Universal Behavior in Nonlinear Systems

by Mitchell J. Feigenbaum

Universal numbers,

$\delta = 4.6692016...$

and

$\alpha = 2.502907875...$,

determine quantitatively the transition from smooth to turbulent or erratic behavior for a large class of nonlinear systems.

There exist in nature processes that can be described as complex or chaotic and processes that are simple or orderly. Technology attempts to create devices of the simple variety: an idea is to be implemented, and various parts executing orderly motions are assembled. For example, cars, airplanes, radios, and clocks are all constructed from a variety of elementary parts each of which, ideally, implements one ordered aspect of the device. Technology also tries to control or minimize the impact of seemingly disordered processes, such as the complex weather patterns of the atmosphere, the myriad whorls of turmoil in a turbulent fluid, the erratic noise in an electronic signal, and other such phenomena. It is the complex phenomena that interest us here.

When a signal is noisy, its behavior from moment to moment is irregular and has no simple pattern of prediction. However, if we analyze a sufficiently long record of the signal, we may find that signal amplitudes occur within narrow ranges a definite fraction of the time. Analysis of another record of the signal may reveal the same fraction. In this case, the noise can be given a *statistical* description. This means that while it is impossible to say what amplitude will appear next in succession, it is possible to estimate the probability or likelihood that the signal will attain some specified range of values. Indeed, for the last hundred years disorderly processes have been taken to be statistical (one has

given up asking for a precise causal prediction), so that the goal of a description is to determine what the probabilities are, and from this information to determine various behaviors of interest—for example, how air turbulence modifies the drag on an airplane.

We know that perfectly definite causal and *simple* rules can have statistical (or random) behaviors. Thus, modern computers possess "random number generators" that provide the statistical ingredient in a simulation of an erratic process. However, this generator does nothing more than shift the decimal point in a rational number whose repeating block is suitably long. Accordingly, it is possible to predict what the nth generated number will be. Yet, in a list of successive generated numbers there is such a seeming lack of order that all statistical tests will confer upon the numbers a pedigree of randomness. Technically, the term "pseudorandom" is used to indicate this nature. One now may ask whether the various complex processes of nature themselves might not be merely pseudorandom, with the full import of randomness, which is untestable, a historic but misleading concept. Indeed our purpose here is to explore this possibility. What will prove altogether remarkable is that some very simple schemes to produce erratic numbers behave *identically* to some of the erratic aspects of natural phenomena. More specifically, there is now cogent evidence that the problem of how a fluid changes over from smooth to turbulent

flow can be solved through its relation to the simple scheme described in this article. Other natural problems that can be treated in the same way are the behavior of a population from generation to generation and the noisiness of a large variety of mechanical, electrical, and chemical oscillators. Also, there is now evidence that various Hamiltonian systems—those subscribing to classical mechanics, such as the solar system—can come under this discipline.

The feature common to these phenomena is that, as some external parameter (temperature, for example) is varied, the behavior of the system changes from simple to erratic. More precisely, for some range of parameter values, the system exhibits an orderly *periodic* behavior; that is, the system's behavior reproduces itself every *period* of time T. Beyond this range, the behavior fails to reproduce itself after T seconds; it almost does so, but in fact it requires *two* intervals of T to repeat itself. That is, the period has *doubled* to 2T. This new periodicity remains over some range of parameter values until another critical parameter value is reached after which the behavior *almost* reproduces itself after 2T, but in fact, it now requires 4T for reproduction. This process of successive period doubling recurs continually (with the range of parameter values for which the period is 2^nT becoming successively smaller as n increases) until, at a certain value of the parameter, it has doubled *ad infinitum*, so that the behavior is no longer periodic. Period doubling is then a characteristic route for a system to follow as it changes over from simple periodic to complex aperiodic motion. All the phenomena mentioned above exhibit period doubling. In the limit of aperiodic behavior, there is a unique and hence *universal* solution common to all systems undergoing period doubling. This fact implies remarkable consequences. For a given system, if we

denote by $\wedge_n$ the value of the parameter at which its period doubles for the nth time, we find that the values $\wedge_n$ converge to $\wedge_\infty$ (at which the motion is aperiodic) *geometrically* for large n. This means that

$$\wedge_\infty - \wedge_n \propto \delta^{-n} \qquad (1)$$

for a fixed value of δ (the *rate* of onset of complex behavior) as n becomes large. Put differently, if we define

$$\delta_n \equiv \frac{\wedge_{n+1} - \wedge_n}{\wedge_{n+2} - \wedge_{n+1}} , \qquad (2)$$

δ_n (quickly) approaches the constant value δ. (Typically, δ_n will agree with δ to several significant figures after just a few period doublings.) What is quite remarkable (beyond the fact that there is always a geometric convergence) is that, for all systems undergoing this period doubling, the value of δ is *predetermined* at the universal value

$$\delta = 4.6692016 ... \quad . \qquad (3)$$

Thus, this definite number must appear as a natural rate in oscillators, populations, fluids, and all systems exhibiting a period-doubling route to turbulence! In fact, most measurable properties of *any* such system in this aperiodic limit now can be determined, in a way that essentially bypasses the details of the equations governing each specific system because the theory of this behavior is universal over such details. That is, so long as a system possesses certain *qualitative* properties that enable it to undergo this route to complexity, its *quantitative* properties are determined. (This result is analogous to the results of the modern theory of critical phenomena, where a few qualitative properties of the system undergoing a phase transition, notably the dimensionality, determine *universal* critical ex-

ponents. Indeed at a *formal* level the two theories are identical in that they are fixed-point theories, and the number δ, for example, can be viewed as a critical exponent.) Accordingly, it is sufficient to study the simplest system exhibiting this phenomenon to comprehend the general case.

Functional Iteration

A random number generator is an example of a simple iteration scheme that has complex behavior. Such a scheme generates the next pseudorandom number by a definite transformation upon the present pseudorandom number. In other words, a certain function is reevaluated successively to produce a sequence of such numbers. Thus, if f is the function and x_0 is a starting number (or "seed"), then x_0, x_1, ..., x_n, ..., where

$$x_1 = f(x_0)$$

$$x_2 = f(x_1)$$

$$.$$
$$.$$
$$.$$

$$x_{n+1} = f(x_n) \qquad (4)$$

$$.$$
$$.$$
$$.$$

is the sequence of generated pseudorandom numbers. That is, they are generated by *functional iteration*. The nth element in the sequence is

$$x_n = f(f(... f(f(x_0)) ...)) \equiv f^n(x_0) , \qquad (5)$$

where n is the total number of applications of f. [$f^n(x)$ is not the nth power of f(x); it is the nth *iterate* of f.] A property of iterates worthy of mention is

$$f^n(f^m(x)) = f^m(f^n(x)) = f^{m+n}(x) , \qquad (6)$$

since each expression is simply m + n applications of f. It is understood that

$$f^0(x) = x . \qquad (7)$$

It is also useful to have a symbol, $\circ$, for functional iteration (or composition), so that

$$f^n \circ f^m = f^m \circ f^n = f^{m+n} . \qquad (8)$$

Now f^n in Eq. (5) is itself a definite and computable function, so that x_n as a function of x_0 is known in principle.

If the function f is *linear* as, for example,

$$f(x) = ax \qquad (9)$$

for some constant a, it is easy to see that

$$f^n(x) = a^n x , \qquad (10)$$

so that, for this f,

$$x_n = a^n x_0 \qquad (11)$$

is the solution of the *recurrence relation* defined in Eq. (4),

$$x_{n+1} = a x_n . \qquad (12)$$

Should $|a| < 1$, then x_n geometrically converges to zero at the rate $1/a$. This example is special in that the linearity of f allows for the explicit computation of f^n.

We must choose a *nonlinear* f to generate a pseudorandom sequence of numbers. If we choose for our nonlinear f

$$f(x) = a - x^2 , \qquad (13)$$

then it turns out that f^n is a polynominal in x of order 2^n. This polynomial rapidly becomes unmanageably large; moreover, its coefficients are polynomials in a of order up to 2^{n-1} and become equally difficult to compute. Thus even if $x_0 = 0$, x_n is a polynomial in a of order 2^{n-1}. These polynomials are nontrivial as can be surmised from the fact that for certain

values of a, the sequence of numbers generated for almost all starting points in the range $(a - a^2, a)$ possess *all* the mathematical properties of a random sequence. To illustrate this, the figure on the cover depicts the iterates of a similar system in two dimensions:

$$x_{n+1} = y_n - x_n^2$$

$$y_{n+1} = a - x_n . \qquad (14)$$

Analogous to Eq. (4), a starting coordinate pair (x_0, y_0) is used in Eq. (14) to determine the next coordinate (x_1, y_1). Equation (14) is reapplied to determine (x_2, y_2) and so on. For some initial points, all iterates lie along a definite elliptic curve, whereas for others the iterates are distributed "randomly" over a certain region. It should be obvious that no explicit formula will account for the vastly rich behavior shown in the figure. That is, while the iteration scheme of Eq. (14) is trivial to specify, its nth iterate as a function of (x_0, y_0) is unavailable. Put differently, applying the simplest of *nonlinear* iteration schemes to itself sufficiently many times can create vastly complex behavior. Yet, precisely because the same operation is reapplied, it is conceivable that only a select few self-consistent patterns might emerge where the consistency is determined by the key notion of iteration and *not* by the particular function performing the iterates. These self-consistent patterns do occur in the limit of infinite period doubling and in a well-defined intricate organization that can be determined *a priori* amidst the immense complexity depicted in the cover figure.

The Fixed-Point Behavior of Functional Iterations

Let us now make a direct onslaught against Eq. (13) to see what it possesses. We want to know the behavior of the system after many iterations. As we

already know, high iterates of f rapidly become very complicated. One way this growth can be prevented is to have the first iterate of x_0 be precisely x_0 itself. Generally, this is impossible. Rather this condition *determines* possible x_0's. Such a self-reproducing point is called a *fixed point* of f. The sequence of iterates is then $x_0, x_0, x_0, ...$ so that the behavior is *static*, or if viewed as periodic, it has period 1.

It is elementary to determine the fixed points of Eq. (13). For future convenience we shall use a modified form of Eq. (13) obtained by a translation in x and some redefinitions:

$$f(x) = 4\lambda x (1 - x) , \qquad (15)$$

so that as λ is varied, $x = 0$ is always a fixed point. Indeed, the fixed-point condition for Eq. (15),

$$x^* = f(x^*) = 4\lambda x^* (1 - x^*) , \qquad (16)$$

gives as the two fixed points

$$x^* = 0, x_0^* = 1 - 1/4\lambda . \qquad (17)$$

The maximum value of f(x) in Eq. (15) is attained at $x = \frac{1}{2}$ and is equal to λ. Also, for $\lambda > 0$ and x in the interval $(0,1)$, f(x) is always positive. Thus, if λ is anywhere in the range $[0,1]$, then any iterate of any x in $(0,1)$ is also always in $(0,1)$. Accordingly, in all that follows we shall consider only values of x and λ lying between 0 and 1. By Eq. (16) for $0 \le \lambda < \frac{1}{4}$, only $x^* = 0$ is within range, whereas for $\frac{1}{4} \le \lambda \le 1$, both fixed points are within the range. For example, if we set $\lambda = \frac{1}{2}$ and we start at the fixed point $x_0^* = \frac{1}{2}$ (that is, we set $x_0 = \frac{1}{2}$), then $x_1 = x_2 = ... = \frac{1}{2}$; similarly if $x_0 = 0$, $x_1 = x_2 = ... = 0$, and the problem of computing the nth iterate is obviously trivial.

What if we choose an x_0 *not* at a fixed point? The easiest way to see what happens is to perform a graphical analysis. We graph $y = f(x)$ together with $y = x$.

Where the lines intersect we have $x = y = f(x)$, so that the intersections are precisely the fixed points. Now, if we choose an x_0 and plot it on the x-axis, the ordinate of $f(x)$ at x_0 is x_1. To obtain x_2, we must transfer x_1 to the x-axis before reapplying f. Reflection through the straight line $y = x$ accomplishes precisely this operation. Altogether, to iterate an initial x_0 successively,

1. move *vertically* to the graph of $f(x)$,
2. move *horizontally* to the graph of $y = x$, and
3. repeat steps 1, 2, etc.

Figure 1 depicts this process for $\lambda = \frac{1}{2}$. The two fixed points are circled, and the first several iterates of an arbitrarily chosen point x_0 are shown. What should be obvious is that if we start from any x_0 in $(0,1)$ ($x = 0$ and $x = 1$ excluded), upon continued iteration x_n will converge to the fixed point at $x = \frac{1}{2}$. No matter how close x_0 is to the fixed point at $x = 0$, the iterates diverge away from it. Such a fixed point is termed *unstable.* Alternatively, for almost all x_0 near enough to $x = \frac{1}{2}$ [in this case, all x_0 in $(0,1)$], the iterates converge towards $x = \frac{1}{2}$. Such a fixed point is termed *stable* or is referred to as an *attractor* of period 1.

Now, if we don't care about the *transient* behavior of the iterates of x_0, but only about some regular behavior that will emerge eventually, then knowledge of the stable fixed point at $x = \frac{1}{2}$ satisfies our concern for the *eventual* behavior of the iterates. In this restricted sense of eventual behavior, the existence of an attractor determines the solution *independently* of the initial condition x_0 provided that x_0 is within the *basin of attraction* of the attractor; that is, that it *is* attracted. The attractor satisfies Eq. (16), which is explicitly independent of x_0. This condition is the basic theme of universal behavior: if an attractor exists, the eventual behavior is independent of the starting point.
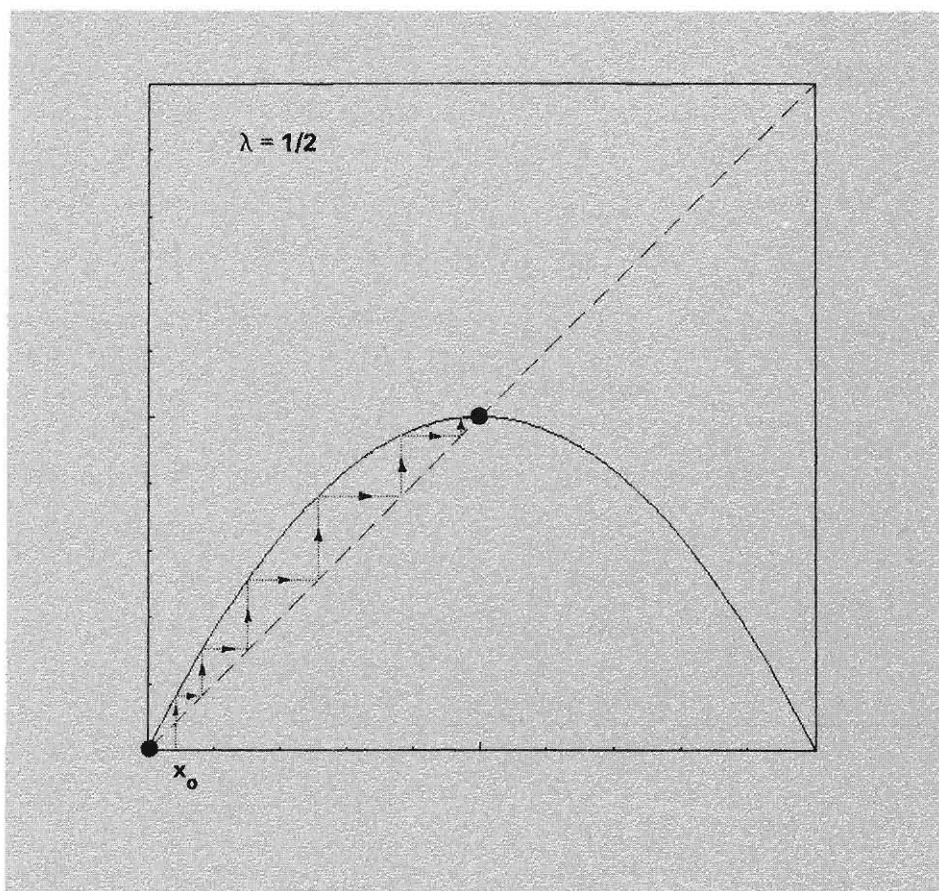

Fig. 1. Iterates of x_0 at $\lambda = 0.5$.

