AN INTRODUCTION TO DYNAMICAL SYSTEMS AND MATHEMATICAL MODELLING

Donal O'Shea
Department of Mathematics
Mount Holyoke College

MONOGRAPH SERIES OF THE NEW LIBERAL ARTS PROGRAM

The New Liberal Arts (NLA) Program of the Alfred P. Sloan Foundation has the goal of assisting in the introduction of quantitative reasoning and concepts of modern technology within liberal education. The Program is based on the conviction that college graduates should have been introduced to both areas if they are to live in the social mainstream and participate in the resolution of policy issues.

The New Liberal Arts Program has led to significant, new courses and course changes in many colleges and universities. MIT Press and McGraw-Hill are jointly publishing an NLA series of books. These monographs are planned to provide teaching/learning materials for other educational developments. A complete list of the monographs can be found on the last page of this book.

NLA Monograph Series

Co-Directors: John G. Truxal and Marian Visich, Jr.
Production Editor: Carol Galdi

Preface and Acknowledgements

The primary aim of this monograph is to explore some models which have been central in the development of mathematical modelling and which exemplify the process of using mathematics to understand the world. I wanted to show how mathematics can be, and is, used as a speculative tool. Computer exploration makes this exploratory side of mathematics accessible to students with a solid high school background. I hope, of course, that there is enough material so that the better student will be able to create his or her own models, but this is not my main purpose.

There are a number of reasons I feel that this apprach is of value. First, there are a number of good books devoted to constructing simple models and I think I have relatively little to add. Secondly, I think it is important to see truly significant examples of models – such models have really shaped our thinking. If one thinks in terms of the distinction between a course devoted to reading great literature and a "how to write" course, then the approach taken here is analogous to that in the former. (Both courses are necessary.) Third, I think it is important to stress the importance of evaluating assumptions – this is the sort of thing an informed citizen needs to be able to do. The fitting of coefficients to get the best agreement between model and reality is not something our average student is, or ever will be, called upon to do.

We focus primarily on dynamical systems, which three centuries of experience have shown to be the most useful class of mathematical models. Historically, the discovery and use of dynamical systems have been intimately bound up with the calculus and considerable knowledge of calculus has been required to use them effectively. The current monograph supposes *no* knowledge of calculus. Instead, we systematically use the computer to explore the behavior of dynamical systems. We supply sample programs together with a detailed explication of what the program is doing, in the hope that an interested reader will be able to modify these programs and write his or her own. Classroom experience has shown that this goal is not too ambitious. I also wanted to to convey the current activity in the area of dynamical systems and give the reader a sense of what it is that mathematicians do and the interplay between mathematical research and the ways in which we perceive our world.

The material in this monograph and its treatment owe much to my colleagues. I would particularly like to single out those in the Five College Calculus in Context project: Harriet Pollatsek and Lester Senechal of Mount Holyoke College, James Callahan of Smith College, David Cox of Amherst College, Ken Hoffman of Hampshire College and Frank Wattenberg of the University of Massachusetts. Virtually everything that is novel in this monograph is due to them. I would also like to thank the Sloan New Liberal Arts program for creating a climate on campus which encouraged the curricular experimentation of which this monograph is a result. Support from the NLA program allowed me to take the time to successively rework the course so as to make it accessible to almost all first year students. Sadly, I cannot even take credit for the idea that an honest course was possible at this level: the idea is due to the directors of the NLA program, John Truxal and Mike Visich of the State University of New York at Stony Brook, who also provided support and encouragement, and to Ken Hoffman. To all these people, to Sam Goldberg of Oberlin who provided much thoughtful criticism, and to five groups of students at Mount Holyoke who put up with successive iterations of this material, thanks.

Lastly, especial thanks to Mary, for being her inimitable self, and to my children, Seamus, Brendan, Sarah and Kathleen, for being themselves. I dedicate this monograph to them.

Contents

Chapter 1 — Of Snakes and Birds

What is Mathematical Modelling?

Mathematical modelling is the process of trying to model some "chunk of reality", perhaps some situation or some constellation of phenomena, in mathematics with a view to using mathematical reasoning to explain, explore, or predict aspects of that reality.

By a *mathematical model* we mean a "dictionary" which sets up a correspondence between the chunk of reality and a set of mathematical objects, together with a set of "operations or rules" which allow one to pass between mathematical objects in the set. This definition is necessarily fuzzy: we will elaborate on it as we go along.

We intend to let the main ideas emerge from an examination of successful models. The last decade has seen a growing awareness of the environmental and ecological changes brought about by humans. Accordingly, we begin with an examination of some simple ecological models which have shed light on the ways in which different species can interact.

Snakes on Guam

Consider first the island of Guam, a small U .S . protectorate half way been New Guinea and Japan. Guam is 30 miles long by 6-10 miles wide, home to 110,000 people. The main industries are the military, tourism, textiles, and petrochemical refining. It is an idyllic place, with temperatures between 72°F and 88°F year round, gentle green hills and sandy beaches.

In 1950, Guam was home to 18 species of birds. The forests were alive with their songs. In the early 1970's, biologists noted a sharp decline in the numbers of some species. By the late 1970's it was obvious to all that something serious was happening. Fewer and fewer birds of any type were observed. By the early 1980's, the Fish and Wildlife Service of the U .S . Department of the Interior started to investigate in earnest and several species were put on the endangered list – sadly, however, most of the birds on the list were already extinct. Today, there are almost no birds at all on Guam.

What was the cause? In Hawaii many birds had died has a result of an avian virus, and it was widely felt that some such virus was causing the destruction. In 1984, a young Fisheries biologist, Julie Savidge, discovered that the cause was something rather different: the brown tree snake, a snake introduced to Guam, apparently by accident, in the late 1940's or early 1950's (the first snakes were sighted in 1952).

The brown tree snake is native to New Guinea and the northeastern coast of Australia. Growing to lengths of over 8 feet, it is nocturnal, very aggressive, and a superb climber. Since the snake is nocturnal and sleeps during the day, it is difficult to spot.

Nevertheless, it was apparent by the end of the 1970's and the early 1980's that the number of snakes had vastly increased. The snake would climb the guy wires to power poles, seeking birds nesting on top of the poles, and often short out transformers. From 1978 to 1982, the number of power outages caused by the snake increased fivefold from less than 20 to nearly 100, causing millions of dollars in damages. In 1985, in what can only be considered as a gesture of defeat, one of the main power lines was shut off at night, effectively abandoning it to the snakes.

The snake is mildly venomous – having fangs, but far back in its mouth, so that it has to really chew to inject its venom. At this writing, no deaths from the snake have been reported. Nevertheless, about one in a thousand visits to hospital emergency rooms on Guam are for brown tree snake bites. One hospital treated about 50 victims bitten by the snake last year, many of them children. The snake has been found everywhere – it can crawl into houses, entering via vents in roofs, open windows, and sewer lines. In one house, it bit a six-week old infant that was sleeping between its parents. In another, parents found their two month old son in his crib with a five foot snake tightly coiled around his neck and repeated bites on his arms and legs.

Modelling the Snake Population

It is clear that the snake population underwent a period of explosive growth in the late 70's and early 80's. In order to understand how the snake population might grow, we let x be the number of snakes on the island. We are interested in how x varies from year to year. Let x' denote the rate of change of x per year. Thus, $x = 10000, x' = 3$ would mean that there are 10000 snakes and the amount is increasing by 3 snakes/year. If the number of snakes were decreasing by 3 snakes/year, we would write $x' = -3$.

One could be a little more subtle, and let x be the total weight of the snakes, measured in units of average snake mass. Since the average brown snake weighs about one kilogram, x will be the number of kilograms of snakes, which is approximately the number of snakes. The advantage of this interpretation is that there is no trouble saying what it means for x or x' to be 3.5, for example. We will continue to think of x as the number of snakes, but if fractional values of snakes bother you, you might consider that we have adopted this convention.

Since Guam is isolated, we don't have to worry about snakes immigrating or emigrating. The rate of change x' of the snakes will be equal to the average number of births per year minus the average number of deaths of per year. Not a great deal is known about the reproductive patterns of the brown snake. As a first approximation we might imagine that both the number of deaths and the number of births is roughly proportional to the number of snakes (that is, twice as many snakes would result in twice as many deaths and twice as many births). Thus, the rate of change of the number of snakes is proportional to the number present. (This means that if, for example, the net increase in a population of 100 snakes is ten per year, then a population of 300 would increase by thirty in a year.) In terms of formulas, our first approximation means that $x' = ax$ where a is some positive constant which would need to be measured.

We can try to estimate the number a. A female snake held in captivity was observed to lay 4 eggs (none of which hatched). Let's be conservative and say that a female snake in the wild on Guam will give rise to one snake that makes it to adulthood a year (the number may well be more: on Guam there are no natural predators which feed on the snakes – there are also indications that a female can lay clutches of up to 12 eggs twice a year). The brown snake lives about ten years, so we may as well assume that one tenth of the snakes die per year. The number will be slightly higher: some snakes are electrocuted, run over by cars, or killed by humans. Let's ignore this, since this number is small and probably balanced by the fact that the average snake actually lives longer than ten years

Let's suppose that half the snakes are female. Then, in a population of x snakes, the number of females will be $\frac{1}{2}x$. On average, each will give rise to one snake per year, so we will have $\frac{1}{2}x$ new snakes per year. The number of deaths will be $\frac{1}{10}x$ snakes per year. Thus the rate of change of the snakes will be

$$x' = \left(\frac{1}{2} - \frac{1}{10}\right)x$$

and $a = \frac{1}{2} - \frac{1}{10} = .5 - .1 = .4$. This approximation is clearly unrealistic if there are a lot of snakes. (In fact, the positive growth rate means that the number of snakes will increase without bound.) They'll eat all the food and start to starve. Moreover, if they are really crowded, disease will spread more easily, and it's reasonable to assume that the death rate will go up. After some point, the number of brown snakes will start to decrease, so rate of change will be negative. (this number might be very high – there are lots of brown snakes on Guam and their numbers don't seem to be declining). How can we take this into account?

