any reality, for a dynamical system its phase space and its dynamics in the phase space are inseparable. Without the dynamics there is no phase space and without the phase space there is no dynamics to speak of. So the embedding to be constructed below is not only to find a place for each phase point in the renormalizable space but also to find a place for the orbit of each point the dynamics is represented by in the renormalizable space.

The idea starts by following an idea of Poincaré. When Poincaré tried to tackle the three-body problem that started the modern era of dynamical systems, he also introduced another type of map to change a continuous flow-like object, i.e. the orbit p_t, to a discrete mapping. In this case you don't even need any cross section. You simple take and fix any time interval, say τ, and then define for each point p the next point τ-time away on the orbit: $F(p) = p_\tau$. In this way an orbit of the equations is an orbit of the map and vice versa. More specifically, two phase points are related by the system if and only if the two points are related by a time-τ Poincaré map for some infinitely many such τs. Although different time interval τ defines a different map, but any two of them are automatically conjugate to each other because of this time invariance property $(p_t)_s = p_{t+s}$ which implies $p_\tau = (p_1)_{(\tau-1)}$ and symmetrically $p_1 = (p_\tau)_{(1-\tau)}$. For this reason we only need to consider one of them, and for definitiveness we may just take the unit to be the time interval, and call the map the time-one map. We can even get away with calling any such a map the time-one map because of the conjugacy. The important thing is the phase space

defined by the system is perfectly preserved by the phase space defined by any one of the maps. So if we can embed every orbit of the time-one map into the renormalizable space then we can consider we have embedded the system into the renormalizable space.

For aesthetic reasons we often fix a small τ and call the map the time-one map F. With τ small enough the orbit of the map F does not appear to be jumpy — the initial point p and its first iterate $p_1 = F(p)$ will not be too far apart so that the trajectory of the map moves about like the trajectory of the flow p_t. With a large τ although the orbit of the map is always on the orbit of the flow the adjacent iterates can be far apart. So from now on we will only consider the time-one map F of such a fixed small time τ.

4.3.4 Everything Is A Vibrating String

We are now ready to start the construction of the embedding. It is based on how we will represent each phase point of the map. We will break down the construction or the explanation into three steps or three preparatory parts, all are illustrated in Fig.4.8 and Fig.4.9. In step one, we first convert every phase point p into a curve in a fixed interval, say $[0, 1]$. We call each of the end points $0, 1$ a node. We pick and fix three points in the interval and call them the antinodes. For definitiveness we space them apart equally. The vertical lines through antinodes are called the antinode bars. The first from the left is the X-antinode bar, the second the Y-antinode bar, and so on. For each phase point $p = (X, Y, Z)$ with the corresponding coordinates, we place a point on the X-antinode bar with exactly the coordinate value X, a point on the Y-antinode bar with the Y value, and so on. For the Lorenz system we end this placement with the Z value on the Z-antinode bar. In general we will do this on all the antinode bars with the corresponding coordinate values. We will always place the $(0, 0)$ point on the 0-node bar and the $(1, 0)$ point on the 1-node bar, and call them the end nodes interchangeably with the end points of the interval $[0, 1]$. We now complete the first step by connecting these points by line segments in the assumed order: starting from the 0-node to the right adjacent X-antinode point, then to the next right antinode point, and so on, ending at the right 1-node. This defines a polygonal curve with the end nodes always fixed.

As a consequence, as a phase point moves on an orbit $\{p_n\}$ with $p_n = F(p_{n-1})$, the motion of the orbit becomes the motion of a standing wave with two stationary nodes at the end. Its crest and trough occur only on the antinode bars as the coordinate antinode points gliding up and down on the coordinate