What makes $x = 0$ unstable, but $x = \frac{1}{2}$ stable? The reader should be able to convince himself that $x = 0$ is unstable because *the slope of $f(x)$ at $x = 0$ is greater than 1.* Indeed, if x^* is a fixed point of f and the derivative of f at x^*, $f'(x^*)$, is smaller than 1 in absolute value, then x^* is stable. If $|f'(x^*)|$ is greater than 1, then x^* is unstable. Also, only *stable* fixed points can account for the eventual behavior of the iterates of an arbitrary point.

We now must ask, "For what values of λ are the fixed points attracting?" By Eq. (15), $f'(x) = 4\lambda(1 - 2x)$ so that

$$f'(0) = 4\lambda \tag{18}$$

and

$$f'(x_0^*) = 2 - 4\lambda. \tag{19}$$

For $0 < \lambda < \frac{1}{4}$, only $x^* = 0$ is stable. At $\lambda = \frac{1}{4}$, $x_0^* = 0$ and $f'(x_0^*) = 1$. For $\frac{1}{4} < \lambda < \frac{3}{4}$, x^* is unstable and x_0^* is stable, while at $\lambda = \frac{3}{4}$, $f'(x^*_0) = -1$ and x_0^* also has become unstable. Thus, for $0 < \lambda < \frac{3}{4}$, the eventual behavior is known.

Period 2 from the Fixed Point

What happens to the system when λ is in the range $\frac{3}{4} < \lambda < 1$, where there are no attracting fixed points? We will see that as λ increases slightly beyond $\lambda = \frac{3}{4}$, f undergoes period doubling. That is, instead of having a stable cycle of period 1 corresponding to one fixed point, the system has a stable cycle of period 2; that is, the cycle contains two points. Since these two points are fixed points of the function f^2 (f applied twice) and since stability is determined by the slope of a function at its *fixed* points, we must now focus on f^2. First, we examine a graph of f^2 at λ just below $\frac{3}{4}$. Figures 2a and b show f and f^2, respectively, at $\lambda = 0.7$.

To understand Fig. 2b, observe first that, since f is symmetric about its maximum at $x = \frac{1}{2}$, f^2 is also symmetric about $x = \frac{1}{2}$. Also, f^2 must have a fixed point whenever f does because the second iterate of a fixed point is still that same point. The main ingredient that determines the period-doubling behavior of f as λ increases is the relationship of the slope of f^2 to the slope of f. This relationship is a consequence of the chain rule. By definition

$$x_2 = f^2(x_0) ,$$

where

$$x_1 = f(x_0), \; x_2 = f(x_1) .$$

We leave it to the reader to verify by the chain rule that

$$f^{2\prime}(x_0) = f'(x_0)f'(x_1) \qquad (20)$$

and

$$f^{n\prime}(x_0) = f'(x_0)f'(x_1) \dots f'(x_{n-1}) , \qquad (21)$$

an elementary result that determines period doubling. If we start at a fixed point of f and apply Eq. (20) to $x_0 = x^*$, so that $x_2 = x_1 = x^*$, then

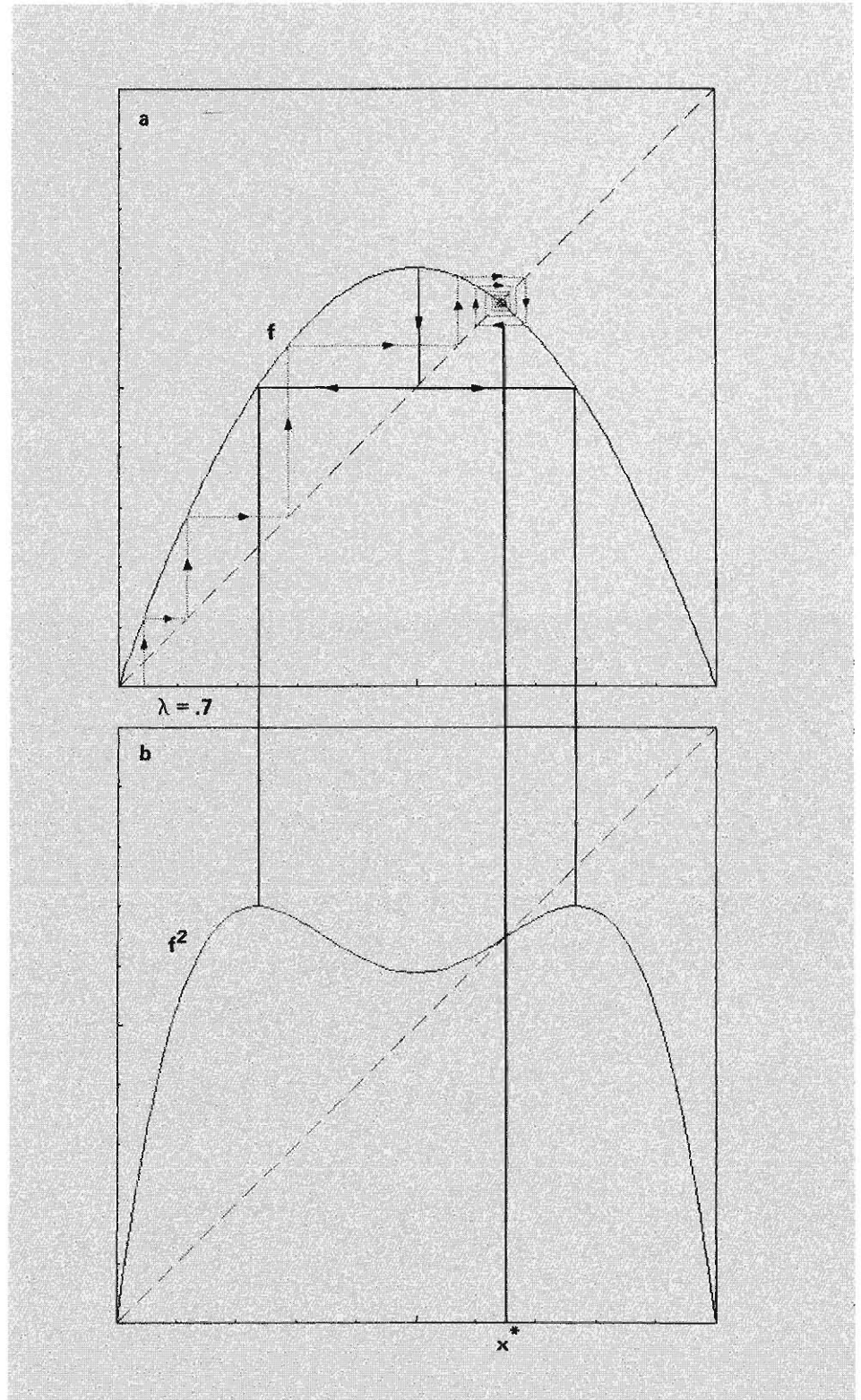$$f^{2\prime}(x^*) = f'(x^*)f'(x^*) = [f'(x^*)]^2 . \qquad (22)$$

Fig. 2. $\lambda = 0.7$. x^* is the stable fixed point. The extrema of f^2 are located in (a) by constructing the inverse iterates of $x = 0.5$.
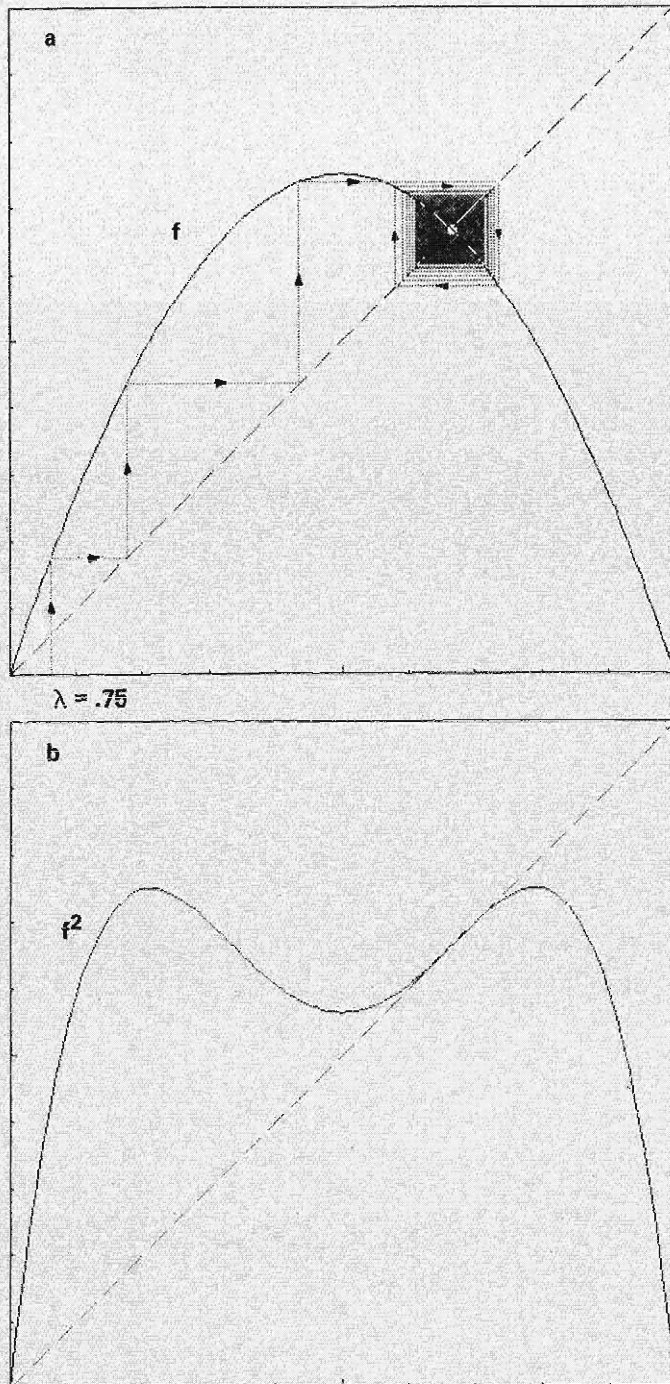
$\lambda = .75$

Since at $\lambda = 0.7$, $|f'(x^*)| < 1$, it follows from Eq. (22) that

$$0 < f^{2'}(x^*) < 1 \ .$$

Also, if we start at the extremum of f, so that $x_0 = \frac{1}{2}$ and $f'(x_0) = 0$, it follows from Eq. (21) that

$$f^{n'}(\tfrac{1}{2}) = 0 \qquad (23)$$

for all n. In particular, f^2 is extreme (and a minimum) at $\frac{1}{2}$. Also, by Eq. (20), f^2 will be extreme (and a maximum) at the x_0 that will iterate under f to $x = \frac{1}{2}$, since then $x_1 = \frac{1}{2}$ and $f'(x_1) = 0$. These points, the *inverses* of $x = \frac{1}{2}$, are found by going *vertically* down along $x = \frac{1}{2}$ to $y = x$ and then *horizontally* to $y = f(x)$. (Reverse the arrows in Fig. 1, and see Fig. 2a.) Since f has a maximum, there are *two* horizontal intersections and, hence, the two maxima of Fig. 2b. *The ability of f to have complex behaviors is precisely the consequence of its double-valued inverse*, which is in turn a reflection of its possession of an extremum. A monotone f, one that always increases, *always* has simple behaviors, whether or not the behaviors are easy to compute. A *linear* f is always monotone. The f's we care about always fold over and so are *strongly* nonlinear. This folding non-linearity gives rise to universality. Just as linearity in any system implies a definite method of solution, folding nonlinearity in any system also implies a definite method of solution. In fact folding non-linearity in the aperiodic limit of period doubling in any system is solvable, and many systems, such as various coupled nonlinear differential equations, possess this nonlinearity.

To return to Fig. 2b, as $\lambda \to \frac{3}{4}$ and the maximum value of f increases to $\frac{3}{4}$, $f'(x^*) \to -1$ and $f^{2'}(x^*) \to +1$. As λ increases beyond $\frac{3}{4}$, $|f'(x^*)| > 1$ and $f^{2'}(x^*) > 1$, so that f^2 must develop two new fixed points beyond those of f; that is, f^2 will cross $y = x$ at two more points. This transition is depicted in Figs. 3a and b for f and f^2, respectively, at $\lambda =$

Fig. 3. $\lambda = 0.75$. (a) depicts the slow convergence to the fixed point. f^2 osculates about the fixed point.

0.75, and similarly in Figs. 4a and b at $\lambda = 0.785$. (Observe the exceptionally slow convergence to x^* at $\lambda = 0.75$, where iterates approach the fixed point not geometrically, but rather with deviations from x^* inversely proportional to the square root of the number of iterations.) Since x_1^* and x_2^*, the new fixed points of f^2, are *not* fixed points of f, it must be that f sends one into the other:

$$x_1^* = f(x_2^*)$$

and

$$x_2^* = f(x_1^*) \ .$$

Such a *pair of points*, termed a *2-cycle*, is depicted by the limiting unwinding circulating square in Fig. 4a. Observe in Fig. 4b that the slope of f^2 is in excess of 1 at the fixed point of f and so is an unstable fixed point of f^2, while the two new fixed points have slopes smaller than 1, and so are *stable*; that is, every two iterates of f will have a point attracted toward x_1^* if it is sufficiently close to x_1^* or toward x_2^* if it is sufficiently close to x_2^*. This means that the sequence under f,

$$x_0, \ x_1, \ x_2, \ x_3, \ ... \ ,$$

eventually becomes arbitrarily close to the sequence

$$x_1^*, \ x_2^*, \ x_1^*, \ x_2^*, \ ... \ ,$$
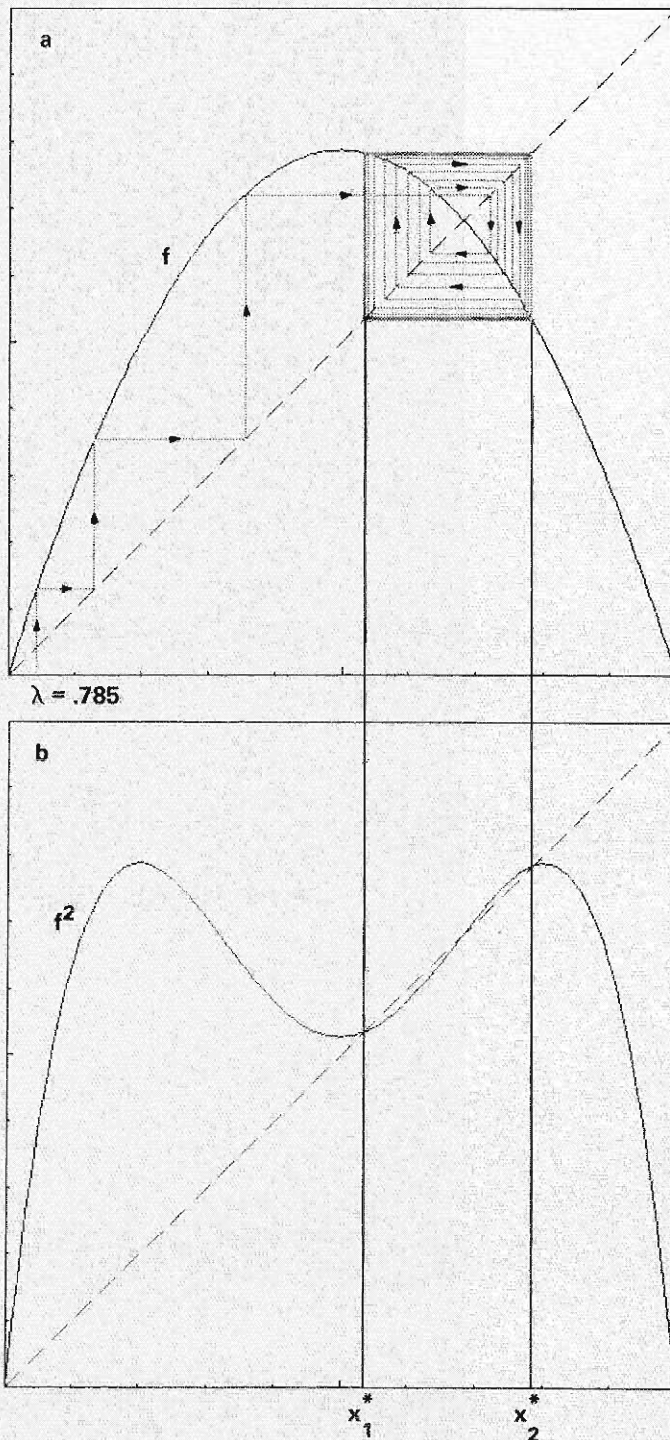
so that this is a stable 2-cycle, or an at-