There are a number of different plausible ways and one cannot say *a priori* what's best – one would have to consult an ecologist and even then the answer might very well be equivocal. In the case of the brown tree snake, not enough is known about it. With this cautionary note in mind, let's see what sort of equation has the property that the rate of change is positive when x is small, and negative when x is large. Again, for definiteness, let's say that the population will start to decrease when x is bigger than 1,000,000 say. (In some areas there are as many as 30,000 per square mile; however, the average density seems to be around 5000 snakes per square mile. As Guam has an area of about 200 square miles, this gives a total population of 1,000,000, which is a lot of snakes!!) Probably the simplest quantity that is positive when x is smaller than 1,000,000 and negative when x is larger is $1000000 - x$. Since we want the rate of change to be proportional to x when x is small, we can't write $x' = 1000000 - x$ (there are other reasons that this is nonsense: among other things, it says that the closer x is to 0, the more the population approaches a growth rate of 1,000,000 snakes/year, which is absurd.) A better idea is to consider the quantity $1 - (x/1000000)$. This is positive for $x < 1,000,000$ and negative for $x > 1,000,000$. We still can't write $x' = 1 - (x/1000000)$ because x' is not proportional to x for small x (in fact, this equation would imply that a population of zero snakes has a growth rate of one snake/year, a clear absurdity – the snakes are bad, but not that bad!). A good alternative is to write

$$x' = .4x(1 - .000001x). \tag{1}$$

Upon multiplying through, we get

$$x' = .4x - .0000004x^2. \tag{1'}$$

For x small the first term will be much bigger than the second and the rate of change of the population will be roughly proportional to x with a rate of increase close to $.4x$ snakes/year. (If for example, $x = 1000$, the first term would be 400 and the second term .4, so that $x' = 399.6$ is very nearly $.4x = 400$; the approximation would be even better if x were less than 1000.) From the first form of the equation, we see that the population will grow (that is, $x' > 0$) as long as x is less than 1,000,000. When x is equal to 1,000,000, we will have $x' = 0$, so the population will not change. When $x > 1,000,000$, we will have $x' < 0$, so that the population will decrease. This suggests that the population will level off at 1,000,000 individuals.

Here is another interpretation of the above equation. We can think of the term proportional to x^2 as a 'crowding' factor. This comes from the fact that, to a first approximation, we can think of the number of encounters between two individuals in a population of x individuals as proportional to x^2. (The reason is that the number of ways you can choose 2 individuals out of a population of x individuals is $x(x-1)/2$. The first individual can be chosen in x different ways and the second in $x-1$ different ways, so there are $x(x-1)$ different ways to choose two individuals if one distinguishes between which is chosen first and which second. If one does not distinguish between the first and second choices, then the number of ways to choose is $x(x-1)/2$ and, for x large, this is pretty close to $x^2/2$). This approximation will be valid even if we think of x as the total mass of individuals in units of average individual size, because x is close to the actual number of individuals). Now, the rate that disease is spread is roughly proportional to the number of contacts, as is the number of times two snakes will get to the same bird. So, we can think of equation $(1')$ as saying that the rate of change of the population is proportional to the number of individuals decreased by a crowding factor proportional to the square of the population.

Other Models?

Before proceeding, we want to emphasize that we arrived at equation (1) (or, equivalently, equation $(1')$) by taking the simplest mathematical expressions consistent with the suppositions that the population increases at the rate of .4 snakes/year times the population when the population is small and decreases once the population exceeds 1,000,000. There are lots of other equations which are consistent with these suppositions. For example, we might have

$$x' = .4x(1-.000001x)^3$$

or

$$x' = \frac{.4x}{1-.000001x}$$

or something even more complicated. It is reasonable to choose the simplest possible mathematical expressions when we lack other data, but it is irresponsible to assume that reality must be that way.

Logistic Growth

Equation $(1')$ has the general form

$$x' = ax - bx^2 \tag{2}$$

where a and b are positive constants. This equation turns out to capture certain features common to many populations. Because of this, and because it is simple, it plays an important role in mathematical ecology. We can rewrite equation 2 as

$$x' = ax(1 - \frac{b}{a}x)$$

or, even more suggestively, as

$$x' = ax\left(1 - \frac{x}{a/b}\right.$$

We see that the population will increase (that is, $x' > 0$) as long as the population is less than $\frac{a}{b}$. It will decrease ($x' < 0$) if the population is greater than $\frac{a}{b}$ and it will remain the same if $x = \frac{a}{b}$. (It is clear that to make biological sense, we must have b *much* smaller that a.) Another feature is that when x is small, the population increase will be roughly proportional to the population, each individual contributing on average a individuals per year or other time unit (from which it is clear that a should be small). Equation (2) is called the *logistic model* of population growth. When a population grows in accordance with equation (2), we say that it *grows logistically* with growth rate a and *carrying capacity* $\frac{a}{b}$. The carrying capacity represents the maximum number of individuals the environment can support.

The Number of Snakes over Time

Equation ($1'$) allows us to say, in principle at least, how many snakes there are at any time, once we know the number of snakes at a given time. To see this, imagine that $x = 100$ at some time. Then according to equation $1'$, the rate of change x' of the number of snakes is

$$.4 \times 100 - .0000004 \times (100)^2 = 40 - .0004 = 39.9996$$

snakes per year. Thus, after one year, we'd expect to have 139.9996 snakes (our original 100, plus the change of 39.9996). If you like, you interpret the fractional numbers as referring to weight. We hasten to point out that here, and in what follows, we are only carrying the decimal places so that you can check the computations – in actual fact, we have made so many assumptions that it is ridiculous to pretend that we're getting anything close to three or four place decimal point accuracy. At this time, the rate of change of the number of snakes (using equation $1'$ again) would be

$$x' = .4 \times 139.9996 - .0000004 \times (139.9996)^2 = 55.992$$

snakes per year.

So, at the end of the second year, we'd have $x = 140.000 + 55.992 = 195.992$ snakes and the rate of change of the number of snakes would be

$$x' = .4 \times 195.992 - .0000004 \times (195.992)^2 = 78.381$$

snakes per year.

At the end of the third year the number of snakes would be 195.992 + 78.381 = 274.373. This computation could easily be continued to give us a reading for the number of snakes as many years ahead into the future as we desire.

Before going wild, and figuring out how many snakes there will be 10, 20, 30, ... years later, let us note that the computation above isn't quite right. Do you see why?

The reason is that the rate of change of the snakes changes whenever the number of snakes changes. Our computation hasn't taken this into account – in fact, we've tacitly assumed that the rate of growth of the number of snakes remains constant throughout the year. Consider, for example, what happens to the number of snakes between the end of the second year and the end of the third year. Assuming that there were 195.992 snakes at the end of the second year, we computed that $x' = 78.381$ snakes per year at the end of the second year (the end of the second year is, of course, the same as the beginning of the third year). We then used these numbers to get the number of snakes at the end of the third year by assuming that the number of snakes grew at the constant rate of 78.381 snakes per year throughout the third year. However, halfway through the third year, the number of snakes would have been at least

$$195.992 + \frac{1}{2}\text{ years} \times 78.381\text{ snakes per year} = 235.183\text{ snakes}.$$

Hence, the rate of change would have been

$$.4 \times 235.183 + .0000004 \times (235.183)^2 = 94.095\text{ snakes per year},$$

which is quite a bit higher than it was at the beginning of the year. If we'd used this number to estimate the number of snakes at the end of the third year, we'd get at least

$$235.183 + \frac{1}{2}\text{ years} \times 94.095\text{ snakes per year} = 282.230\text{ snakes},$$

which is different than the number 274.373 computed above!

In fact, the computation we have just done suffers from exactly the same defect we pointed out earlier. Namely, it assumes that the rate of change of the number of snakes is constant for half a year. In actual fact, after even only a month, the number of snakes will have increased and, hence, the rate of change will be different our estimate does not take this into account. We could compute the number of snakes after each month.

If x_0 is the number of snakes at some time, then the rate of change of the snakes

$$x_0' = .4x_0 - .0000004x_0^2 \quad \text{snakes/year.}$$

Thus, after one month there would be

$$x_0 + \frac{1}{12}x_0' = x_0 + \frac{1}{12}(.4x_0 - .0000004x_0^2) \tag{3}$$

snakes. We can then take this as our new value of x_0 and compute the number of snakes a month later still and so on.

Of course, we are assuming in equation (3) that the rate of change is constant over a month, which isn't the case. It's easy to modify equation (3) to assume that the rate of change of the snakes is constant over the period of a week. Doing this, after one week we would have

$$x_0 + \frac{1}{52}x_0' = x_0 + \frac{1}{52}(.4x_0 - .0000004x_0^2) \tag{4}$$

snakes. Continuing with the assumption that the rate of change does not vary during the course of a week, we could figure out the number of snakes after two weeks by applying the formula above twice (the second time replacing x_0 by the number computed using formula (4)), and after three weeks, by applying it three times, and so on.