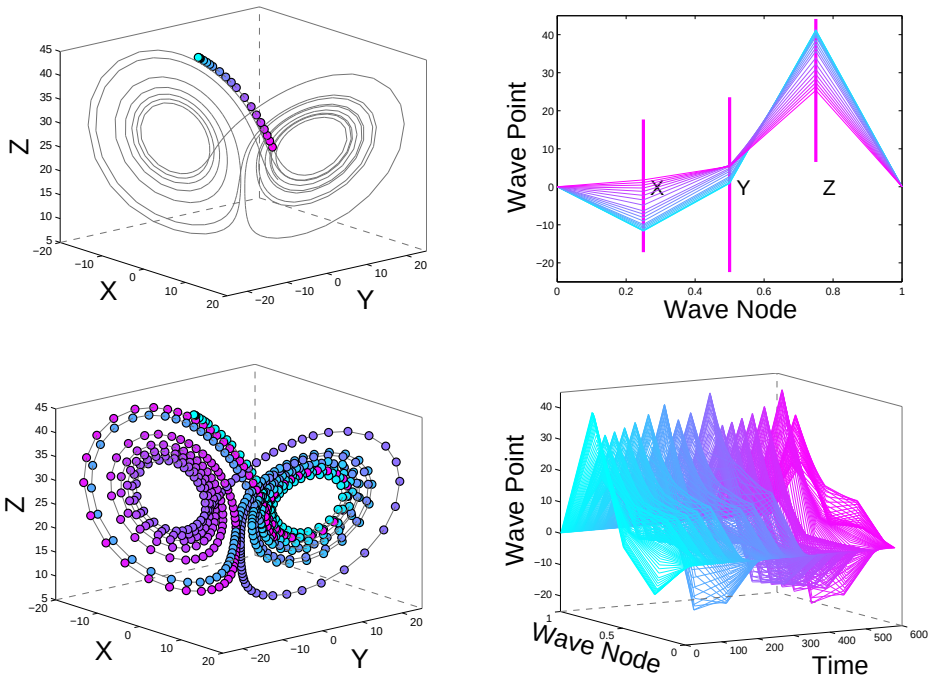


Figure 4.8: Top left: An orbit of 30 points on the Lorenz attractor. Top right: the orbit is represented by a standing wave. Color codes the iterate point and the corresponding wave profile. The colored segments of the antinode bars represent the ranges of the attractor in its respective ordinates. Bottom: the same except the orbit is 600 iterates long and its standing wave representation, color coordinated, is also viewed along the iterative time dimension.

bars. When the iterative time dimension is added to the standing wave representation an orbit of the map becomes an undulating surface. A plot of the orbit in the phase space is one form of visualization. A plot of the orbit in the standing wave space is another form of visualization. If the Cartesian coordinate representation of the Lorenz system can invoke a picture of the fluttering of a butterfly in the Amazon of Brazil, then the standing wave representation of the system should surely complete the picture of the ensuing stormy wave of a hurricane in the gulf of Mexico, a metaphor first used by Lorenz to describe his newly discovered chaos phenomenon and had since found a place in the collective consciousness of the popular culture. The former visualization is limited by

the spacial dimension of a system. The latter has none — we can always take a look at the orbital wave of any system in one glance. So we can see and then say every dynamics, deterministic or otherwise, is a vibrating string or wave.