Fig. 4. $\lambda = 0.785$. (a) shows the outward spiralling to a stable 2-cycle. The elements of the 2-cycle, x_1^* and x_2^*, are located as fixed points in (b).
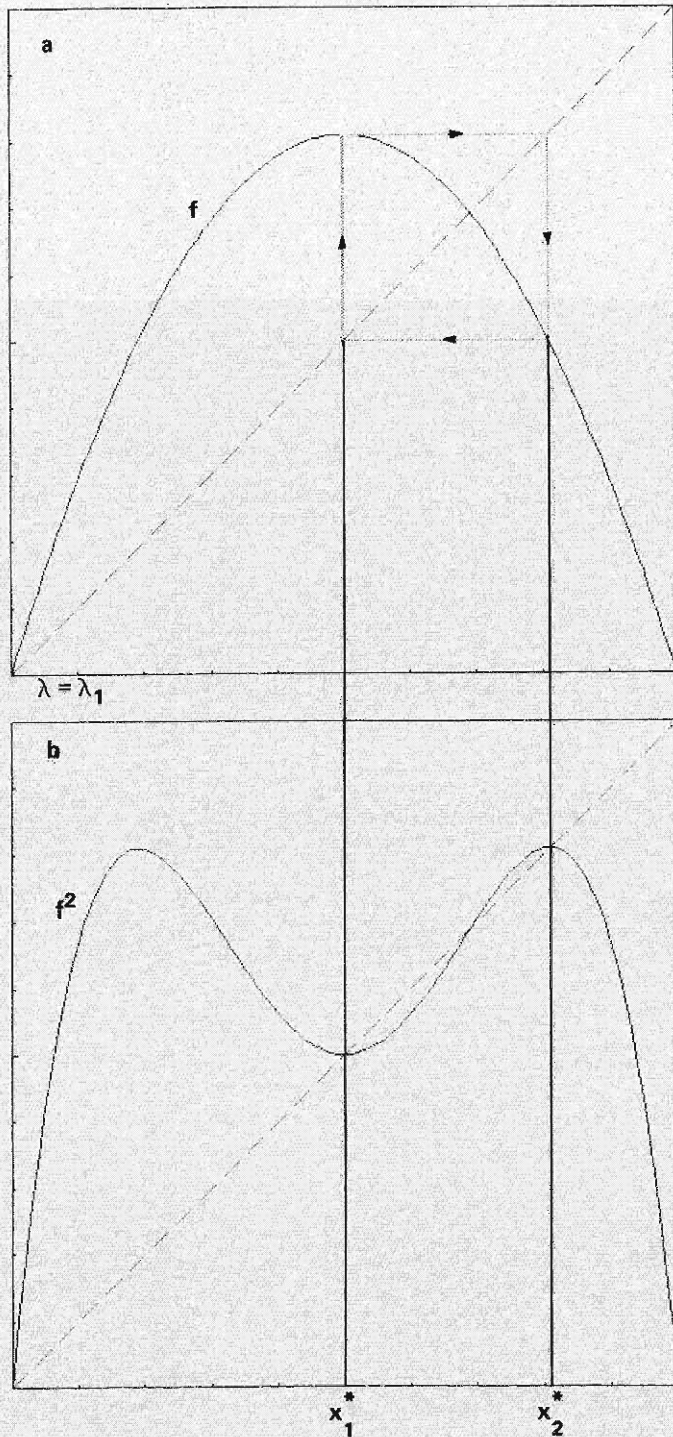
Fig. 5. $\lambda = \lambda_1$. *A superstable 2-cycle.* x_1^* *and* x_2^* *are at extrema of* f^2.

tractor of period 2. Thus, we have observed for Eq. (15) the first period doubling as the parameter λ has increased.

There is a point of paramount importance to be observed; namely, f^2 has the same slope at x_1^* and at x_2^*. This point is a direct consequence of Eq. (20), since if $x_0 = x_1^*$, then $x_1 = x_2^*$, and vice versa, so that the product of the slopes is the same. More generally, if x_1^*, x_2^*, ..., x_n^* is an n-cycle so that

$$x_{r+1}^* = f(x_r^*) \qquad r = 1, 2, ..., n - 1$$

and

$$x_1^* = f(x_n^*), \qquad\qquad (24)$$

then *each* is a fixed point of f^n with identical slopes:

$$x_r^* = f^n(x_r^*) \qquad r = 1, 2, ..., n \qquad (25)$$

and

$$f^{n\prime}(x_r^*) = f'(x_1^*) ... f'(x_n^*). \qquad (26)$$

From this observation will follow period doubling *ad infinitum*.

As λ is increased further, the minimum at $x = \frac{1}{2}$ will drop as the slope of f^2 through the fixed point of f increases. At some value of λ, denoted by λ_1, $x = \frac{1}{2}$ will become a fixed point of f^2. Simultaneously, the right-hand maximum will also become a fixed point of f^2. [By Eq. (26), both elements of the 2-cycle have slope 0.] Figures 5a and b depict the situation that occurs at $\lambda = \lambda_1$.

Period Doubling *Ad Infinitum*

We are now close to the end of this story. As we increase λ further, the minimum drops still lower, so that both x_1^* and x_2^* have negative slopes. At some parameter value, denoted by $\wedge_2$, the slope at *both* x_1^* and x_2^* becomes equal to -1. Thus at $\wedge_2$ the same situation has developed for f^2 as developed for f at $\wedge_1 = \frac{3}{4}$. This transitional case is depicted in Figs. 6a and b. Accordingly, just as the fixed point of f at $\wedge_1$ issued into being a 2-cycle, so too does *each* fixed point of f^2 at $\wedge_2$ create a 2-cycle, which in turn is a 4-cycle of f. That is, we have now encountered the second period doubling.

The manner in which we were able to follow the creation of the 2-cycle at $\wedge_1$ was to anticipate the presence of period 2, and so to consider f^2, which would resolve the cycle into a pair of fixed points. Similarly, to resolve period 4 into fixed points we now should consider f^4. Beyond being the fourth iterate of f, Eq. (8) tells us that f^4 can be computed from f^2:

$$f^4 = f^2 \circ f^2 .$$

From this point, we can abandon f itself, and take f^2 as the "fundamental" function. Then, just as f^2 was constructed by iterating f with itself we now iterate f^2 with itself. The manner in which f^2 reveals itself as being an iterate of f is the slope equality at the fixed points of f^2, which we saw imposed by the chain rule. Since the operation of the chain rule is "automatic," we actually needed to consider only the fixed point of f^2 nearest to $x = \frac{1}{2}$; the behavior of the other fixed point is slaved to it. Thus, at the level of f^4, we again need to focus on only the fixed point of f^4 nearest to $x = \frac{1}{2}$; the other *three* fixed points are similarly slaved to it. Thus, a recursive scheme has been unearthed. We now increase λ to λ_2, so that the fixed point of f^4 nearest to $x = \frac{1}{2}$ is again at $x = \frac{1}{2}$ with slope 0.
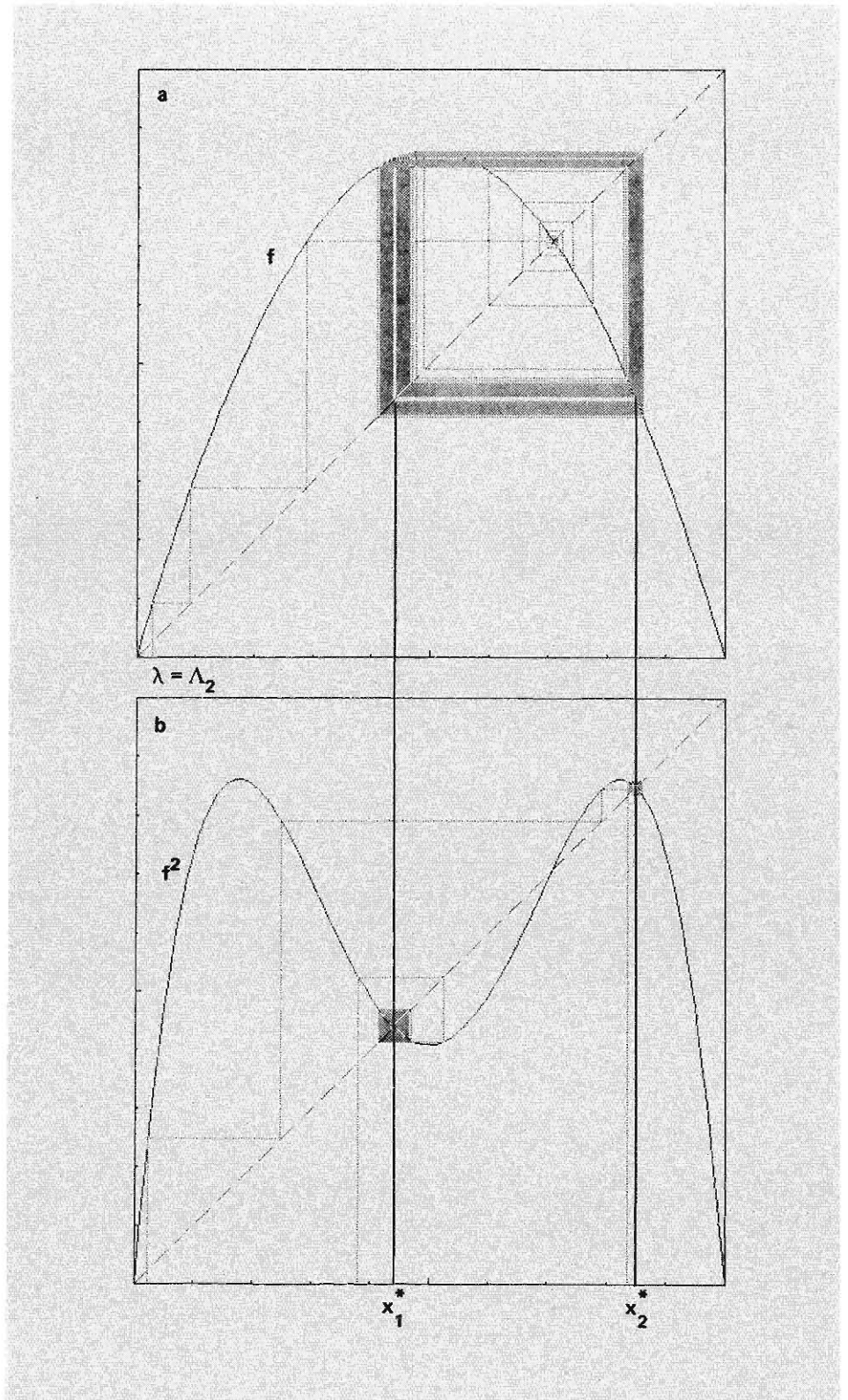

Fig. 6. $\lambda = \wedge_2$. x_1^* *and* x_2^* *in (b) have the same slow convergence as the fixed point in Fig. 3a.*
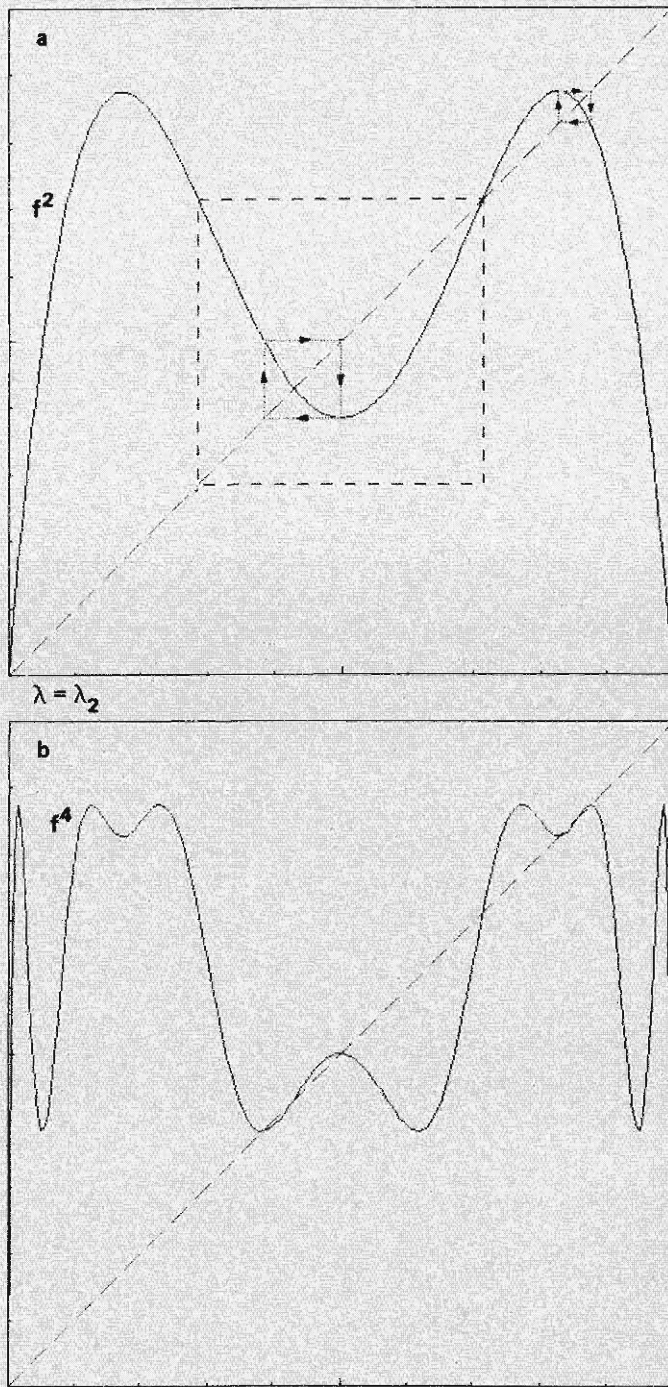
Fig. 7. $\lambda = \lambda_2$. A superstable 4-cycle. The region within the dashed square in (a) should be compared with all of Fig. 5a.