Assuming that the rate of change does not change over the period of a week is certainly better than assuming that the rate of change is constant over the period of a month (which in turn is better than assuming that the rate of change is constant over the period of a year), but is still an approximation in the sense that the rate of change is changing constantly. We could if we wished compute over an even shorter time period to get a better approximation. In general, if we assume the rate of change is constant over a time interval of

Δt years, and we start with x_0 snakes, the number of snakes x_1 after Δt years is

$$x_1 = x_0 + \Delta t\, x_0' = x_0 + \Delta t\,(.4x_0 - .0000004x_0^2) \qquad (5).$$

After $2\Delta t$ years, then number x_2 of snakes is

$$x_2 = x_1 + \Delta t\, x_1' = x_1 + \Delta t\,(.4x_1 - .0000004x_1^2).$$

After $3\Delta t$ years, then number x_3 of snakes is

$$x_3 = x_2 + \Delta t\, x_2' = x_2 + \Delta t\,(.4x_2 - .0000004x_2^2),$$

and so on. Note that we are just repeatedly applying formula (5), which followed immediately from formula (1'), where we take the result obtained and feed it back in. This process of repeatedly applying an equation is called **iteration** of the equation.

Incidentally, the notation Δt is traditional – one typically uses the upper case Greek letter Δ (read "delta") to denote a small amount of some quantity: Δt refers to a small interval of time; Δx would refer to a small number of snakes.

We get better and better approximations by choosing Δt smaller and smaller. The price we pay is that if we want to know the number of snakes after, say, 10 years, we have to apply equation (5) more and more times as Δt gets smaller. If $\Delta t = 1$, then we have have to iterate equation (5) 10 times; if instead of assuming that the rate of change stayed constant over a period of a year, we assumed only that the the rate of change was constant over a day (in which case $\Delta t = \frac{1}{365}$, we would have to iterate equation (5) 3650 times. This is not difficult in principle, but it is, in practice, a lot of arithmetic!

Using a Computer

Fortunately, it is easy to get a computer to do the computations for us. We just have to write out a clear set of instructions, called a *program*, which tells the computer what to do at each step. The repetitive sort of task involved in iterating a single equation is ideally suited for a computer because, while the number of computations is large, the instructions will be simple. We just want the computer to do the same thing over and over again.

Instructions must be given to a computer in a language that it understands. There are dozens of different computer languages, some of them better suited for different types of tasks than others. (Some of you may be familiar with the more common of these: Fortran, C, Pascal, Basic and variants.) Most of the languages are designed to be as much like English as possible in order to make them easy (for humans) to remember and understand. Here, for example, is a QuickBasic program to compute the number of snakes after ten years assuming that the rate of change is constant over a month (so $\Delta t = \frac{1}{12}$) and that we start with 10 snakes. (Incidentally, QuickBasic is a variant or "dialect"of Basic – other variants include TruBasic and Basica: the program below would be the same in any of these dialects. We have chosen to write all programs in QuickBasic, because it is easier to use than Basic, it's cheap, it is available on IBM-compatible PC's and Apple products, and it also looks very like Pascal, another program in widespread use.)

```
LET x=10
LET deltat=1/12
FOR N = 1 TO 120
xprime = .4*x-.0000004*(x*x)
x = x + deltat*xprime
NEXT N
PRINT x
```

You should look over the program to see what it does. The first two lines define the starting values of the variables. The variables are `x`, and `deltat`. Note that a variable need not be a single letter. We take advantage of this to name the variables so as to give a (human) reader a clue to what the program does — *the computer does not understand English words and doesn't care what we call the variables.* Even if it is perfectly clear to you, or to any other intelligent human being, what to do with the variable based on its name, it is not clear to the computer. The computer must be told explicitly what to do.

A computer reads a program line by line starting with the top line. In the above program the first line tells the machine that `x` is to be set equal to 10. Thereafter, unless it were told otherwise, everytime the computer encounters `x` it will substitute the number 10. If a later line gives a different value of `x`, say 3.14, the computer will substitute 3.14 for `x` from then on and forget that `x` was ever 10. The third line "`FOR N=1 TO 120`" is different from the preceding ones. It signals the beginning of what is called a *loop* which will be repeated 120 times. More precisely, it tells the computer to define a new variable `N`, which will only take integer values, and set it equal to 1. The computer then continues to read down the lines. It sets `xprime` equal to $.4 \times 10 - .0000004 \times 10^2$; that is, 3.9999996 (remember that it thinks `x` is 10). This is the rate of change of x, but the computer does not know this (nor would it make any difference if it did). Note that the computer interprets the asterisk * as a multiplication symbol. You must put this in if you want to multiply two numbers – it will not automatically assume that juxtaposition means multiplication. Reading the next line, the computer sets `x` equal to $10 + \frac{1}{12} \times 3.9999996 = 10.033$. (Remember, that it thinks `deltat` $= 1/12$. Also, for brevity, we are only going to carry three decimal places – the computer will, of course, carry more). Henceforth it thinks that `x` is 10.033 and has no memory of the fact that it once was 10. The computer encounters the line `NEXT N`, and something different happens. As you might expect, it sets `N` equal to the next integer (that is, 2) – but, then, before proceeding on to the next line, it goes back to the line where `N` was defined and checks whether the new value of `N` (that is, 2) is less than or equal to the second number (in our case 120) in the line defining `N`. If it is (as it is in our case: 2 is certainly less than or equal to 120), the machine starts reading at the line following the `FOR` statement (that is, the line following the statement "`FOR N = 1 TO 120`")). Thus it will read the next two statements again (but thinking that `x` = 10.033). Upon encountering the first line after the FOR statement , it sets `xprime` $= .4 \times 10.033 - .0000004 \times (10.033)^2 = 4.013$. When it encounters the second line it sets $\mathtt{x} = 10.033 + \frac{1}{12} \times 4.013 = 10.368$. Now, it encounters the line `NEXT N` the second time; it thinks that `N` is 2, so it sets `N` equal to 3 (the next number), goes back to the line "`FOR N = 1 TO 120`", checks whether 3 is less than 120, which it is, and consequently starts reading the line following the `FOR` statement. In this way, the next two lines are read 120 times. Finally, when `N` = 120 and the computer encounters the line "`NEXT N`", it sets `N` $= 121$ and, since 121 is not less or equal to 120, it does not read the line following the `FOR` statement, but goes to the line after "`NEXT N`". This line tells it to print the value which it is currently using for `x`, all previous values having been forgotten.

Let's run the program and see what happens. We get $x = 511.2602$. We mentioned that this will actually be a little off, because we are assuming that the rate of change of the snakes is constant over a period of a month, whereas the rate of change is actually increasing throughout the month. Let's see what we would get if we recomputed the rate of change every one thousandth of a year. Then we would take `deltat = 1/1000` and replace the `FOR` statement by `FOR N = 1 TO 10000` (because we would need 10,000 iterations of .001 years to get to ten years. This gives the program

```
LET x=10
LET deltat=1/1000
FOR N = 1 TO 10000
xprime= .4*x-.0000004*(x*x)
x = x + deltat*xprime
NEXT N
PRINT x
```

Note that it takes a few seconds longer (because there are so many more steps to compute). We get $x = 545.2544$. This is different enough so that we should probably switch to `deltat = 1/1000`. However, we stick to `deltat = 1/12` because it is so much faster. You should redo all the calculations below with `deltat = 1/1000` to check for yourself that what we are saying is not too far out of line, So let's return to taking `deltat = 1/12`. We can modify the program to see what happens after twenty years. Doing this, we get the following.

```
LET x=10
LET deltat=1/12
FOR N = 1 TO 240
xprime=-4*x-.0000004*(x*x)
x = x + deltat*xprime
NEXT N
PRINT x
```

Running the program, we find $x = 25517.95$. After thirty years $x = 579200.6$ and after forty years $x = 987498.9$. It is clear that the snake population is growing explosively twenty to forty years after they have been established. In fact, we can print a little table showing the growth of snakes against the number of years. We could do this by moving the `PRINT` statement inside the loop in the last program as follows.

```
LET x=10
LET deltat=1/12
FOR N = 1 TO 240
xprime=.4*x-.0000004*(x*x)
x = x + deltat*xprime
PRINT x
NEXT N
```

We would then get 240 values of x printed out (each time the machine went throught the loop it would print out a value of x). Try it! We could also keep track of the month by changing the `PRINT` statement to `PRINT N, x`. This is better, but the large number of lines makes it hard to see what is going on at a glance.

Instead, let us write out a program that will print out the number of snakes every 5 years over a 50 year period. There are many ways to do this – perhaps the simplest is to put in two loops, one which prints out the year and the number of snakes every 5 years (that is, every sixtieth time we compute a new x) and another that loops through every month. The following program does this.

```
LET x=10
LET deltat=1/12
PRINT "Year", "Snakes"
PRINT " ", " "
PRINT 0, 10
FOR J = 1 TO 10
    FOR N = 1 TO 60
        xprime= .4*x-.0000004*(x*x)
        x = x + deltat*xprime
    NEXT N
PRINT 5*J, x
NEXT J
```

Running it, we get Table 1, where the first column is the number of years after the time we start looking and the second column gives the number of snakes in that year. In order to keep things straight, we added the line "`PRINT "Year", "Snakes"`" which will have the effect of printing the word "Year" as the first entry in the first column and the word "Snakes" as the first entry in the second column. The next line ("`PRINT " ", " "`") will result in an empty line. These devices label the columns of our table.