4.3.5 Wave Running Up A Stair

Next, the step two of the construction, or the part two preparatory explanation of the embedding, see Fig.4.9. Here we will represent each phase point p by a point or member of the X_0-subset of the center-unstable manifold of the identity map ψ_0. In fact, we will represent it in a neighborhood of the fixed point $u_{\lambda,0}$ of the u-family for any $\lambda > 1$. Recall that for any $\lambda > 1$, the line through the origin $(0, 0)$ of slope λ defines the graph of $u_{\lambda,0}$ in the active interval $[0, 1/\lambda]$ and the value of $u_{\lambda,0}$ in the silent interval $(1/\lambda, 1]$ is zero. For a simpler notation we drop the subscript 0 and use $u_\lambda = u_{\lambda,0}$. We know u_λ is in the chaotic set X_0 and is a fixed point of the renormalization, $R(u_\lambda) = u_\lambda$. For each phase point p, we will denote by g the element of X_0 that is the outcome of this second step. First, g is exactly u_λ except on first backward interval $[c_{-1}, c_0]$ of the critical point, for which $c_0 = 1/\lambda$ and $c_{-1} = 1/\lambda^2$. In this interval, we will replicate the standing wave representation of the phase point with one essential modification. First, instead of the horizontal line between the stationary nodes we will tilt the line along the line of u_λ, making the cobweb points (c_{-1}, c_0), $(c_0, 1)$ the new stationary wave nodes. We still place three antinodes between them. The essential difference is, instead of using the system's native coordinate ranges we first scale them all to a uniform small range, place them immediately above the λ-sloped line. We will scale them short enough so that the new antinode bars run like steps in a stair — the first coordinate's antinode bar is below the second coordinate's antinode bar, and so on. Obviously, the more antinodes are the shorter antinode bars need be to squiz them into a stair run of a fixed height. It is doable always for any finite number of antinodes. In fact one can even accommodate countably many antinodes. Finally, we use the same scaling as for the antinode bars to scale and place the coordinates of the phase point p on their corresponding antinode bars. Connect these antinode points and the two end nodes with line segments, from the left node to the next right antinode, and on and on to the last right node on the top of the unit square box. Because the antinode bars run like a stair, the polygonal piece is monotone increasing. Now the construction of the element g is completed by fixing this monotone piece to the rest of g. Clearly g is in the chaotic subset X_0.

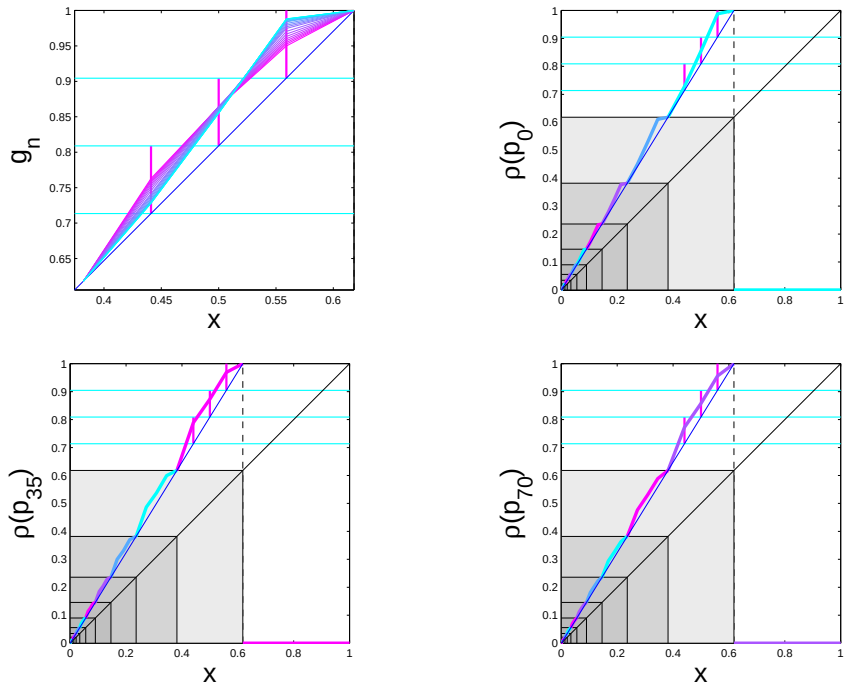


Figure 4.9: Top left: The renormalizable representations for the first 30 points of the same orbit of the Lorenz equations as Fig.4.8, color coordinated with its first two plots. Only the part in the subinterval $[c_{-1}, c_0]$ is shown. The part of g_n not shown is the line segment λx in the rest left interval and $g_n = 0$ in the right interval $(c_0, 1]$. Top right: The full view of the conjugate image of the initial point p_0 as the concatenation of the renormalizable elements, $g_0, g_1, \ldots$, in the forward orbital order of the system. Bottom: The same embedded elements in X_0 for the 35th and the 70th iterates. Only four different colors are used to cycle through the concatenation construction. The conjugacy relationship $\rho(p_n) = R^n(\rho(p))$ always holds.