Figures 7a and b depict this situation for f^2 and f^4, respectively. When λ increases further, the maximum of f^4 at $x = \frac{1}{2}$ now moves up, developing a fixed point with negative slope. Finally, at $\wedge_3$ when the slope of this fixed point (as well as the other three) is again -1, each fixed point will split into a pair giving rise to an 8-cycle, which is now stable. Again, $f^8 = f^4 \bigcirc f^4$, and f^4 can be viewed as fundamental. We define λ_3 so that $x = \frac{1}{2}$ again is a fixed point, this time of f^8. Then at $\wedge_4$ the slopes are -1, and another period doubling occurs. Always,

$$f^{2^{n+1}} = f^{2^n} \bigcirc f^{2^n} . \tag{27}$$

Provided that a constraint on the range of λ does not prevent it from decreasing the slope at the appropriate fixed point past -1, this doubling must recur *ad infinitum*.

Basically, the mechanism that f^{2^n} uses to period double at $\wedge_{n+1}$ is the same mechanism that $f^{2^{n+1}}$ will use to double at $\wedge_{n+2}$. The function $f^{2^{n+1}}$ is constructed from f^{2^n} by Eq. (27), and similarly $f^{2^{n+2}}$ will be constructed from $f^{2^{n+1}}$. Thus, there is a definite operation that, by acting on functions, creates functions; in particular, the operation acting on f^{2^n} at $\wedge_{n+1}$, (or better, f^{2^n} at λ_n) will determine $f^{2^{n+1}}$ at λ_{n+1}. Also, since we need to keep track of f^{2^n} only in the interval including the fixed point of f^{2^n} closest to $x = \frac{1}{2}$ and since this interval becomes increasingly small as λ increases, the part of f that generates this region is also the restriction of f to an increasingly small interval about $x = \frac{1}{2}$. (Actually, slopes of f at points farther away also matter, but these merely set a "scale," which will be eliminated by a rescaling.) The behavior of f away from $x = \frac{1}{2}$ is immaterial to the period-doubling behavior, and in the limit of large n only the *nature of f's maximum* can matter. This means that in the infinite period-doubling limit, all functions with a quadratic extremum will have identical behavior. [$f''(\frac{1}{2}) \neq 0$ is the

13

generic circumstance.| Therefore, the operation on functions will have a *stable fixed point* in the space of functions, which will be the common universal limit of high iterates of any specific function. To determine this universal limit we must enlarge our scope vastly, so that the role of the starting point, x_0, will be played by an arbitrary *function*; the attracting fixed point will become a universal function obeying an equation implicating only itself. The role of the function in the equation $x_0 = f(x_0)$ now must be played by an *operation* that yields a new function when it is performed upon a function. In fact, the heart of this operation is the functional composition of Eq. (27). If we can determine the exact operator and actually can solve *its* fixed-point problem, we shall understand why a special number, such as δ of Eq. (3), has emerged independently of the specific system (the starting function) we have considered.

The Universal Limit of High Iterates

In this section we sketch the solution to the fixed-point problem. In Fig. 7a, a dashed square encloses the part of f^2 that we must focus on for all further period doublings. This square should be compared with the unit square that comprises all of Fig. 5a. If the Fig. 7a square is reflected through $x = \frac{1}{2}$, $y = \frac{1}{2}$ and then *magnified* so that the circulation squares of Figs. 4a and 5a are of equal size, we will have in each square a piece of a function that has the same kind of maximum at $x = \frac{1}{2}$ and falls to zero at the right-hand lower corner of the circulation square. Just as f produced this second curve of f^2 in the square as λ increased from λ_1 to λ_2, so too will f^2 produce another curve, which will be similar to the other two when it has been magnified suitably and reflected twice. Figure 8 shows this superposition for the first *five* such functions; at the resolution of the figure, observe that the last three

A DISCOVERY

The inspiration for the universality theory came from two sources. First, in 1971 N. Metropolis, M. Stein, and P. Stein (all in the LASL Theoretical Division) discovered a curious property of iterations: as a parameter is varied, the behavior of iterates varies in a fashion independent of the particular function iterated. In particular for a large class of functions, if at some value of the parameter a certain cycle is stable, then as the parameter increases, the cycle is replaced successively by cycles of doubled periods. This period doubling continues until an infinite period, and hence erratic behavior, is attained.

Second, during the early 1970s, a scheme of mathematics called dynamical system theory was popularized, largely by D. Ruelle, with the notion of a "strange attractor." The underlying questions addressed were (1) how could a purely causal equation (for example, the Navier-Stokes equations that describe fluid flow) come to demonstrate highly erratic or statistical properties and (2) how could these statistical properties be computed. This line of thought merged with the iteration ideas, and the limiting infinite "cycles" of iteration systems came to be viewed as a possible means to comprehend turbulence. Indeed, I became inspired to study the iterates of functions by a talk on such matters by S. Smale, one of the creators of dynamical system theory, at Aspen in the summer of 1975.

My first effort at understanding this problem was through the complex analytic properties of the generating function of the iterates of the quadratic map

$$x_{n+1} = \lambda x_n (1 - x_n) .$$

This study clarified the mechanism of period doubling and led to a rather different kind of equation to determine the values of λ at which the period doubling occurs. The new equations were intractable, although approximate solutions seemed possible. Accordingly, when I returned from Aspen, I numerically determined some parameter values with an eye toward discerning some patterns. At this time I had never used a large computer—in fact my sole computing power resided in a programmable pocket calculator. Now, such machines are very slow. A particular parameter value is obtained iteratively (by Newton's method) with each step of iteration requiring 2^n iterates of the map. For a 64-cycle, this means 1 minute per step of Newton's method. At the same time as n increased, it became an increasingly more delicate matter to locate the desired solution. However, I immediately perceived the λ_n's were converging geometrically. This enabled me to predict the next value with increasing accuracy as n increased, and so required just one step of Newton's method to obtain the desired value. To the best of my knowledge, this observation of geometric convergence has never been made independently, for the sim-

ple reason that the solutions have always been performed automatically on large and fast computers!

That a geometric convergence occurred was already a surprise. I was interested in this for two reasons: first, to gain insight into my theoretical work, as already mentioned, and second, because a convergence rate is a number invariant under all smooth transformations, and so of mathematical interest. Accordingly, I spent a part of a day trying to fit the convergence rate value, 4.669, to the mathematical constants I knew. The task was fruitless, save for the fact that it made the number memorable.

At this point I was reminded by Paul Stein that period doubling isn't a unique property of the quadratic map, but also occurs, for example, in

$$x_{n+1} = \lambda \sin \pi x_n .$$

However, my generating function theory rested heavily on the fact that the nonlinearity was simply quadratic and not transcendental. Accordingly, my interest in the problem waned.

Perhaps a month later I decided to determine the λ's in the transcendental case numerically. This problem was even slower to compute than the quadratic one. Again, it became apparent that the λ's converged geometrically, and altogether amazingly, the convergence rate was the same 4.669 that I remembered by virtue of my efforts to fit it.

Recall that the work of Metropolis, Stein, and Stein showed that precise qualitative features are independent of the specific iterative scheme. Now I learned that precise quantitative features also are independent of the specific function. This discovery represents a complete inversion of accustomed ritual. Usually one relies on the fact that similar equations will have qualitatively similar behavior, but quantitative predictions depend on the details of the equations. The universality theory shows that qualitatively similar equations have the identical *quantitative* behavior. For example, a system of differential equations naturally determines certain maps. The computation of the actual analytic form of the map is generally well beyond present mathematical methods. However, should the map exhibit period doubling, then precise quantitative results are available from the universality theory because the theory applies independently of which map it happens to be. In particular, certain fluid flows have now been experimentally observed to become turbulent through period doubling (subharmonic bifurcations). From this one fact we know that the universality theory applies—and indeed correctly determines the precise way in which the flow becomes turbulent, without any reference to the underlying Navier-Stokes equations.

curves are coincident. Moreover, the scale reduction that f^2 will determine for f^4 is based solely on the functional composition, so that if these curves for f^{2^n}, $f^{2^{n+1}}$, *converge* (as they obviously do in Fig. 8), the scale reduction from level to level will *converge to a definite constant*. But the width of each circulation square is just the distance between $x = \frac{1}{2}$ when it is a fixed point of f^{2^n} and the fixed point of f^{2^n} next nearest to $x = \frac{1}{2}$ (Figs. 7a and b). That is, asymptotically, *the separation of adjacent elements of period-doubled attractors is reduced by a constant value from one doubling to the next*. Also from one doubling to the next, this next nearest element *alternates* from one side of $x = \frac{1}{2}$ to the other. Let d_n denote the algebraic distance from $x = \frac{1}{2}$ to the nearest element of the attractor cycle of period 2^n, in the 2^n-cycle at λ_n. A positive number α scales this distance down in the 2^{n+1}-cycle at λ_{n+1}:

$$\frac{d_n}{d_{n+1}} \sim -\alpha . \tag{28}$$

But since rescaling is determined only by functional composition, there is some function that composed with itself will *reproduce* itself reduced in scale by $-\alpha$. The function has a quadratic maximum at $x = \frac{1}{2}$, is symmetric about $x = \frac{1}{2}$, and can be scaled by hand to equal 1 at $x = \frac{1}{2}$. Shifting coordinates so that $x = \frac{1}{2} \rightarrow x = 0$, we have

$$-\alpha g(g(x/\alpha)) = g(x) . \tag{29}$$

Substituting $g(0) = 1$, we have

$$g(1) = -\frac{1}{\alpha} . \tag{30}$$

Accordingly, Eq. (29) is a definite equation for a function g depending on x through x^2 and having a maximum of 1 at $x = 0$. There is a unique smooth solution to Eq. (29), which determines

$$\alpha = 2.502907875 \dots . \tag{31}$$

Knowing α, we can predict through Eq. (28) a definite scaling law binding on the iterates of any scheme possessing period doubling. The law has, indeed, been amply verified experimentally. By Eq. (29), we see that the relevant operation upon functions that underlies period doubling is functional composition followed by magnification, where the magnification is determined by the fixed-point condition of Eq. (29) with the function g the fixed point in this space of functions. However, Eq. (29) does not describe a stable fixed point because we have not incorporated in it the parameter increase from λ_n to λ_{n+1}. Thus, g is not the limiting function of the curves in the circulation squares, although it is intimately related to that function. The full theory is described in the next section. Here we merely state that we can determine the limiting function and thereby can *determine the location of the actual elements of limiting 2^n-cycles*. We also have established that g is an unstable fixed point of functional composition, where the rate of divergence away from g is precisely δ of Eq. (3) and so is computable. Accordingly, there is a full theory that determines, in a precise quantitative way, the aperiodic limit of functional iterations with an *unspecified* function f.
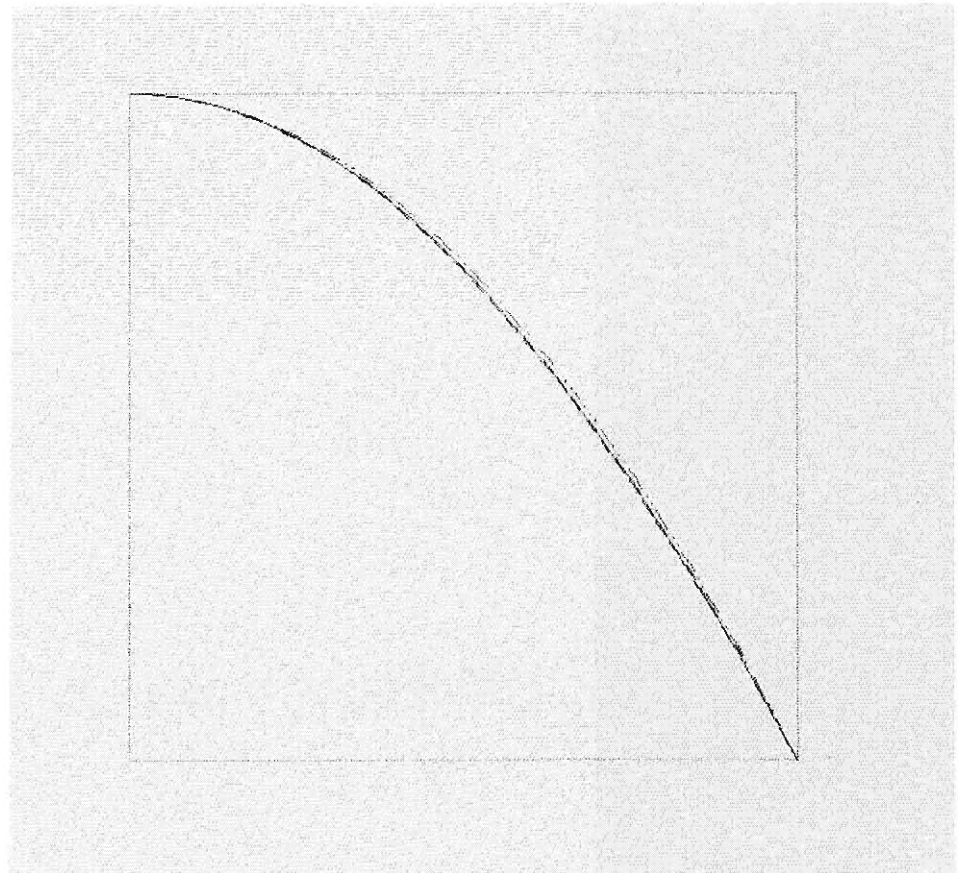

Fig. 8. The superposition of the suitably magnified dotted squares of $f^{2^{n-1}}$ at λ_n (as in Figs. 5a, 7a, ...).

Some Details of the Full Theory

Returning to Eq. (28), we are in a position to describe theoretically the universal scaling of high-order cycles and the convergence to a universal limit. Since d_n is the distance between $x = \frac{1}{2}$ and the element of the 2^n-cycle at λ_n nearest to $x = \frac{1}{2}$ and since this nearest element is the 2^{n-1} iterate of $x = \frac{1}{2}$ (which is true because these two points were coincident before the nth period doubling began to split them apart), we have

$$d_n = f^{2^{n-1}} (\lambda_n, \tfrac{1}{2}) - \tfrac{1}{2} \,. \qquad (32)$$

For future work it is expedient to perform a coordinate translation that moves $x = \frac{1}{2}$ to $x = 0$. Thus, Eq. (32) becomes

$$d_n = f^{2^{n-1}} (\lambda_n, 0) \,. \qquad (33)$$

Equation (28) now determines that the rescaled distances,

$$r_n \equiv (-\alpha)^n \, d_{n+1} \,.$$

will converge to a definite finite value as $n \to \infty$. That is,

$$\ell\text{im} \; (-\alpha)^n f^{2^n} (\lambda_{n+1}, 0) \qquad (34)$$
$$\scriptstyle n \to \infty$$

must exist if Eq. (28) holds.