Year	Snakes
0	10
5	71.51555
10	511.2602
15	3645.426
20	25517.95
25	158370.9
30	579200.6
35	911976.2
40	987498.9
45	998346.2
50	999783.4

Table 1. Computer output representing the number of years elapsed and the number of snakes on Guam

Notice that the bulk of the growth occurs between 25 years and 35 years after the original 10 snakes. In the case of Guam, if the snakes were introduced in the late 1940's or early 1950's (so say there were 10 in 1950), then this model indicates that the population explosion of snakes would have taken place between 1975 and 1985 (which accords well with observation) – the number of snakes would have almost quadrupled between 1975 and 1980, and almost doubled between 1980 and 1985.

The Logistic Graph

We can with minor modifications to the above programs, plot the number of snakes against time. The following program will plot the number of snakes over a sixty year time period.

```
SCREEN 12
WINDOW (0, 0)-(720, 1000000)
LET x=10
LET deltat=1/12
FOR N = 1 TO 720
xprime= .4*x-.0000004*(x*x)
```

```
x = x + deltat*xprime
PSET (N, x)
NEXT N
```

A number of lines in the program need explanation. We want graphical output, so we have to tell the computer what we are plotting on (in our case, a computer monitor). The first line tells the computer the resolution of the screen that we are plotting on: the program line `SCREEN 12` indicates a high resolution VGA color graphics screen. This command will vary from machine to machine and from language to language. If we think of the screen as a piece of graph paper, the second line tells the computer what the range of numbers to plot on the horizontal axis and what range to plot on the vertical axis. In other words, it specifies the scale. It does this by specifying the coordinates of the lower left hand point and the coordinates of the upper right hand point on the screen: the command `WINDOW (0, 0)-(720, 1000000)` says that the coordinates of the lower left hand point are (0,0) and those of the upper right hand point to are (720,1000000). (By convention, the first number a in the pair of coordinates (a, b) of a point refers to the position on the horizontal axis and the second number b to the position on the vertical axis.) This means the horizontal axis will correspond to the numbers from 0 to 720 and the vertical axis to the numbers from 0 to 1000000. (Remember that you have tell a computer *everything*: if you tell a person to graph something, they fiddle and choose the axes appropriately – in a computer program you have to explicitly say what you want.) If, instead, you had wanted the horizontal axis of the screen to correspond to the numbers between -25 and 362 and the vertical axis to the numbers between 1000 and 1200, the appropriate line would have been `WINDOW (-25, 1000) - (362, 1200)`.

The other line that needs comment is the eighth line `PSET (N, x)`. This tells the computer to plot the point `(N, x)` (that is, light up the pixel corresponding to the point N units along the horixontal axis and x units along the vertical axis). With the choice of coordinates made by `WINDOW` statement in the program displayed above, the command `PSET (0,0)` would cause the point in the lower left corner of the screen to light up and the command PSET (360, 500000) would cause the point in the center of the screen to light up. In the program above, every time the program goes through the loop (that is does an iteration) it lights up a pixel on the screen. The collection of these lit points is what we plot out below. Actually, 720 points does not produce a nice unbroken curve (there are too few dots) – so the curve below was produced using 7200 iterations (that is, we replaced the fourth line of the program above with `deltat = 1/120` and the fifth line with `FOR N = 1 TO 7200`).

Running the program gives the following picture.

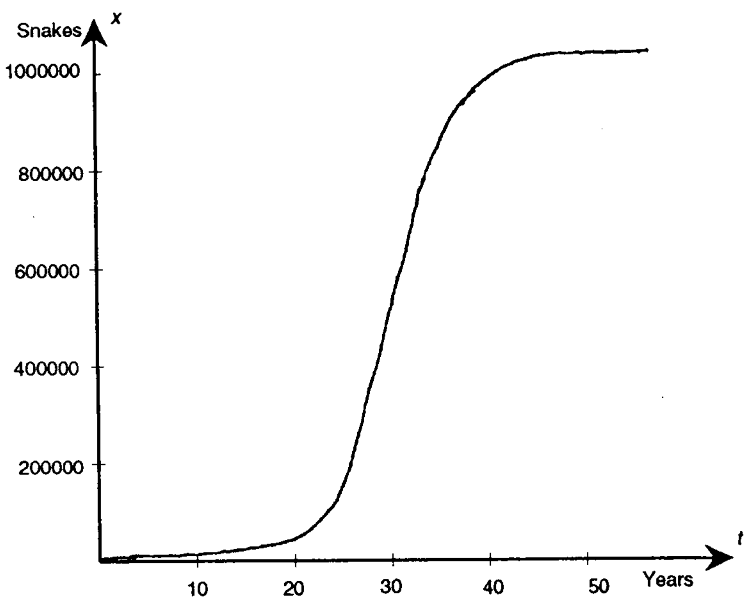

Figure 1. x as a function of time if $x(0) = 10$

The S-shaped curve is often called a **logistic curve** – in our case it represents the growth of the number of snakes over time. It is typical of the type of growth exhibited by a quantity that satisfies a logistic growth law. At first the quantity grows rather slowly, then there is a rapid increase to nearly the carrying capacity, after which the growth slows down. This is consistent with what seems to have happened on Guam. If we suppose that there were around ten snakes on Guam in 1950, then the model predicts a rapid (in fact, five- or sixfold) increase in the number of snakes between 1976 and 1983. If we imagine that the number of power outages caused by snakes is proportional to the number of snakes, then this prediction is in good agreement with the datum that the number of outages quintupled between 1978 and 1982. Of course, we should not take these numbers too seriously: good data on the number of snakes on Guam in the 60's and 70's is just not available. Moreover, the snakes required nearly thirty years to spread to all parts of the island – and our model takes no account of the distribution of the snakes on the island. In fact, our model is quite crude – it does not diferentiate between young and mature snakes (which would be important if we wanted to model the effect of snakes on large birds) and it incorporates assumptions about the birth and death rates of the snakes which are little better than guesswork. Nonetheless, it does explain how such a rapid increase in power outages and snake sightings could have occurred.

Summary

We did three things in this chapter.

- We identified a quantity in which we were interested.
- We wrote down an equation for the rate of change of this quantity.
- We used this equation for the rate of change to describe how the quantity behaves as time passes.

The quantity in which we were interested was the number of snakes x. The rate of change x' of the number of snakes (in units of snakes/year) was given by the equation $x' = .4x(1 - .0000001x)$. The growth in x as a function of time is described by the table in Table 1 and the graph in Figure 1.

These three steps are common to almost all modelling (involving dynamical systems). In fact, if there is any strategy common to all modelling efforts, it is the following: 1) identify the quantity (or quantities) of interest, 2) write down equations for the rates of change of these quantities, and 3) use these equations (and a computer) to get values of these quantities at any time.

These are the main principles. We also saw a number of more specific things.

- The rate equation we got belonged to a class of commonly used ecological models: namely, the logistic growth models $x' = ax - bx^2$
- The process by which we obtained the rate equation was filled with guesswork. At best we can say that the equation is plausible – the determination of a and b was an estimate subject to large amounts of error. Nonetheless, the process of working out a rate equation resulted in our identifying a number of interesting features of the situation: for example, the quantity a/b is called the carrying capacity and represents the number of individuals of a species that the environment can support. We will see many more examples of interesting quantities emerging from a modelling effort.
- There are a variety of computer languages – with the use if a manual it is easy to write a set of instructions or program to perform a given repetitive task.

There is one mathematical principle which is tacit in what we have done in this chapter and which will play a central role in all that follows. Namely:

- **If you know a quantity at some time and know the rate of change of that quantity for all values of the quantity (or at all times), then you can determine the value of that quantity at any time.**

We will elaborate on this principle later in this monograph. As we will see, experience shows that it is often easier to write equations for the rate of change of a quantity than for the behavior of the quantity itself over time.

Incidentally, the logistic growth model fits a good many populations encountered in practice and in the laboratory. For an account, see the first chapter of Waltman's book *Competition Models in Population Biology* (see bibliography). For more on the brown snake, see the papers of Fritts and Savidge cited in the bibliography.

Exercises on Logistic Growth

1. If we want to accurately get x from the equation for the rate of change x', then we have to compute x over shorter and shorter time intervals. In the text we saw that if we computed the number of snakes after 10 years, assuming that the rate of change is constant over a period of 1/12 of a year, then we obtained 511.2602 snakes. On the other hand, if we computed assuming the rates of change are constant over a period of a thousandth of a year, then we get 545.2544 snakes. Compute the number of snakes at the end of ten years using a time interval of one ten-thousandth of a year. Do the same thing for a time interval of one one-hundred-thousandth of a year. Based on these answers, what is the actual number of snakes, correct to one decimal place?

2. If $x' = 3x + 2$ and x is equal to 3 at time 0, what is x, correct to two decimal places, fifteen time units later? (You will need to write a computer program; make sure you make your choice of Δt explicit – you will need to experiment with several different choices until you feel confident of the accuracy.

3. What would the equation for the rate of change of the number of snakes be if we assume that each female snake produces on average three snakes each year that make it to adulthood? How long would it take a population of 10 snakes to grow to over 900,000 under this assumption.

4. What would the rate of change of the snakes be, if the average snake lived only 5 years (instead of 10). How long would it take a population of 10 such snakes to grow to over 900,000 snakes under this assumption?

Chapter 2– Ecological Models

Modelling Species Interactions

The dwindling number of birds on Guam was noticed before the burgeoning number of snakes. Generally, the growth of one species affects other species in different ways. Can we try to imagine what some of these are?