4.3.6 Zero Is The Origin Of Everything

Finally, the last step. This is to define the conjugacy $\rho : \mathbb{R}^k \to X_0$. For every phase point $p = p_0$ from the phase space $\mathbb{R}^k$ of the system, we first generate its orbit $p_0, p_1, p_2, \ldots$. For each iterate p_n we generate its X_0-counterpart g_n described above. The conjugating image of p_0 point is not the g_0 element because in just one application of the renormalization every g_n becomes the fixed point

u_λ. That is, all the g-elements are in the super stable manifold of the u-fixed point. Instead, we complete the last step by defining the embedding $\rho(p)$ as the concatenation of the g-sequence as follows,

$$\rho(p) = g_0 \vee g_1 \vee g_2 \vee \cdots = \vee_{n=0}^{\infty} g_n.$$

Because the node points of the g_n elements are fixed on the cobweb points of the critical point c_0, the standing wave part of g_1 is scaled down by the concatenation to fit it inside the next backward interval $[c_{-2}, c_{-1}]$, and that part of g_2 is inside the next $[c_{-3}, c_{-2}]$ interval, and so on. That is, all the future iterate representations are crammed toward and packed at the origin, and any g_n representation can be brought on to a full view by the crop and enlarge action of the renormalization. Because of this construction, we see immediately that the renormalization of the conjugate image of the point is exactly the embedding of the iterate of the point. That is, $R(\rho(p_0)) = \rho(p_1) = \rho(F(p_0))$, namely the following conjugacy diagram commutes:

$$\begin{array}{ccc} \mathbb{R}^k & \xrightarrow{F} & \mathbb{R}^k \\ {\scriptstyle\rho}\downarrow & & \downarrow{\scriptstyle\rho} \\ X_0 & \xrightarrow[R]{} & X_0 \end{array}$$

This completes the representation of every dynamical system, large or small, deterministic or nondeterministic, inside the chaotic space X_0 of the spike renormalization R, and hence a proof of the theorem.

4.3.7 Forever Mystery

It is hard to not notice some of the metaphysical and metaphilosophical points about the theorem. The first is somewhat obvious. At a first glance one may think all physical laws or models described by partial differential equations are not subject to the embedding because they are infinitely dimensional. But at a second look one realizes there is no escape for them from the renormalization square. In fact every continuous partial differential equation is an idealization of some discrete process, and any implementation or simulation of such an equation must take on a finite dimensional discretization as its numerical approximation. Such discretization always results in a finite dimensional system or map which in turn is the subject of the spike renormalization conjugation.

The second point is paradoxical. One may wonder if some infinite dimensional partial differential equation is large enough to contain all finite dimensional systems, and question should the embedding theorem be such a big deal.

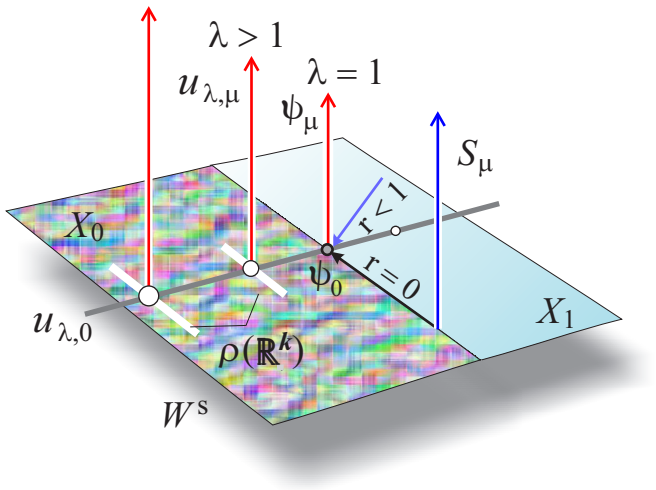


Figure 4.10: A summary of the dynamical structure of the renormalizable space $\mathcal{U}$. The identity element, the center-manifold, the first nature number are the literal and metaphorical center of the renormalization.