However, from Fig. 8 we know something stronger than Eq. (34). When the nth iterated function is *magnified* by $(-\alpha)^n$, it converges to a definite function. Equation (34) is the value of this function at $x = 0$. After the magnification, the convergent functions are given by

$$(-\alpha)^n f^{2^n} (\lambda_{n+1}, x/(-\alpha)^n) \,.$$

Thus,

$$g_1(x) \equiv \ell\text{im} (-\alpha)^n f^{2^n} (\lambda_{n+1}, x/(-\alpha)^n) \qquad (35)$$
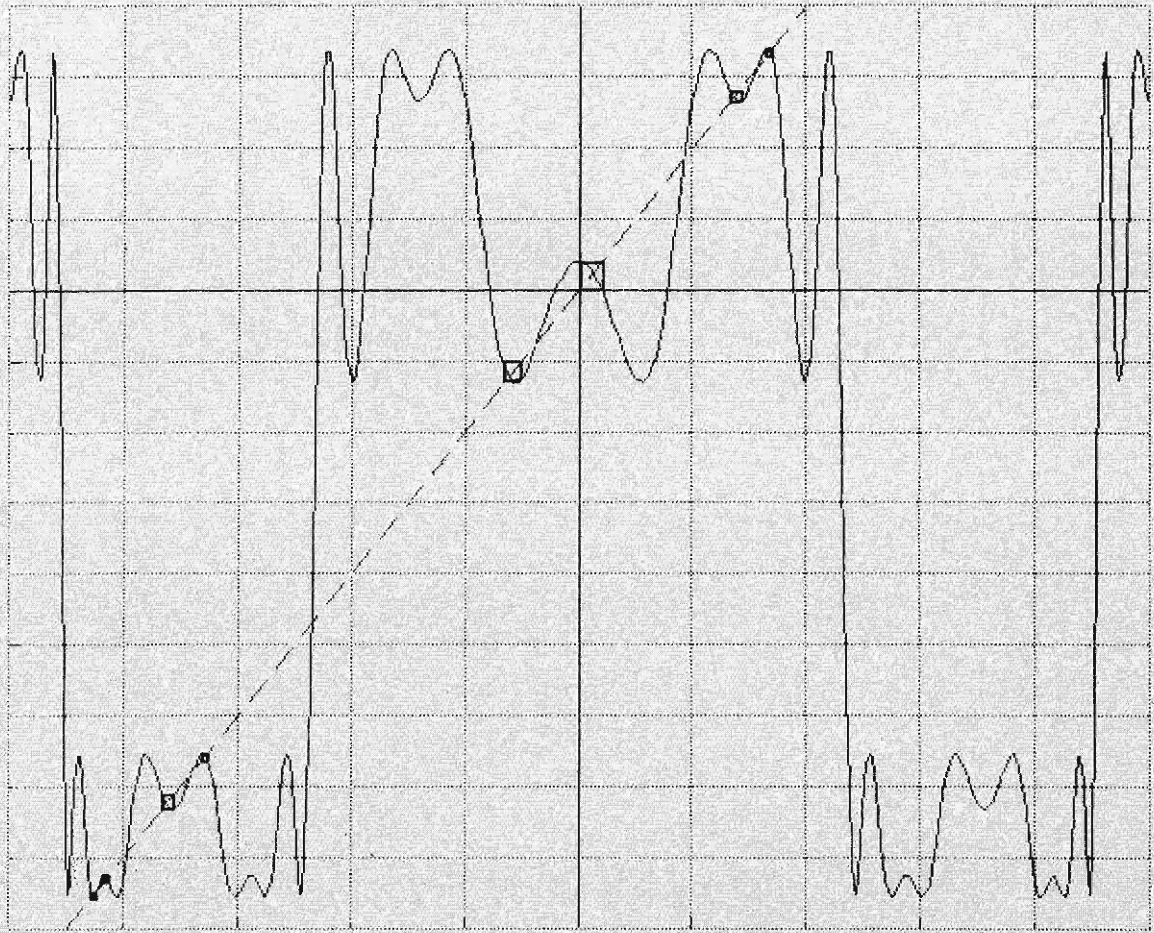$$\scriptstyle n \to \infty$$

Fig. 9. The function g_1. The squares locate cycle elements.

is the limiting function inscribed in the square of Fig. 8. The function $g_1(x)$ is, by the argument of the restriction of f to increasingly small intervals about its maximum, the *universal* limit of all iterates of all f's with a quadratic extremum. Indeed, it is numerically easy to ascertain that g_1 of Eq. (35) is always the same function independent of the f in Eq. (32).

What is this universal function good for? Figure 5a shows a crude approximation of g_1 [n = 0 in the limit of Eq. (35)], while Fig. 7a shows a better approximation (n = 1). In fact, the extrema of g_1 near the fixed points of g_1 support circulation squares each of which contains two points of the cycle. (The two squares shown in Fig. 7a locate the four elements of the cycle.) That is, g_1 determines the location of elements of high-

order 2^n-cycles near $x = 0$. Since g_1 is *universal*, we now have the amazing result that the location of the actual elements of highly doubled cycles is universal! The reader might guess this is a *very* powerful result. Figure 9 shows g_1 out to x sufficiently large to have 8 circulation squares, and hence locates the 15 elements of a 2^n-cycle nearest to $x = 0$. Also, the universal value of the scaling parameter α, obtained numerically, is

$$\alpha = 2.502907875 \ldots . \tag{36}$$

Like δ, α is a number that can be *measured* [through an experiment that observes the d_n of Eq. (28)] in any phenomenon exhibiting period doubling.

If g_1 is universal, then of course its iterate g_1^2 also is universal. Figure 7b

depicts an early approximation to this iterate. In fact, let us define a new universal function g_0, obtained by scaling g_1^2:

$$g_0(x) \equiv -\alpha g_1^2(-x/\alpha) . \tag{37}$$

(Because g_1 is universal and the iterates of our quadratic function are all symmetric in x, both g_1 and g_0 are symmetric functions. Accordingly, the minus sign within g_1^2 can be dropped with impunity.) From Eq. (35), we now can write

$$g_0(x) = \lim_{n \to \infty}(-\alpha)^n f^{2^n}(\lambda_n, x/(-\alpha)^n) . \tag{38}$$

[We introduced the scaling of Eq. (37) to provide one power of α per period doubling, since each successive iterate of f^{2^n}

reduces the scale by $\alpha|$.

In fact, we can generalize Eqs. (35) and (38) to a *family* of universal functions g_r:

$$g_r(x) = \ell im_{n \to \infty}(-\alpha)^n f^{2^n}(\lambda_{n+r}, x/(-\alpha)^n) . \quad (39)$$

To understand this, observe that g_0 locates the cycle elements as the fixed points of g_0 at extrema; g_1 locates the same elements by determining two elements per extremum. Similarly, g_r determines 2^r elements about each extremum near a fixed point of g_r. Since each f^{2^n} is always magnified by $(-\alpha)^n$ for each r, the scales of all g_r are the same. Indeed, g_r for $r > 1$ looks like g_1 of Fig. 9, except that each extremum is slightly higher, to accommodate a 2^r-cycle. Since each extremum must grow by convergently small amounts to accommodate higher and higher 2^r-cycles, we are led to conclude that

$$g(x) = \ell im_{r \to \infty} g_r(x) \quad (40)$$

must exist. By Eq. (39),

$$g(x) = \ell im_{n \to \infty}(-\alpha)^n f^{2^n}(\lambda_\infty, x/(-\alpha)^n) . \quad (41)$$

Unlike the functions g_r, $g(x)$ is obtained as a limit of f^{2^n}'s at a *fixed value* of λ. Indeed, this is the special significance of λ_∞; it is an isolated value of λ at which repeated iteration and magnification lead to a convergent function.

We now can write the equation that g satisfies. Analogously to Eq. (37), it is easy to verify that all g_r are related by

$$g_{r-1}(x) = -\alpha g_r(g_r(-x/\alpha)) . \quad (42)$$

By Eq. (40), it follows that g satisfies

$$g(x) = -\alpha g(g(x/\alpha)) . \quad (43)$$

The reader can verify that Eq. (43) is in-variant under a magnification of g. Thus, the theory has nothing to say about absolute scales. Accordingly, we must fix this by hand by setting

$$g(0) = 1 . \quad (44)$$

Also, we must specify the nature of the maximum of g at $x = 0$ (for example, quadratic). Finally, since g is to be built by iterating a $- x^2$, it must be both smooth and a function of x through x^2. With these specifications, Eq. (43) has a *unique* solution. By Eqs. (44) and (43),

$$g(0) = 1 = -\alpha g(g(0)) = -\alpha g(1) ,$$

so that

$$\alpha = -1/g(1) . \quad (45)$$

Accordingly, Eq. (43) determines α together with g.

Let us comment on the nature of Eq. (43), a so-called functional equation. Because g is smooth, if we know its value at a finite number of points, we know its value to some approximation on the interval containing these points by any sufficiently smooth interpolation. Thus, to some degree of accuracy, Eq. (43) can be replaced by a finite coupled system of nonlinear equations. Exactly then, Eq. (43) is an infinite-dimensional, nonlinear vector equation. Accordingly, we have obtained the solution to one-dimensional period doubling through our infinite-dimensional, explicitly universal problem. Equation (43) must be infinite-dimensional because it must keep track of the infinite number of cycle elements demanded of any attempt to solve the period-doubling problem. Rigorous mathematics for equations like Eq. (43) is just beyond the boundary of present mathematical knowledge.

At this point, we must determine two items. First, where is δ? Second, how do we obtain g_1, the real function of interest for locating cycle elements? The two problems are part of one question. Equation (42) is itself an iteration scheme. However, unlike the elements in Eq. (4), the elements acted on in Eq. (42) are *functions*. The analogue of the function of f in Eq. (4) is the operation in function space of functional composition followed by a magnification. If we call this operation T, and an element of the function space ψ, Eq. (42) gives

$$T|\psi|(x) = -\alpha\psi^2(-x/\alpha) . \quad (46)$$

In terms of T, Eq. (42) now reads

$$g_{r-1} = T|g_r| , \quad (47)$$

and Eq. (43) reads

$$g = T|g| . \quad (48)$$

Thus, g is precisely the fixed point of T. Since g is the limit of the sequence g_r, we can obtain g_r for large r by linearizing T about its fixed point g. Once we have g_r in the linear regime, the exact repeated application of T by Eq. (47) will provide g_1. Thus, we must investigate the stability of T at the fixed point g. However, it is obvious that T is *unstable* at g: for a large enough r, g_r is a point arbitrarily close to the fixed point g; by Eq. (47), successive iterates of g_r under T move away from g. How unstable is T? Consider a one-parameter family of functions f_λ, which means a "line" in the function space. For each f, there is an isolated parameter value λ_∞, for which repeated applications of T lead to convergence towards g |Eq. (41)|. Now, the function space can be "packed" with all the lines corresponding to the various f's. The set of all the points on these lines specified by the respective λ_∞'s determines a "surface" having the property that repeated applications of T to any point on it will converge to g. This is the surface of stability of T (the "stable manifold" of T through g). But through each point of this surface issues out the corresponding line, which is one-

dimensional since it is parametrized by a single parameter, λ. Accordingly, T is *unstable* in only *one* direction in function space. Linearized about g, this line of instability can be written as the one-parameter family

$$f_\lambda(x) = g(x) - \lambda h(x) , \qquad (49)$$

which passes through g (at $\lambda = 0$) and deviates from g along the unique direction h. But f_λ is just one of our transformations [Eq. (4)]! Thus, as we vary λ, f_λ will undergo period doubling, doubling to a 2^n-cycle at $\wedge_n$. By Eq. (41), λ_∞ for the family of functions f_λ in Eq. (49) is

$$\lambda_\infty = 0 . \qquad (50)$$

Thus, by Eq. (1)

$$\lambda_n \sim \delta^{-n} . \qquad (51)$$

Since applications of T by Eq. (47) iterate in the opposite direction (diverge away from g), it now follows that the rate of instability of T along h must be precisely δ.

Accordingly, we find δ and g_1 in the following way. First, we must linearize the operation T about its fixed point g. Next, we must determine the stability directions of the linearized operator. Moreover, we expect there to be precisely one direction of instability. Indeed, it turns out that infinitesimal deformations (conjugacies) of g determine *stable* directions, while a unique unstable direction, h, emerges with a stability rate (eigenvalue) precisely the δ of Eq. (3). Equation (49) at λ_r is precisely g_r for asymptotically large r. Thus g_r is known asymptotically, so that we have entered the sequence g_r and can now, by repeated use of Eq. (47), step down to g_1. All the ingredients of a full description of high-order 2^n-cycles now are at hand and evidently are universal.

Although we have said that the function g_1 universally locates cycle elements near $x = 0$, we must understand that it doesn't locate all cycle elements. This is possible because a finite distance of the scale of g_1 (for example, the location of the element nearest to $x = 0$) has been magnified by α^n for n diverging. Indeed, the distances from $x = 0$ of all elements of a 2^n-cycle, "accurately" located by g_1, are reduced by $-\alpha$ in the 2^{n+1}-cycle. However, it is obvious that some elements have no such scaling: because $f(0) = a_n$ in Eq. (13), and $a_n \rightarrow a_\infty$, which is a definite nonzero number, the distance from the origin of the element of the 2^n-cycle farthest to the right certainly has not been reduced by $-\alpha$ at each period doubling. This suggests that we must measure locations of elements on the far right with respect to the farthest right point. If we do this, we can see that these distances scale by α^2, since they are the images through the quadractic maximum of f at $x = 0$ of elements close to $x = 0$ scaling with $-\alpha$. In fact, if we image g_1 through the maximum of f (through a quadratic conjugacy), then we shall indeed obtain a new universal function that locates cycle elements near the right-most element. The correct description of a highly doubled cycle now emerges as one of universal local clusters.

We can state the scope of universality for the location of cycle elements precisely. Since $f(\lambda_1, x)$ exactly locates the two elements of the 2^1-cycle, and since $f(\lambda_1, x)$ is an approximation to g_1 [n $= 0$ in Eq. (35)], we evidently can locate both points exactly by appropriately sealing g_1. Next, near $x = 0$, $f^2(\lambda_2, x)$ is a better approximation to g_1 (suitably scaled). However, in general, the more accurately we scale g_1 to determine the smallest 2-cycle elements, the greater is the error in its determination of the right-most elements. Again, near $x = 0$, $f^4(\lambda_3, x)$ is a still better approximation to g_1. Indeed, the suitably scaled g_1 now can determine several points about $x = 0$ accurately, but determination of the right-

most elements is still worse. In this fashion, it follows that g_1, suitably scaled, can determine 2^r points of the 2^n-cycle near $x = 0$ for r $\ll$ n. If we focus on the neighborhood of one of these 2^r points at some definite distance from $x = 0$, then by Eq. (35) the larger the n, the larger the *scaled* distance of this region from $x = 0$, and so, the poorer the approximation of the location of fixed points in it by g_1. However, just as we can construct the version of g_1 that applies at the right-most cycle element, we also can construct the version of g_1 that applies at this chosen neighborhood. Accordingly, the universal description is set through an acceptable tolerance: if we "measure" f^{2^n} at some definite n, then we can use the actual location of the elements as foci for 2^n versions of g_1, each applicable at one such point. For all further period doubling, we determine the new cycle elements through the g_1's. In summary, the *more accurately we care to know the locations* of arbitrarily high-order cycle elements, the *more parameters we must measure* (namely, the cycle elements at some chosen order of period doubling). This is the sense in which the universality theory is asymptotic. Its ability to have serious predictive power is the fortunate consequence of the high convergence rate $\delta(\sim 4.67)$. Thus, typically after the first two or three period doublings, this asymptotic theory is already accurate to within several percent. If a period-doubling system is *measured* in its 4- or 8-cycle, its behavior throughout and symmetrically beyond the period-doubling regime also is determined to within a few percent.