The interactions among species in a complex environment are intricate, so we confine our studies to two species. This will give us a sense of the issues involved in trying to model environments. The two species can interact in a number of different ways. They could compete in the sense that they eat the same food. Or they might be symbiotic, each species being able to survive without the other, but both doing better when the other is present. Or one species might feed on the other. We are interested in what will happen to the populations in each of these cases. Will one species kill off the other? Will the populations settle down to some steady state values in which both survive? Or will the populations fluctuate in some manner? To see what happens, we construct models for the different ways in which the species interact.

Competing Species

We know (or at least have a pretty good idea about) what happens when brown snakes and birds share the same environment, so let's instead consider two imaginary species of birds, the red-breasted berry guzzlers and the white-tufted berry chompers, on an isolated island. Let's suppose that both species of birds eat the same types of berries and that the berries grow in abundance on the island. Can we formulate some hypotheses regarding the rates of change of the populations?

Let x be the number of red breasted berry guzzlers (or, if you prefer, the biomass of berry guzzlers, measured in units of average guzzler weight). Let's measure time in months. So let x' denote the rate of change of the population per month. If there are no chompers around (and, of course, no other species of birds or animals, such as brown snakes, to interfere with the guzzlers), then we assume that the guzzlers grow logistically.

In particular, when the population is small, we assume that x' is proportional to x. For the sake of definiteness, let's suppose that the average lifespan of a guzzler is 3 years or, since we are measuring time in months, 36 months. Suppose further that each year every couple gives birth to (on average) 1 bird. Then, on the average, we would expect $\frac{1}{36}x$ birds to die and $\frac{1}{12} \cdot \frac{1}{2}x$ to be born each month. Thus, when x is small, we want

$$x' = (\frac{1}{24} - \frac{1}{36})x$$

Let's assume that the carrying capacity of the island is 10000 birds (so that the rate of change of x becomes negative if there are more than 10000 birds). Since we are assuming a logistic growth law we must have

$$x' = .014x(1 - \frac{x}{10000})$$

or, upon multiplying through,

$$x' = .014x - .0000014x^2. \tag{1}$$

As in the last chapter, we can, if we want, think of the term proportional to x^2 as a 'crowding' factor, and view equation (1) as saying that the rate of change of the population is proportional to the number of individuals decreased by a crowding factor proportional to the square of the population.

Before moving on, let us plot how x, viewed as a function of time, changes in accordance with equation (1). We st[illegible] [illegible]th $x = 100$ at time $t = 0$ and see what happens over the course of 100 years. The following program (in QuickBasic) will perform the computations. We do the computations at intervals of .1 months and plot the results for the 1200 months (that is, 100 years).

```
DEFDBL A-Z
SCREEN 12
WINDOW (0, 0)-(12000, 10000)
LET x = 100
LET deltat = .1
FOR N = 1 TO 12000
  xprime = .014 * x * (1 - .0001 * x)
  x = x + deltat * xprime
  PSET (N, x)
  NEXT N
```

The first line of this program needs explanation. Unless you tell it otherwise, QuickBasic (like BASIC and FORTRAN) will only store the first eight digits of a number – this is called single precision. We have numbers ranging from .0000014 (the coefficient of x_1^2) to 10000 (the maximum value of x_1) and we are asking the computer to do 12000 computations, so it behooves us to store more digits. (In the course of 12000 computations, it is easy to imagine that the small errors caused by rounding numbers to seven digits will compound and become significant). The next step up is called *double precision* and refers to having the computer store 16 digits. The first line of the program `DEFDBL A-Z` states that every variable beginning with the letters `a` to `z` should be treated as a double precision number. This means that 16 digits are stored. (This does not make much difference in this example, but it is good to get in the habit of using double precision when using a program which has a lot of iterations – this is absolutely crucial in some of the programs which follow). Running the program gives the graph sketched in Figure 1.

We drew the axes in Figure 1 by hand, although it would have been easy to add a few lines to the program to have the computer draw them.

Now let us consider the white-tufted berry chompers, the other species of birds on our island which also feeds on the abundant

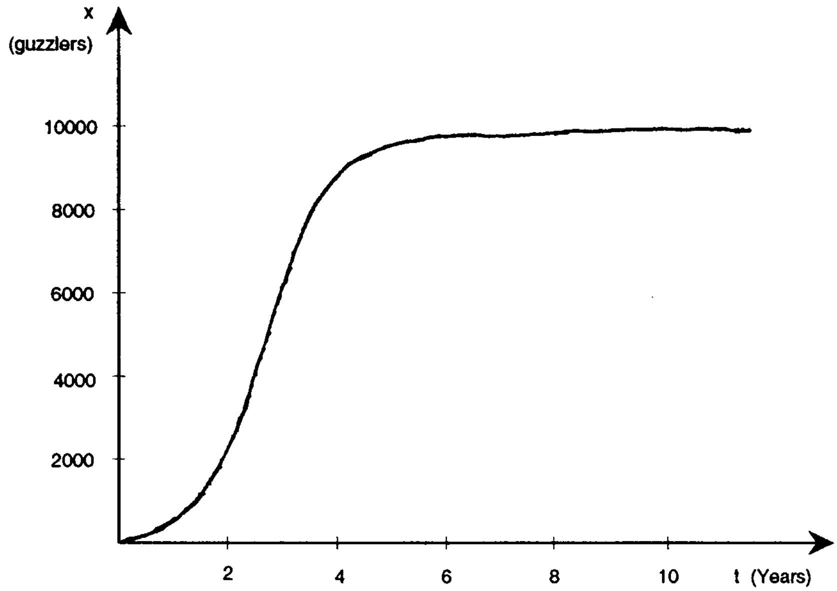

Figure 1. x as a function of time if
$x' = .014x(1 - .0001x)$ and $x(0) = 100$

berries. Let y be number of chompers (or, if you prefer, the biomass of the chompers in units of the average indivual berry chomper size). Again, for the sake of argument, let's suppose that without the guzzlers, the berry chompers would also grow logistically (so $y' = ay - by^2$ for some specific values of a and b).

Suppose also that when the population of chompers is relatively small and there are no guzzlers around, the average chomper lives for two years and that each couple gives birth, on average, to two berry chompers a year. Thus, if the chomper population is not too large, the number of births per month is $\frac{2}{12}\frac{1}{2}y$ (where we are assuming that $\frac{1}{2}y$ is the number of couples). Since each bird lives an average of 24 months, in a population of y birds, we can expect an average of $\frac{1}{24}y$ deaths per month. Thus if y is not too large, the rate of change y'

of y is equal to $(\frac{2}{24} - \frac{1}{24})y = .042y$. Let's suppose that the average chomper is smaller than a guzzler, so that the island could support 12500 chompers. Under these assumptions we have

$$y' = .042y(1 - \frac{1}{12500}y)$$

or, doing the arithmetic,

$$y' = .042y - .0000034y^2. \qquad (2)$$

We remark that in rounding off to get the coefficients of y and y^2 to two digits, we have actually changed the carrying capacity slightly from 12500 to $\frac{.042}{.0000034} \approx 12353$. These slight differences will not concern us: there is no way our model could ever be precise enough to make differences at this scale meaningful. systems, one never knows the values of the coefficients – in this case, the growth rate and the carrying capacity – exactly. Even in well understood situations, it would be rare to have more than two-digit accuracy in the coefficients. The simplest way to determine how sensitive your answers are to rounding off is to run the computer program with different values of the coefficients – for example, you might try running equation (2) with the values .041 in place of .042 and .0000033 in place of .0000034.

What happens when both guzzlers and chompers are present on the island? If they both eat the same food, then there will be less to go around. One way we could try to take this into account is to assume that rate of increase of one species is reduced by some amount proportional to the number of interactions between the species. (One might imagine, for instance, that if a bird of each species arrived at the same berry bush, then each bird would go away with less food; on the other hand, if the populations of both were small and on different parts of the island, so there was no interaction between the species, it is difficult to imagine that either would affect the growth rate of the other.) To a first approximation, the number of interactions between x individuals and y individuals is proportional to the product xy

(because, again to a first approximation, the number of interactions should be proportional to x and y separately). Thus, to take account of the presence of the other species, we might subtract some number times xy from equation (1) and some, possibly different, number times xy from equation (2). If one accepts this argument, a reasonable model for the interaction of two competing species would be:

$$\begin{aligned} x' &= ax - bx^2 - cxy \\ y' &= dy - ey^2 - fxy, \end{aligned} \tag{3}$$

where a, b, c, d, e, f are positive constants. For the sake of argument let's take a, b, d and e to be as in equations (1) and (2); that is, $a = .014, b = .0000014, d = .042, e = .0000034$. To guess at some sensible values of c and f, note that the term cxy must represent the decrease on average of the number of guzzlers per day resulting from competition. To pull numbers out of the air, we might imagine that if y were about 6000 (that is, about half their carrying capacity without guzzlers), the rate of increase of the guzzlers would be reduced by one third. Thus c might be $\frac{1}{6000} \cdot \frac{1}{3} \cdot (.014) \approx .0000008$. Similarly, we might imagine that if x were about 5000, the rate of increase of the chompers would be reduced by one fourth. Thus f might be $\frac{1}{5000} \cdot \frac{1}{4} \cdot (.042) = .0000021$. We emphasize, that we are just making numbers up here: the exact values are not important (but we do want to get the order of magnitude about right). Note, however, that these types of considerations suggest that b and c (respectively e and f) are roughly of the same order of magnitude (and many orders of magnitude less than a (respectively, d)). We wind up with the model

$$\begin{aligned} x' &= .014x_1 - .0000014x^2 - .0000008xy \\ y' &= .042y - .0000034y^2 - .0000021xy. \end{aligned} \tag{4}$$

Before checking what happens, we pause to ask what we would do about obtaining the numbers a-f for equation (3) in a real situation. We cannot emphasize too strongly that one should consult an ecologist: even measuring population levels in the wild is tricky, much less growth levels and carrying capacities. As you might expect from our discussion above, the numbers a, b, d, e are easier to measure than c and f. One frequently gets values by seeing what happens for different values of $a - f$ and choosing values which give equations whose behaviour is closest to that observed in the real situation. Systems like (3) are also used to test hypotheses: what values of the carrying capacities (b and e) are consistent with what is observed?; if you want to reduce the numbers of some pestiferous species and hold them fixed at some low value by introducing a competing species, what characteristics would the introduced species have to have?; what is the scale of interspecies rivalry (that is, the values $c$and f) consistent with what is observed? and so on.