Upon a momentary reflection one realizes that any partial differential equation we know always has a definitive structure. Say it is very big, like the FitzHuge-Nagumo equation capable of supporting infinitely many traveling waves but not standing waves. Then it is not large enough because it will freeze out any system of standing waves. That is, peculiar and definitive structure of a model is a desirable thing which is also the undoing for all-encompassing. So one may reach a conclusion that an all-inclusive system must be featureless, must lack any definitive structure in order to be flexible. In contrary the exact opposite is the case. As shown in Fig.4.10, the spike renormalization is full of structures. All rates of convergence from $r = 0$ to $r < 1$ are represented. All rates of divergence from $r = 1$ to $r < \infty$ are represented. All real numbers are universal, and all systems are represented inside the renormalization chaos. Strange and intriguing of all is the spike generation S-family, whose electrical-dynamical states enable us to make up this organization for all systems, that arises from the boundary between the chaotic goo X_0 and the dead space X_1, as if it has to have a bird eye view on both worlds.

The third point is metaphysical. We know that all laws of physics are time reversible. That is, the present state of a physical system is determined by any state in the past. Conversely, one can reserve the time and the laws hold per-

fectly in symmetry. This means the present state of the system also determines any past state as well as any future state. Or, any future state that is yet to come has already determined the present state. On the other hand we know or it appears to us that time unfolds only in one direction. Reality proceeds only into the future not the past. Time is irreversible. This time irreversibility is oddly consistent with the spike renormalization. It is an irreversible operation. Also, the conjugacy which finds a place in the renormalization for every process can only do so for the present and future states of the process not its past. It favors the future-determines-the-present bias, a definitive break from the time duplicity symmetry. When a phase point is embedded into the renormalization space all the forward iterates must be used to determine its renormalization representation. The more distant a future state is the closer to the origin the renormalization counterpart locates, and the fuzzier or more blurry the future state looks from the present. Yet, as the renormalization unfolds the future state gets closer, looms larger like a freight train approaching. It is like I can see my lunch menu while attending a snoozing morning meeting, or the state that the meeting will adjourn at 12 o'clock is determined by the state that people need to forage for food by noon. So we see the future coming, limited only by its resolution near the origin from which all states rise. Once it has been the present became the past, gone forever. We can say this about our daily life as well as about the renormalization. Every present state has a unique past, but the renormalization will not preserve it. It encodes only the future. Every present state has a predetermined future but the renormalization will not reveal it all at once. The future becomes so blurred near the origin that for all intents and purposes we may just as well think the future to be uncertain, holding out all the possibilities. The renormalization zero is the origin of everything. This is perhaps the ramification of the renormalization chaos, the effect of the sensitive dependence on initial conditions.

The fourth point is metacosmological. It jives with the parallel universe hypothesis and the anthropic principle. We have already seen from the proof above that the embedding of any dynamical system into the renormalization is not unique. It can be put near the fixed point at the end of every u-family of the renormalization. The u-family line never touch each other except at the origin. If you view the origin as if it were at infinity then these λ-expansive lines would appear to be parallel, all intersecting at the distant infinity. So every system has infinite copies, each inside a parallel universe. Each universe distinguishes itself from the others by a distinct expanding number, along the unstable-manifold of the u-fixed point around which the parallel universe organizes. There are

infinitely many expanding numbers, perhaps ours just makes our universe the way it is.

The last point is mystical. If I follow around every atom of my body I can in theory construct not one but infinitely many orbits inside the renormalization chaos space. I can also in theory construct not one but infinitely many dense orbits each connects all orbits in all universes. It seems we have known it all, solved all mysteries. It even appears that we have figured out who or what God is, having found out the dense orbit has infinitely many manifestations, more than enough for everyone to start their own religion, for however many times in their life time, and by Gödel's logic one is perfectly as good as any others in terms of provability, and that both the Newtonian and the Darwinian camps about God can be perfectly at peace at last. Anything that sounds this good always has a catch. Actually, it is a nuttier catch, a head-scratching type catch. If you show me the life history of a renormalization orbit I would not know what heck it is even though that orbit is of me but constructed by you. There is no way for me to tell, at least not now and perhaps never, what does such an orbit represent or mean, or if a dense orbit is currently near me. I would not even recognize myself let alone anything else. I only know I am somewhere inside that chaos square but I don't know where. Life will always be a mystery.