To make precise dynamical predictions, we do not have to construct all the local versions of g_1; all we really need to know is the local *scaling* everywhere along the attractor. The scaling is $-\alpha$ at $x = 0$ and α^2 at the right-most element. But what is it at an arbitrary point? We can determine the scaling law if we order

elements not by their location on the x-axis, but rather by their order as iterates of x = 0. Because the time sequence in which a process evolves is precisely this ordering, the result will be of immediate and powerful predictive value. It is precisely this scaling law that allows us to compute the spectrum of the onset of turbulence in period-doubling systems.

What must we compute? First, just as the element in the 2^n-cycle nearest to x = 0 is the element halfway around the cycle from x = 0, the element nearest to an arbitrarily chosen element is precisely the one halfway around the cycle from it. Let us denote by $d_n(m)$ the distance between the mth cycle element (x_m) and the element nearest to it in a 2^n-cycle. |The d_n of Eq. (28) is $d_n(0)$|. As just explained,

$$d_n(m) = x_m - f^{2^{n-1}}(\lambda_n, x_m) . \qquad (52)$$

However, x_m is the mth iterate of $x_0 = 0$. Recalling from Eq. (6) that powers commute, we find

$$d_n(m) = f^m(\lambda_n, 0)$$
$$- f^m(\lambda_n, f^{2^{n-1}}(\lambda_n, 0)) . \qquad (53)$$

Let us, for the moment, specialize to m of the form 2^{n-r}, in which case

$$d_n(2^{n-r}) = f^{2^{n-r}}(\lambda_n, 0)$$
$$- f^{2^{n-r}}(\lambda_n, f^{2^{n-1}}(\lambda_n, 0))$$
$$= f^{2^{n-r}}(\lambda_{(n-r)+r}, 0)$$
$$- f^{2^{n-r}}(\lambda_{(n-r)+r}, f^{2^{n-1}}(\lambda_n, 0)) . \qquad (54)$$

For r ≪ n (which can still allow r ≫ 1 for n large), we have, by Eq. (39),

$$d_n(2^{n-r}) \sim (-\alpha)^{-(n-r)}[g_r(0)$$
$$- g_r((-\alpha)^{n-r} f^{2^{n-1}}(\lambda_n, 0))]$$

or

$$d_n(2^{n-r}) \sim (-\alpha)^{-(n-r)}[g_r(0)$$
$$- g_r((-\alpha)^{-r+1} g_1(0))] . \qquad (55)$$

The object we want to determine is the local scaling at the mth element, that is, the ratio of nearest separations at the mth iterate of x = 0, at successive values of n. That is, if the scaling is called σ,

$$\sigma_n(m) \equiv \frac{d_{n+1}(m)}{d_n(m)} . \qquad (56)$$

|Observe by Eq. (28), the definition of α, that $\sigma_n(0) \sim (-\alpha)^{-1}$.| Specializing again to m = 2^{n-r}, where r ≪ n, we have by Eq. (55)

$$\sigma(2^{n-r}) \sim \frac{g_{r+1}(0) - g_{r+1}((-\alpha)^{-r}g_1(0))}{g_r(0) - g_r((-\alpha)^{-r+1}g_1(0))} . \qquad (57)$$

Finally, let us rescale the axis of iterates so that all 2^{n+1} iterates are within a unit interval. Labelling this axis by t, the value of t of the mth element in a 2^n-cycle is

$$t_n(m) = m/2^n . \qquad (58)$$

In particular, we have

$$t_n(2^{n-r}) = 2^{-r} . \qquad (59)$$

Defining σ along the t-axis naturally as

$$\sigma(t_n(m)) \sim \sigma_n(m) \qquad (as\ n \to \infty) ,$$

we have by Eqs. (57) and (59),

$$\sigma(2^{-r-1}) = \frac{g_{r+1}(0) - g_{r+1}((-\alpha)^{-r}g_1(0))}{g_r(0) - g_r((-\alpha)^{-r+1}g_1(0))} . \qquad (60)$$

It is not much more difficult to obtain σ for all t. This is done first for rational t by writing t in its binary expansion:

$$t_{r_1 r_2 r_3 \cdots} = 2^{-r_1} + 2^{-r_2} + \dots .$$

In the 2^n-cycle approximation we require σ_n at the $2^{n-r_1} + 2^{n-r_2} + \dots$ iterate of the origin. But, by Eq. (8),

$$f^{2^{n-r_1} + 2^{n-r_2} + \cdots} = f^{2^{n-r_1}} \circ f^{2^{n-r_2}} \circ \dots .$$

It follows by manipulations identical to those that led from Eq. (54) to Eq. (60) that σ at such values of t is obtained by replacing the individual g_r terms in Eq. (60) by appropriate iterates of various g_r's.

There is one last ingredient to the computation of σ. We know that $\sigma(0) = -\alpha^{-1}$. We also know that $\sigma_n(1) \sim \alpha^{-2}$. But, by Eq. (59),

$$t_n(1) = 2^{-n} \to 0 .$$

Thus σ is discontinuous at t = 0, with $\sigma(0 - \varepsilon) = -\alpha^{-1}$ and $\sigma(0 + \varepsilon) = \alpha^{-2}(\varepsilon \to 0^+)$. Indeed, since $x_{2^{n-r}}$ is always very close to the origin, each of these points is imaged quadratically. Thus Eq. (60) actually determines $\sigma(2^{-r-1} - \varepsilon)$, while $\sigma(2^{-r-1} + \varepsilon)$ is obtained by replacing each numerator and denominator g_r by its square. The same replacement also is correct for each multi-g_r term that figures into σ at the binary expanded rationals.

Altogether, we have the following results. σ(t) can be computed for all t, and it is *universal* since its explicit computation depends only upon the universal functions g_r. σ is *discontinuous* at all the rationals. However, it can be established that the *larger* the number of terms in the binary expansion of a rational t, the smaller the discontinuity of σ. Lastly, as a finite number of iterates leaves t unchanged as n → ∞, σ must be *continuous* except at the rationals. Figure 10 depicts 1/σ(t). Despite the pathological nature of σ, the reader will observe that basically it is constant half the time at α^{-1} and half the time at α^{-2} for 0 < t < ½. In a succeeding approximation, it can be decomposed in each half into two slightly different quarters,
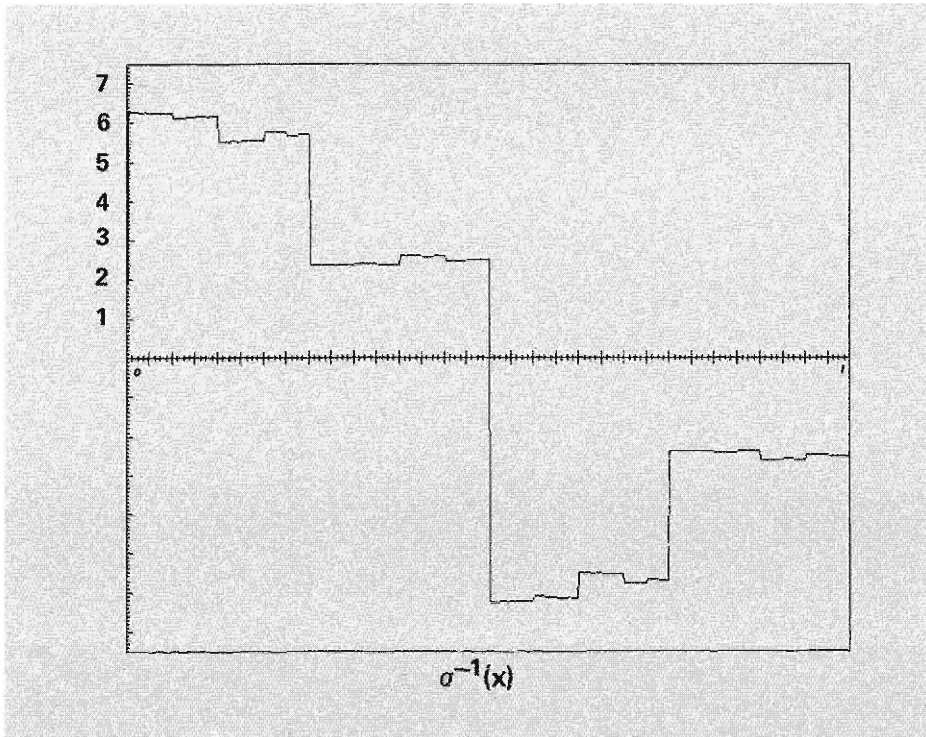
Fig. 10. The trajectory scaling function. Observe that $\sigma (x + 1/2) = - \sigma (x)$.
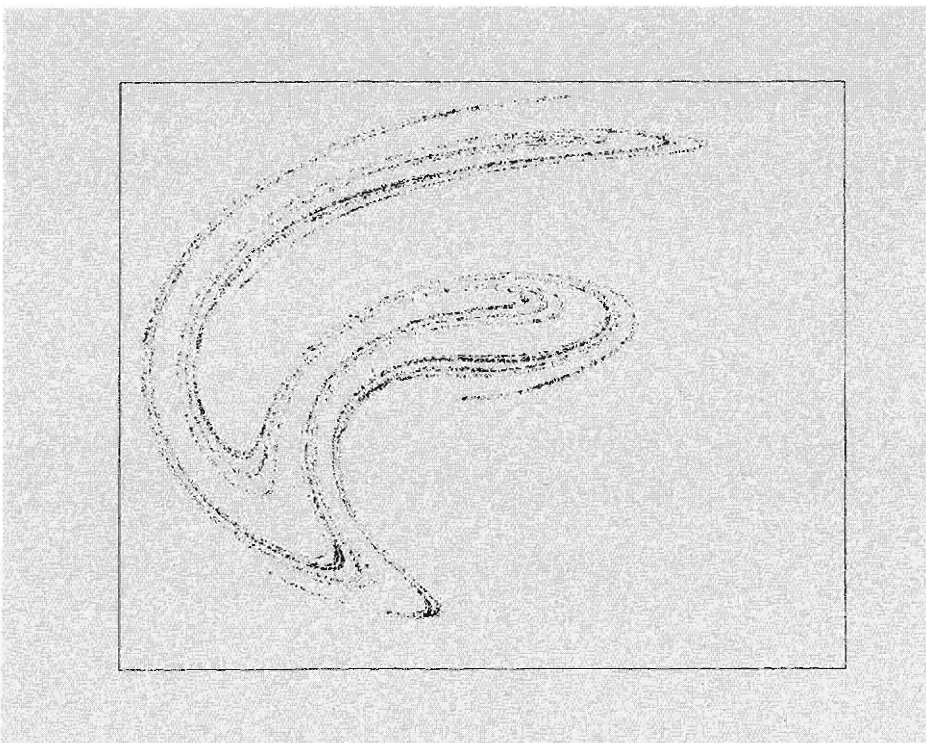

Fig. 11. The plotted points lie on the "strange attractor" of Duffing's equation.

and so forth. [It is easy to verify from Eq. (52) that σ is periodic in t of period 1, and has the symmetry

$$\sigma(t + \tfrac{1}{2}) = -\sigma(t) \ .$$

Accordingly, we have paid attention to its first half $0 < t < \tfrac{1}{2}$.] With σ we are at last finished with one-dimensional iterates per se.

Universal Behavior in Higher Dimensional Systems

So far we have discussed iteration in *one* variable; Eq. (15) is the prototype. Equation (14), an example of iteration in two dimensions, has the special property of preserving areas. A generalization of Eq. (14),

$$x_{n+1} = y_n - x_n^2$$

and

$$y_{n+1} = a + bx_n \qquad (61)$$

with $|b| < 1$, contracts areas. Equation (61) is interesting because it possesses a so-called *strange attractor*. This means an attractor (as before) constructed by folding a curve repeatedly upon itself (Fig. 11) with the consequent property that two initial points very near to one another are, in fact, very far from each other when the distance is measured along the folded attractor, which is the path they follow upon iteration. This means that after some iteration, they will soon be far apart in actual distance as well as when measured along the attractor. This general mechanism gives a system highly sensitive dependence upon its initial conditions and a truly statistical character: since very small differences in initial conditions are magnified quickly, unless the initial conditions are known to *infinite precision*, all known knowledge is eroded rapidly to future ignorance. Now, Eq. (61) enters

into the early stages of statistical behavior through period doubling. Moreover, δ of Eq. (3) is *again* the rate of onset of complexity, and α of Eq. (31) is again the rate at which the spacing of adjacent attractor points is vanishing. Indeed, the one-dimensional theory determines all behavior of Eq. (61) in the onset regime.

In fact, dimensionality is irrelevant. The same theory, the same numbers, etc. also work for iterations in N dimensions, provided that the system goes through period doubling. The basic process, wherever period doubling occurs *ad infinitum*, is functional composition from one level to the next. Accordingly, a modification of Eq. (29) is at the heart of the process, with composition on functions from N dimensions to N dimensions. Should the specific iteration function contract N-dimensional volumes (a dissipative process), then in general there is one direction of slowest contraction, so that after a number of iterations the process is effectively one-dimensional. Put differently, the one-dimensional solution to Eq. (29) is always a solution to its N-dimensional analogue. It is the relevant fixed point of the analogue if the iteration function is contractive.