Let us return to equation (4). What happens? Suppose, we start with 100 berry guzzlers and 200 berry chompers. We write a brief program that sketches the way in which x and y change over 20 years (240 months). We do the computations over time intervals of .1 months.

```
DEFDBL A-Z
SCREEN 12
WINDOW (0, 0)-(2400, 12500)
LET x = 100
LET y = 200
FOR N = 1 TO 24000
 xprime=(.014 -.0000014 * x-.0000008 * y)*x
 x= x + .1 * xprime
 PSET (N, x)
 yprime=(.042-.0000034 * y-.0000021 * x) * y
 y = y + .1*yprime
 PSET (N, y)
 NEXT N
PRINT x, y
```

Upon running this program, we find (see Figure 2) that the chompers increase very fast, reaching a population of almost 12000 by the end of the 20 years, whereas the guzzlers have barely limped up to a population of 649. Will the guzzlers go the way of the

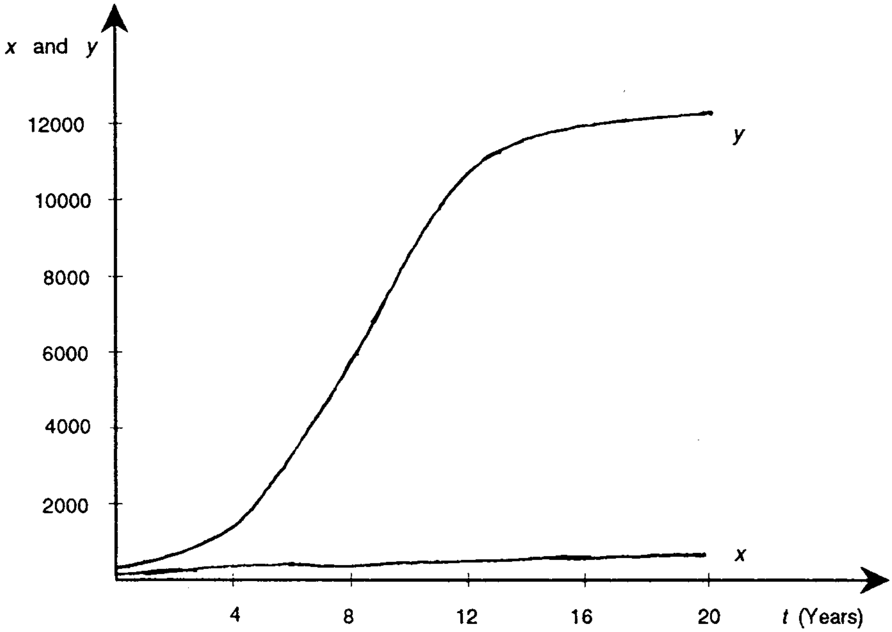

Figure 2. x and y over a 20 year interval

passenger pigeons and the dodo birds or will they rally and reach a respectable population?

To answer the question, we check what happens over the course of 200 years (2400 months). To do this we modify the program slightly, replacing the number 2400 in the `WINDOW` command by 24000, and changing the statement `FOR N = 1 TO 2400` to `FOR N = 1 TO 24000`. Running the program, we obtain Figure 3

The guzzlers do indeed rally, beating back the chompers! After 200 years, the number of chompers drops to 9550 and the number of guzzlers rises to 4539.

Below, we have plotted the values of (x, y) for different starting values $x(0)$ and $y(0)$ of x and y. We suppose $y(0) = 200$ and have sketched what happens to the state (x, y) (in the course of 100 years) when $x(0) = 2000, 4000, 6000$ and 8000, respectively. The following program plots x against y for different values of time, beinning with $x = 2000$ and $y = 200$ at time 0 and ending 100 years (= 1200 months)later. This curve is called a **trajectory**.

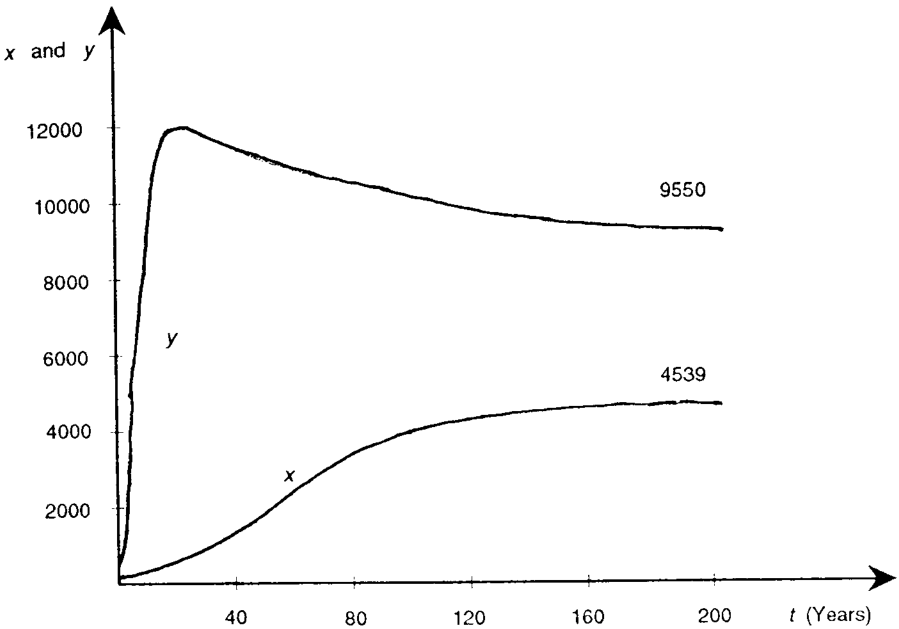

Figure 3. x and y over 200 years

```
DEFDBL A-Z
SCREEN 12
WINDOW (0, 0)-(10000, 13000)
LET x = 2000
LET y = 200
FOR N = 1 TO 12000
 xprime=(.014-.0000014 * x-.0000008 * y)*x
 x = x + .1 * xprime
 yprime=(.042-.0000034 * y-.0000021 * x)* y
 y = y + .1*yprime
 PSET (x, y)
 NEXT N
```

Adding another loop makes it sketch the trajectories from the four different starting values we want.

```
DEFDBL A-Z
SCREEN 12
WINDOW (0, 0)-(10000, 13000)
LET t = 0
FOR K = 1 TO 4
 LET x = 2000 * K
 LET y = 200
 FOR N = 1 TO 12000
  xprime=(.014-.0000014*x-.0000008*y)*x
  x = x + .1 * xprime
  yprime=(.042-.0000034*y-.0000021*x)*y
  y = y + .1 * yprime
  t = t + .1
  PSET (x, y)
 NEXT N
NEXT K
```

We obtain the following picture.

The numerical evidence suggests that we always end up with about 9545 chompers and 4545 guzzlers. To verify this, let's return to equation (4) and try to get some idea of what it says. We will do this by interpreting equation (4) geometrically. Before doing this, we need to recall some notions that you learned in school.

Pairs of Numbers and Points on the Plane

You probably learned in school that there is a one to one correspondence between numbers and points on a line. This is the starting point for one of the more fruitful lines of thought in Western intellectual history, namely Descartes' powerful synthesis, relating numbers and geometry. Descartes noted that every number could be identified with a point on a line. One chooses a point on the line (and calls it the origin), an orientation (that is a "positive" direction from that point), and a unit of length. Then, for example, the number -2.5 corresponds to a point 2.5 units from the origin in the direction opposite

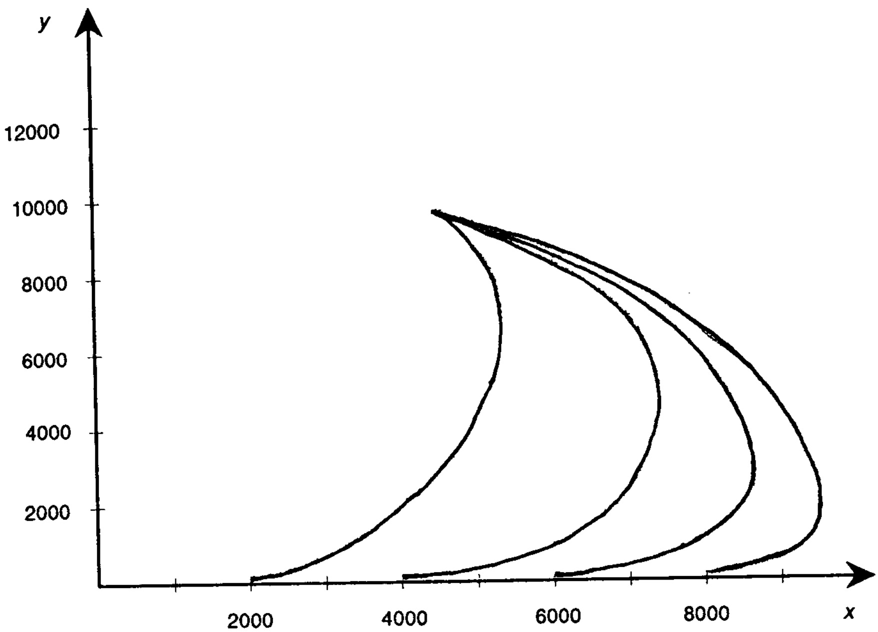

Figure 4. x and y for various $x(0)$ and fixed $y(0) = 200$

to the positive direction (the "negative" direction), the number π to the point π units from the origin in the positive direction. This leads to the familiar concept of the *number line*, a phrase meant to signal that the set of real numbers and the notion of a line are equivalent (see Figure 5).

Descartes went further. He saw that any point in the plane could be given by two numbers. One chooses an origin, a unit of length, two axes intersecting at the origin, and an orientation along each axis. On specifies an arbitrary point by two numbers, the first indicating how far it is along one axis (in the positive or negative direction according to whether the number is positive or negative), the second indicating how far it is along the other axis (with the same sign convention). Conversely, any pair of numbers corresponds to a point of the plane (see Figure 6). The set of all real numbers is usually

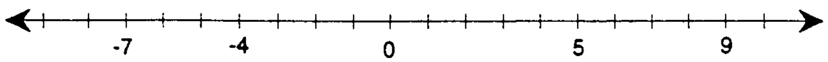

Figure 5. The correspondence between points on the line and numbers

denoted by R and the set of pairs of numbers by $\mathbf{R}^2$. Thus, there is a one-to-one correspondence between points of $\mathbf{R}^2$ and points in the plane.

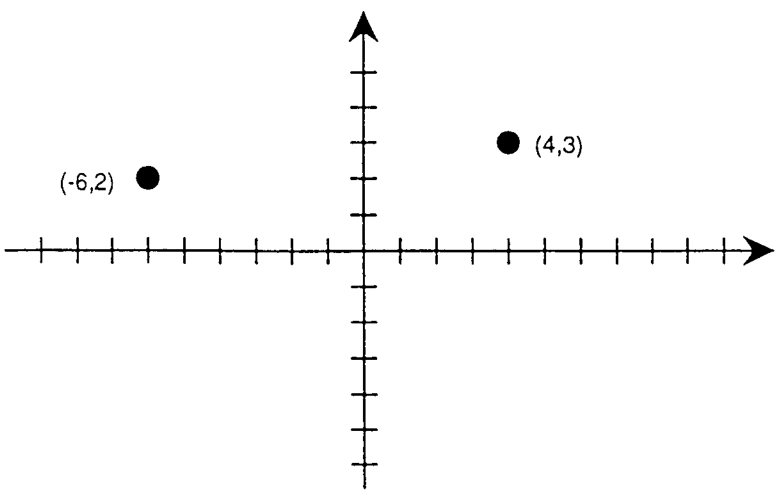

Figure 6. The correspondence between points on the plane and pairs of numbers

In our model, the point (x, y) on $\mathbf{R}^2$ corresponds to having

x guzzlers and y chompers. When we are thinking of a point in $(x, y) \in \mathbf{R}^2$ as corresponding to a situation we are modelling, we refer to it as a **state**. Of course, when modelling populations of two species, not every point of the plane is a state. To be a state (that is to correspond to a possible real situation), a point (x, y) in the plane must be such that $x \geq 0$ and $y \geq 0$. The set of all possible states, that is the first quadrant $\{(x, y) : x \geq 0, y \geq 0\}$, is called the **state space**.

Rates of Change of Pairs of Numbers

We can think of equation (4) as assigning to each point (x, y) of the plane, the numbers (x', y') according to the rule

$$\begin{aligned} x' &= .014x_1 - .0000014x^2 - .0000008xy \\ y' &= .042y - .0000034y^2 - .0000021xy. \end{aligned} \tag{4}$$

Thus, to the point (1000, 2000) corresponding to 1000 guzzlers and 2000 chompers, equation (4) assigns the pair (x', y') where

$$\begin{aligned} x' &= .014(1000) - 0000014(1000)^2 - .0000008(1000)(2000) \\ &= 14 - 1.4 - 1.6 = 11.0\text{guzzlers/month} \end{aligned}$$

and

$$\begin{aligned} y' &= .042(2000) - .0000034(2000)^2 - .0000021(1000)(2000) \\ &= 84 - 13.6 - 4.2 = 66.2\text{chompers/month}. \end{aligned}$$

We think of this pair as the "rate of change" (or "velocity") of the point $(1000, 2000)$. The first number 1000 of the pair is changing at a rate of 11.0 guzzlers/month, the second number 2000 at the rate of 66.2 chompers/month. (Both quantities are increasing – if one were decreasing, its rate of change would be negative.)

In terms of the language introduced a little earlier, equation (4) assigns to each state (that is, to each point (x, y) in the first quadrant of $\mathbf{R}^2$) a rate of change (x', y') of that state. This rate of change is also a pair of numbers, so that we could also picture it as a point in $\mathbf{R}^2$. However, this runs the danger of confusing rates of change of states and states. Instead we picture the rate of change (x', y') of a point (x, y) as an *arrow* beginning at (x, y) and going x' units in the horizontal direction (to the right if $x' > 0$, to the left if $x' < 0$) and y' units in the vertical direction (up if $y' > 0$ and down if $y' < 0$). Thus we picture the rate of change (11, 66.2) of the point (1000, 2000) as an arrow starting at (1000,2000) and going 11 units to the right and 66.2 units up. We have sketched this in Figure 7 below.

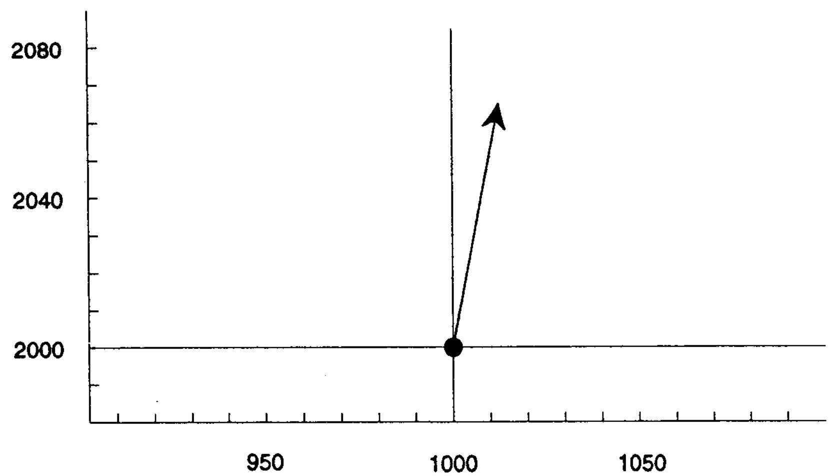

Figure 7. The state (1000, 2000) with velocity (11, 66.2)

If, instead, we wanted to represent a state (210, 315) with rate of change (3, -5), we would draw the arrow starting at the point (210, 315), but pointing down 5 units and to the right 3 units (see Figure 8). Similarly the arrow corresponding to the rate of change (-3, 5) points up and to the left and and that corresponding to (-3, -5) down and to the left.

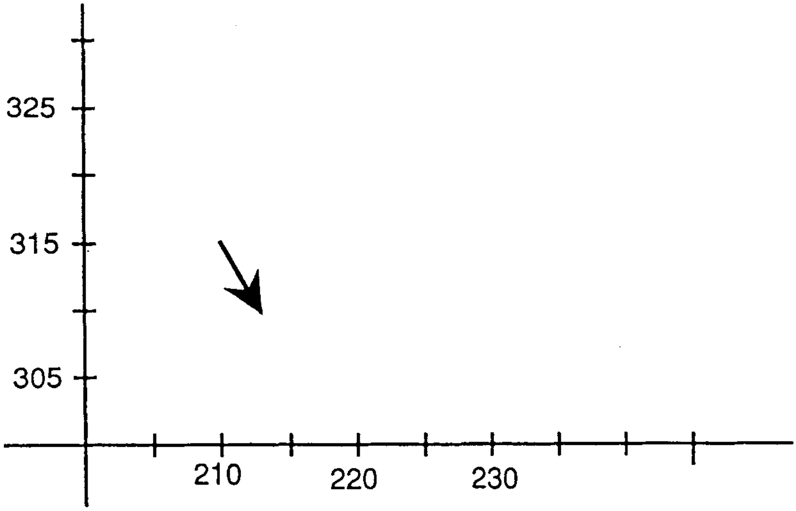

Figure 8. A point (210,315)with velocity (3,-5)

Rate Equations and Vector Fields

We have already stated that equation (4) can be thought of as assigning to each point (x, y) of $\mathbf{R}^2$ a rate of change (x', y'). We have just seen that we can picture the rate of change of a point as an arrow starting at that point. Thus, we can think of equation (4) as assigning to each point of $\mathbf{R}^2$ an arrow which represents the rate of change of the point. That is, we can think of equation (4) as attaching lots of little arrows to points on the plane, one arrow for each point. Since an arrow at a point is often called a **vector**, a collection of arrows attached to points of a set is called a **vector field** on that set. Equation (4) defines a vector field on $\mathbf{R}^2$.

In Figure 4, we plotted the points (x, y) for different values of time starting from some fixed point $(x(0), y(0))$ at time $t = 0$. We called these curves *trajectories*. We think of the state moving along the trajectory – at each point on the trajectory its rate of change is that given by equation (4).