Universal Behavior in Differential Systems

The next step of generalization is to include systems of differential equations. A prototypic equation is Duffing's oscillator, a driven damped anharmonic oscillator,

$$\ddot{x} + k\dot{x} + x^3 = b\sin 2\pi t . \qquad (62)$$

The periodic drive of period 1 determines a natural time step. Figure 12a depicts a period 1 attractor, usually referred to as a *limit cycle.* It is an attractor because, for a range of initial conditions, the solution to Eq. (62) settles down to the cycle. It is period 1 because it repeats the same curve in every period of the drive.
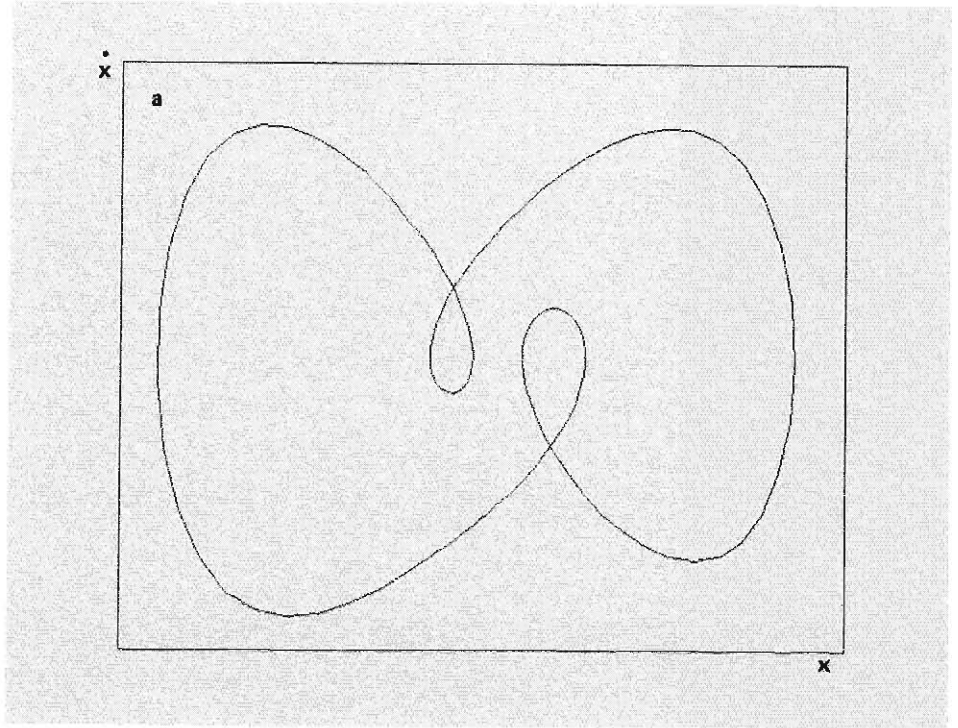

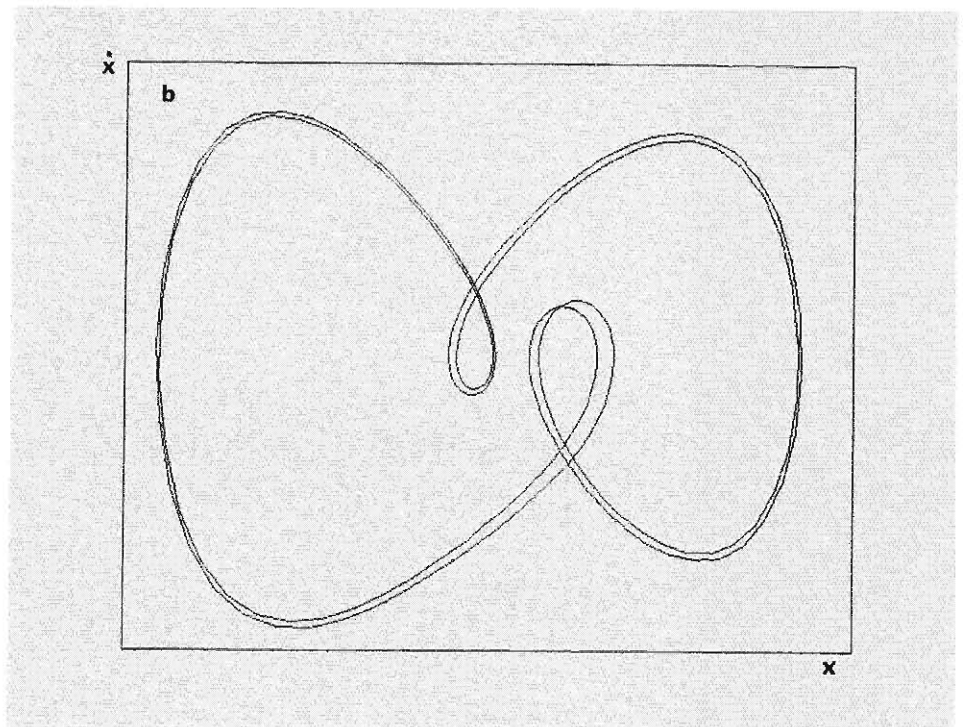Fig. 12a. The most stable 1-cycle of Duffing's equation in phase space $(x,\dot{x})$.

Fig. 12b. The most stable 2-cycle of Duffing's equation. Observe that it is two displaced copies of Fig. 12a.
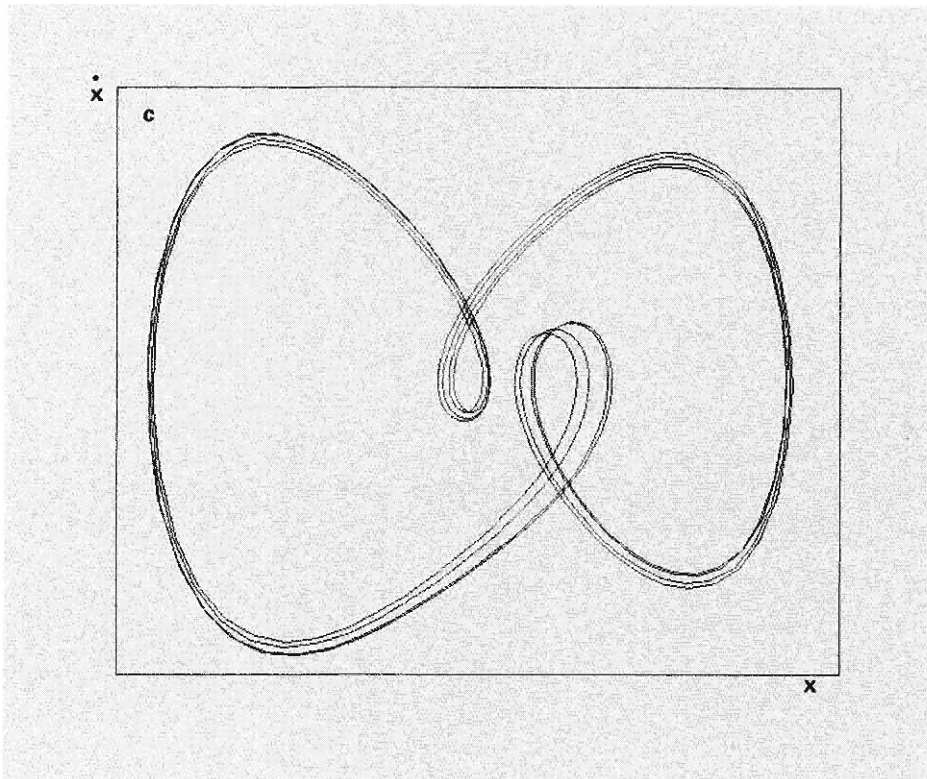
Fig. 12c. The most stable 4-cycle of Duffing's equation. Observe that the displaced copies of Fig. 12b have either a broad or a narrow separation.

Figures 12b and c depict attractors of periods 2 and 4 as the friction or damping constant k in Eq. (62) is reduced systematically. The parameter values k $= \lambda_0, \lambda_1, \lambda_2, ...,$ are the damping constants corresponding to the most stable 2^n-cycle in analogy to the λ_n of the one-dimensional functional iteration. Indeed, this oscillator's period doubles (at least numerically!) *ad infinitum*. In fact, by k $= \lambda_5,$ the δ_3 of Eq. (2) has converged to 4.69. Why is this? Instead of considering the entire trajectories as shown in Fig. 12, let us consider only where the trajectory point is located every 1 period of the drive. The 1-cycle then produces only one point, while the 2-cycle produces a pair of points, and so forth. This *time-one map* |if the trajectory point is $(x, \dot{x})$ now, where is it one period later?| is by virtue of the differential equation a smooth and invertible func-

tion in two dimensions. Qualitatively, it looks like the map of Eq. (61). In the present state of mathematics, little can be said about the analytic behavior of time-one maps; however, since our theory is universal, it makes no difference that we don't know the explicit form. We still can determine the complete quantitative behavior of Eq. (62) in the onset regime where the motion tends to aperiodicity. If we already know, by measurement, the precise form of the trajectory after a few period doublings, we can compute the form of the trajectory as the friction is reduced throughout the region of onset of complexity by carefully using the full power of the universality theory to determine the spacings of elements of a cycle.

Let us see how this works in some detail. Consider the time-one map of the

Duffing's oscillator in the superstable 2^n-cycle. In particular, let us focus on an element at which the scaling function σ (Fig. 10) has the value σ_0, and for which the next iterate of this element also has the scaling σ_0. (The element is not at a big discontinuity of σ.) It is then intuitive that if we had taken our time-one examination of the trajectory at values of time displaced from our first choice, we would have seen the same scaling σ_0 for this part of the trajectory. That is, the differential equations will extend the map-scaling function continuously to a function along the entire trajectory so that, if two successive time-one elements have scaling σ_0, then the entire stretch of trajectory over this unit time interval has scaling σ_0. In the last section, we were motivated to construct σ as a function of t along an interval precisely towards this end.

To implement this idea, the first step is to define the analogue of d_n. We require the spacing between the trajectory at time t and at time $T_n/2$ where the period of the system in the 2^n-cycle is

$$T_n \cong 2^n T_0 . \tag{63}$$

That is, we define

$$d_n(t) \equiv x_n(t) - x_n(t + T_n/2) . \tag{64}$$

(There is a d for each of the N variables for a system of N differential equations.) Since σ was defined as periodic of period 1, we now have

$$d_{n+1}(t) \sim \sigma(t/T_{n+1})d_n(t) . \tag{65}$$

The content of Eq. (65), based on the n-dependence arising solely through the T_n in σ, and not on the detailed form of σ, already implies a strong scaling prediction, in that the ratio

$$\frac{d_{n+1}(t)}{d_n(t)} ,$$

when plotted with t scaled so that $T_n =$

I, is a function *independent* of n. Thus if Eq. (65) is true for *some* σ, whatever it might be, then knowing $x_n(t)$, we can compute $d_n(t)$ and from Eq. (65) $d_{n+1}(t)$. As a consequence of periodicity, Eq. (64) for $n \rightarrow n + 1$ can be solved for $x_{n+1}(t)$ (through a Fourier transform). That is, if we have measured any chosen coordinate of the system in its 2^n-cycle, we can compute its time dependence in the 2^{n+1}-cycle. Because this procedure is recursive, we can compute the coordinate's evolution for all higher cycles through the infinite period-doubling limit. If Eq. (65) is true and σ not known, then by measurement at a 2^n-cycle and at a 2^{n+1}-cycle, σ could be *constructed* from Eq. (65), and hence all higher order doublings would again be determined. Accordingly, Eq. (65) is a very powerful result. However, we know much more. The universality theory tells us that period doubling is universal and that there is a *unique* function σ which, indeed, we have computed in the previous section. Accordingly, by *measuring* x(t) in some chosen 2^n-cycle (the higher the n, the more the number of effective parameters to be determined empirically, and the more precise are the predictions), we now can compute the entire evolution of the system on its route to turbulence.

How well does this work? The empirically determined σ [for Eq. (62)] of Eq. (65) is shown for n = 3 in Fig. 13a and n = 4 in Fig. 13b. The figures were constructed by plotting the ratios of d_{n+1} and d_n scaled respective to T = 16 in Fig. 13a and T = 32 in Fig. 13b. Evidently the scaling law Eq. (65) is being obeyed. Moreover, on the same graph Fig. 14 shows the empirical σ for n = 4 and the recursion theoretical σ of Fig. 10. The reader should observe the detail-by-detail agreement of the two. In fact, if we use Eq. (65) and the theoretical σ with n = 2 as empirical input, the n = 5 frequency spectrum agrees with the empirical n = 5 spectrum to
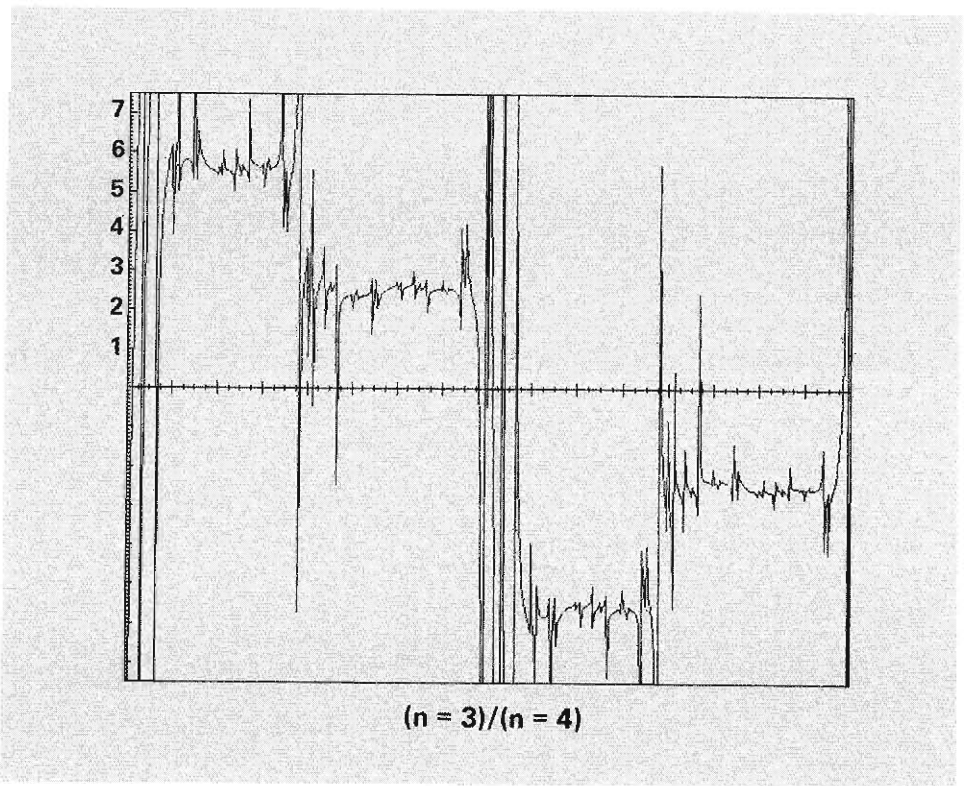

(n = 3)/(n = 4)

Fig. 13a. The ratio of nearest copy separations in the 8-cycle and 16-cycle for Duffing's equation.
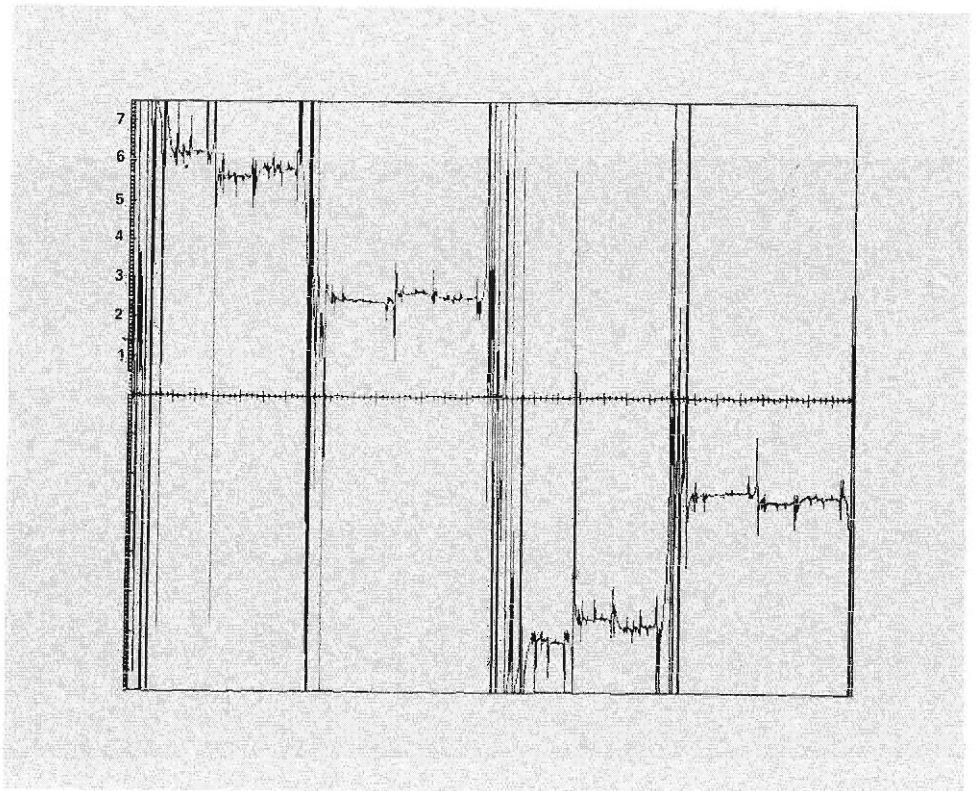

Fig. 13b. The same quantity as in Fig. 13a, but for the 16-cycle and 32-cycle. Here, the time axis is twice as compressed.
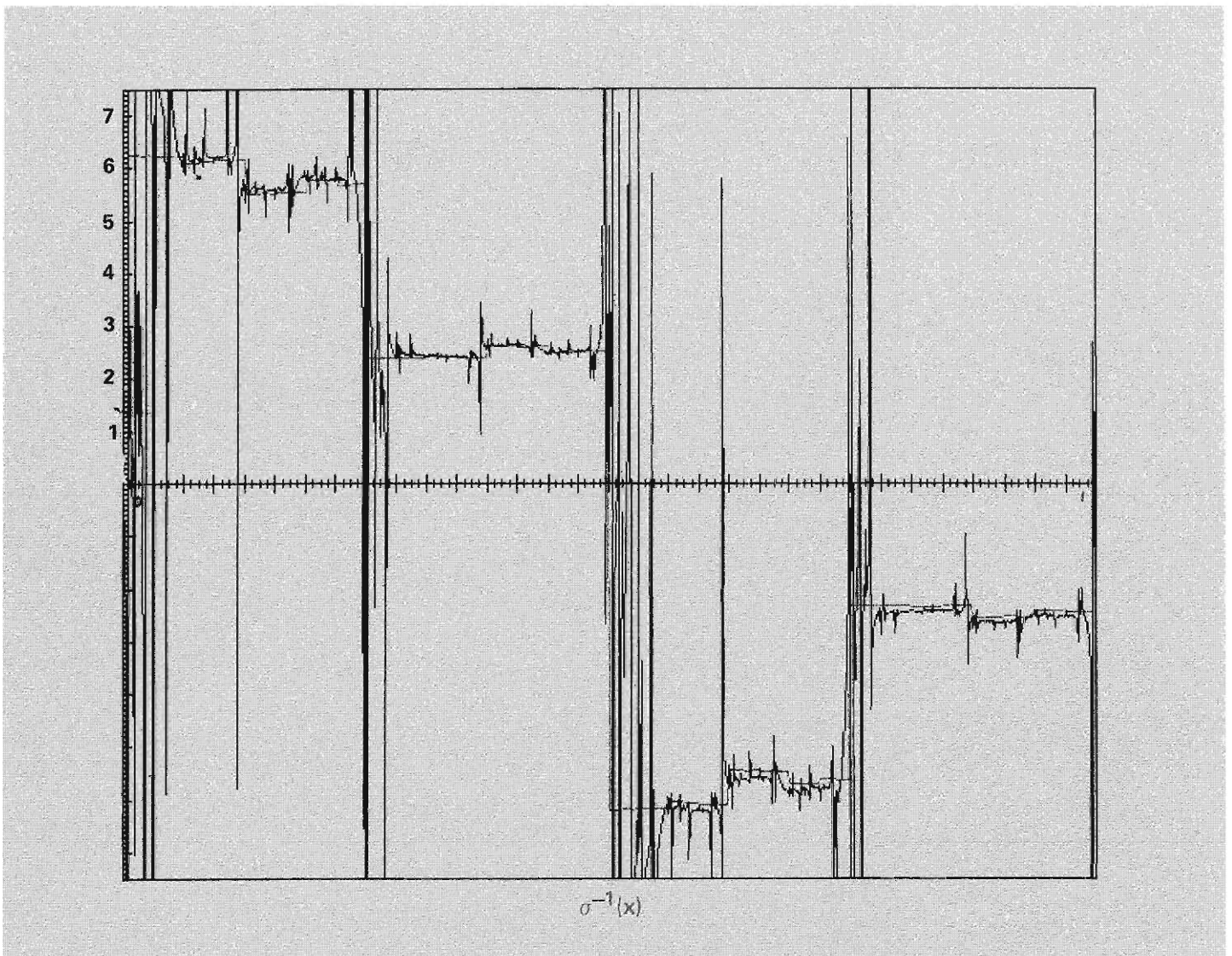
Fig. 14. Figure 13b overlayed with Fig. 10 compares the universal scaling function σ with the empirically determined scaling of nearest copy separations from the 16-cycle to the 32-cycle for Duffing's equation.

within 10%. (The n = 4 determines n = 5 to within 1%.) Thus the asymptotic universality theory is correct *and* is already well obeyed, even by n = 2!

Equations (64) and (65) are solved, as mentioned above, through Fourier transforming. The result is a recursive scheme that determines the Fourier coefficients of $x_{n+1}(t)$ in terms of those of $x_n(t)$ and the Fourier transform of the (known) function σ(t). To employ the formula accurately requires knowledge of the entire spectrum of x_n (amplitude *and* phase) to determine each coefficient of x_{n+1}. However, the formula enjoys an

approximate local prediction, which roughly determines the amplitude of a coefficient of x_{n+1} in terms of the amplitudes (alone) of x_n near the desired frequency of x_{n+1}.

What does the spectrum of a period-doubling system look like? Each time the period doubles, the fundamental frequency halves; period doubling in the continuum version is termed half-subharmonic bifurcation, a typical behavior of coupled nonlinear differential equations. Since the motion *almost* reproduces itself every period of the drive, the amplitude at this original fre-

quency is high. At the first subharmonic halving, spectral components of the odd halves of the drive frequency come in. On the route to aperiodicity they saturate at a certain amplitude. Since the motion more nearly reproduces itself every two periods of drive, the next saturated subharmonics, at the odd fourths of the original frequency, are smaller still than the first ones, and so on, as each set of odd 2^nths comes into being. A crude approximate prediction of the theory is that whatever the system, the saturated amplitudes of each set of successively lower half-frequencies

define a smooth interpolation located 8.2 dB *below* the smooth interpolation of the previous half-frequencies. [This is shown in Fig. 15 for Eq. (62).] After subharmonic bifurcations *ad infinitum*, the system is now no longer periodic; it has developed a continuous broad spectrum down to zero frequency with a definite internal distribution of the energy. That is, the system emerges from this process having developed the beginnings of broad-band noise of a determined nature. This process also occurs in the onset of turbulence in a fluid.

The Onset of Turbulence

The existing idea of the route to turbulence is Landau's 1941 theory. The idea is that a system becomes turbulent through a succession of instabilities, where each instability creates a new degree of freedom (through an indeterminate phase) of a time-periodic nature with the frequencies successively higher and incommensurate (*not* harmonics); because the resulting motion is the superposition of these modes, it is quasi-periodic.

In fact, it is experimentally clear that quasi-periodicity is incorrect. Rather, to produce the observed noise of rapidly decaying correlation the spectrum must become *continuous* (broad-band noise) down to zero frequency. The defect can be eliminated through the production of successive half-subharmonics, which then emerge as an allowable route to turbulence. If the general idea of a succession of instabilities is maintained, the new modes do *not* have indeterminate phases. However, only a small number of modes need be excited to produce the required spectrum. (The number of modes participating in the transition is, as of now, an open experimental question.) Indeed, knowledge of the phases of a small number of amplitudes at an early stage of period doubling suffices to determine the phases of the transition
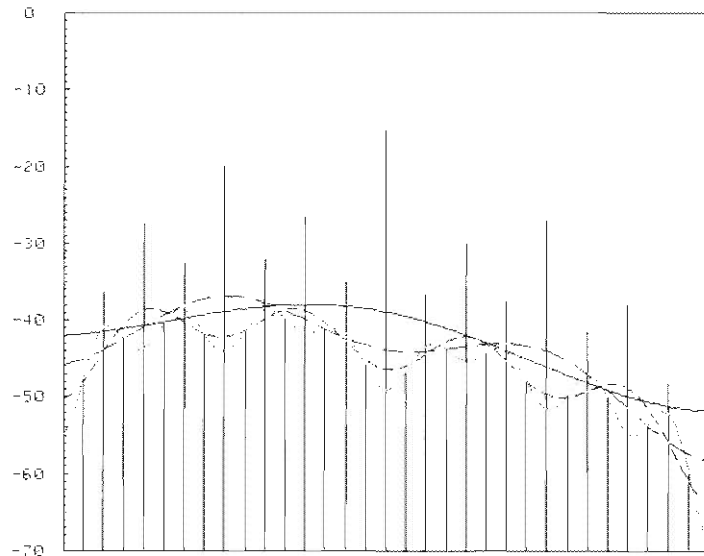

Fig. 15. The subharmonic spectrum of Duffing's equation in the 32-cycle. The dotted curve is an interpolation of the odd 32nd subharmonics. The shorter dashed curve is constructed similarly for the odd 16th subharmonics, but lowered by 8.2 dB. The longer dashed curve of the 8th subharmonics has been dropped by 16.4 dB, and the solid curve of the 4th subharmonics by 24.6 dB.
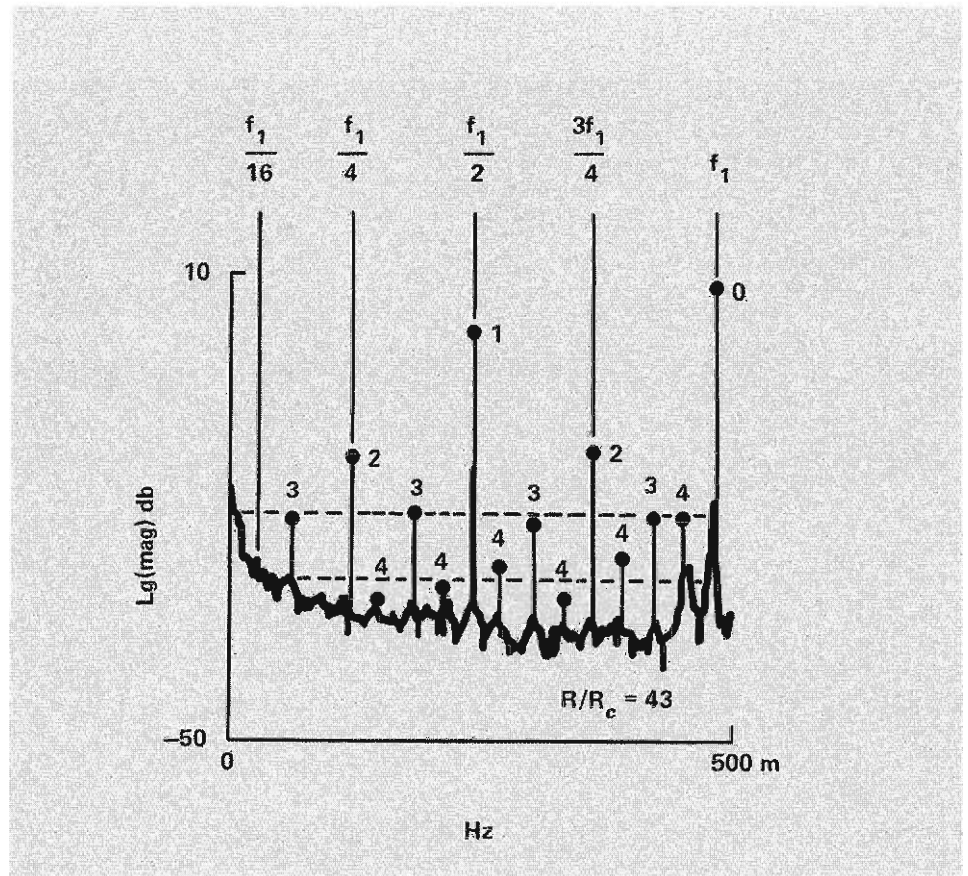

Fig. 16. The experimental spectrum (redrawn from Libchaber and Maurer) of a convecting fluid at its transition to turbulence. The dashed lines result from dropping a horizontal line down through the odd 4th subharmonics (labelled 2) by 8.2 and 16.4 dB.

Mitchell J. Feigenbaum is a mathematical physicist and inventor of the universality theory. He earned his bachelor of electrical engineering degree from the City College of New York in 1964 and his Ph.D. in elementary particle physics from Massachusetts Institute of Technology in 1970. He was a Research Associate at Cornell University and Virginia Polytechnic Institute before becoming a Staff Member in the Theoretical Division at LASL in 1974. In May, 1980, he received a LASL Distinguished Performance Award for his universality theory and development of the first quantitative understanding of the onset of turbulence. He is presently writing a monograph on this work for the series *Progress in Physics* published by Birkhauser Boston, Inc. His other technical interests include field theories and functional integrals, the renormalization group and phase transitions, and nonlinear dynamics generally. His peers at LASL often turn to him as the resident consultant on related technical matters.

spectrum. What is important is that a purely causal system can and does possess essentially statistical properties. Invoking *ad hoc* statistics is unnecessary and generally incompatible with the true dynamics.

A full theoretical computation of the onset demands the calculation of successive instabilities. The method used traditionally is perturbative. We start at the static solution and add a small time-dependent piece. The fluid equations are linearized about the static solution, and the stability of the perturbation is studied. To date, only the first instability has been computed analytically. Once we know the parameter value (for example, the Rayleigh number) for the onset of this first time-varying instability, we must determine the correct form of the solution after the perturbation has grown large *beyond* the linear regime. To this solution we add a new time-dependent perturbative mode, again linearized (now about a time-varying, nonanalytically available solution) to discover the new instability. To date, the second step of the analysis has been performed only numerically. This process, in principle, can be repeated again and again until a suitably turbulent flow has been obtained. At each successive stage, the computation grows successively more intractable.

However, it is just at this point that the universality theory solves the problem; it works only after enough in-stabilities have entered to reach the asymptotic regime. Since just two such instabilities already serve as a good approximate starting point, we need only a few parameters for each flow to empower the theory to complete the hard part of the infinite cascade of more complex instabilities.

Why should the theory apply? The fluid equations make up a set of coupled field equations. They can be spatially Fourier-decomposed to an infinite set of coupled ordinary differential equations. Since a flow is viscous, there is some smallest spatial scale below which no significant excitation exists. Thus, the equations are effectively a finite coupled set of nonlinear differential equations. The number of equations in the set is completely irrelevant. The universality theory is generic for such a dissipative system of equations. Thus it is possible that the flow exhibits period doubling. If it does, then our theory applies. However, to prove that a given flow (or any flow) actually should exhibit doubling is well beyond present understanding. All we can do is experiment.

Figure 16 depicts the experimentally measured spectrum of a convecting liquid helium cell at the onset of turbulence. The system displays measurable period doubling through four or five levels; the spectral components at each set of odd half-subharmonics are labelled with the level. With $n = 2$ taken as asymptotic, the dotted lines show the crudest interpolations implied for the $n = 3$, $n = 4$ component. Given the small amount of *amplitude* data, the interpolations are perforce poor, while ignorance of higher odd multiples prevents construction of any significant interpolation at the right-hand side. Accordingly, to do the crudest test, the farthest right-hand amplitude was dropped, and the oscillations were smoothed away by averaging. The experimental results, -8.3 dB and -8.4 dB, are in surprisingly good agreement with the theoretical 8.2!

From this good experimental agreement and the many period doublings as the clincher, we can be confident that the measured flow has made its transition according to our theory. A measurement of δ from its fundamental definition would, of course, be altogether convincing. (Experimental resolution is insufficient at present.) However, if we work backwards, we find that the several percent agreement in 8.2 dB is an *experimental observation* of α in the system to the same accuracy. Thus, the present method has provided a theoretical calculation of the actual dynamics in a field where such a feat has been impossible since the construction of the Navier-Stokes equations. In fact, the scaling law Eq. (65) transcends these equations, and applies to the *true* equations, whatever they may be.