Long Term Behavior

What's all this have to do with winding up with 9545 chompers and 4545 guzzlers, no matter how many of each we begin with?

First note, that almost every state (x, y) is changing. Those that aren't must satisfy $x' = 0, y' = 0$. Such states are important enough to warrant a separate name. In general, states (x, y) with the property that $x' = 0, y' = 0$ are called **equilibrium states** or **points**. Such states have rate of change (0,0). If we start at such a state, we will always remain at that state.

The equilibrium states of the vector field defined by equation (4) must satisfy

$$\begin{aligned} .014x_1 - .0000014x^2 - .0000008xy &= 0 \\ .042y - .0000034y^2 - .0000021xy &= 0. \end{aligned}$$

Rewriting gives

$$\begin{aligned} (14 - .0014x - .0008y)x &= 0 \\ (42 - .0034y - .0021x)y &= 0. \end{aligned}$$

Thus we have equilibria when $(x, y) = (0, 0)$; when $x = 0$, in which case $42 - .0033y = 0$, so that $y = 12253$; when $y = 0$, in which case $14 - .0014x = 0$, so $x = 10000$; and, finally, when

$$14 - .0014x - .0008y = 42 - .0034y - .0021x = 0,$$

in which case $x \approx 4545$ and $y \approx 9545$. We have sketched these equilibrium points, together with the lines $14 - .0014x - .0008y = 0$ and $42 - .0034y - .0021x = 0$ in Figure 9 below. (Note that the lines can be put into the usual slope-intercept form by dividing by the coefficient of y and transposing. Upon doing this, the equation of first line, call it Line I, becomes $y = -1.75x + 17500$ and that of the second, call it Line II, $y = -.618x + 12353$

Now we want to get an idea of how the other states are changing. We do this by determining which way the arrows point at various

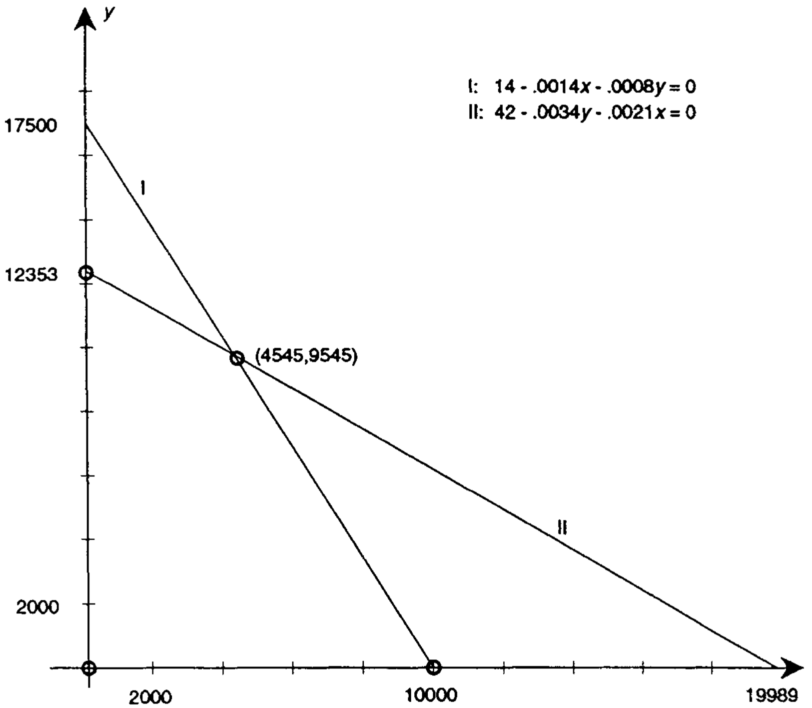

Figure 9. Equilibrium points of the vector field defined by equation (4)

places on the phase space, note that $x' < 0$ at all points of the first quadrant above Line I and $x' > 0$ at all points in the first quadrant below this line. So, in the first quadrant, all arrows above Line I point to the left, all arrows below the line to the right, and all vectors on the line are vertical. Similarly, since y' is less than zero above Line II, equal to zero on it, and greater than zero below it, all arrows in the first quadrant above the Line II point down, all arrows beginning on it are horizontal, and all vectors below it point up. We sketch some representative arrows (enormously magnified) in Figure 10.

The vector field shows clearly, that no matter where we start, the populations of guzzlers and chompers will eventually settle down to constant values (in our case, 4545 and 9545, respectively).

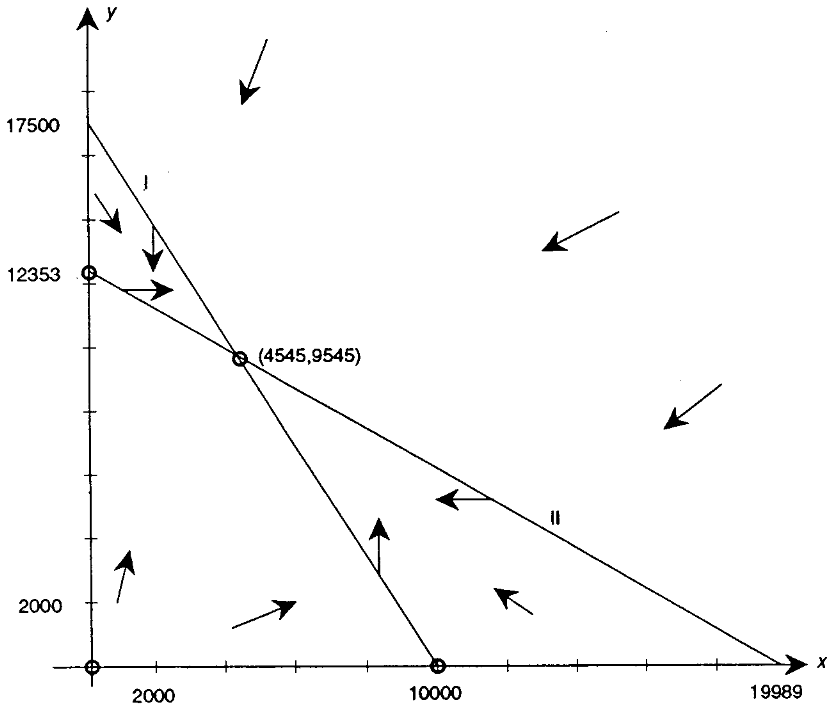

Figure 10. The vector field defined by equation (4)

Long Term Behavior of the General System

By generalizing the above argument, we can imagine the various possibilities for the vector field in the case of the general system represented by equation (3). Checking first where x' and y' are equal to zero, we have

$$
\begin{aligned}
x' &= 0 \text{ if and only if } x(a - bx - cy) = 0 \\
&\quad \text{if and only if } x = 0 \text{ or } a - bx - cy = 0 \\
y' &= 0 \text{ if and only if } y(d - ey - fx) = 0 \\
&\quad \text{if and only if } y = 0 \text{ or } d - ey - fx = 0.
\end{aligned}
$$

Thus the arrows are vertical on the y-axis and on the line $y = -\frac{b}{c}x + \frac{a}{c}$. The arrows are horizontal on the x-axis and on the line

$y = -\frac{f}{e}y + \frac{d}{e}$. We can assume, by interchanging the names of the variables that $\frac{a}{c}$, the y-intercept of the first line, is greater than $\frac{d}{f}$, the y-intercept of the second line (see Figures 11 and 12). We have two possibilities: either the lines meet or they don't. If they meet, we essentially have the situation in Figure 11. Both species coexist and their populations move towards some steady state.

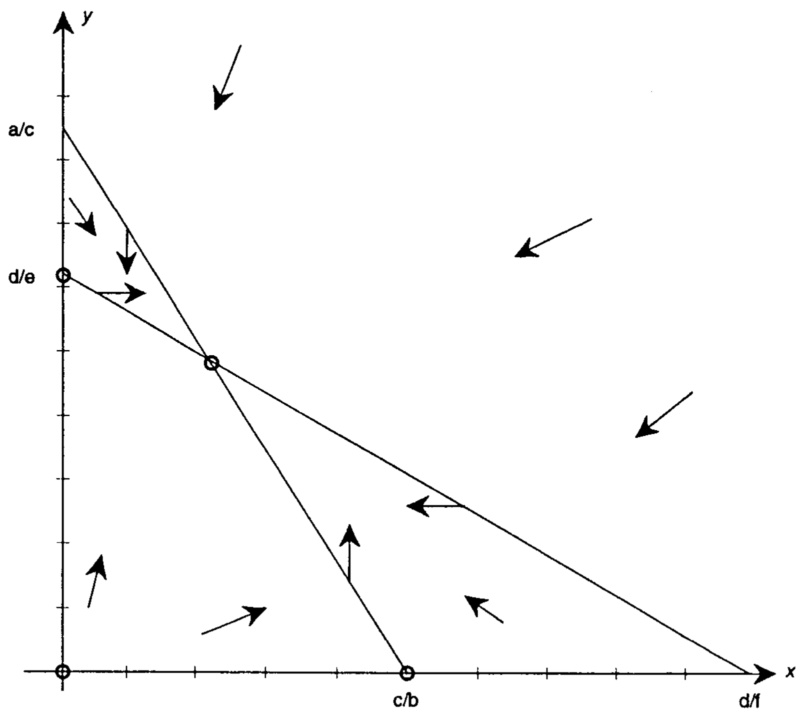

Figure 11. One possibility for the vector field defined by equation (3)

If the lines don't meet, then we get the a diagram of the vector field like that sketched in Figure 12. Notice that in this case, the two species do not co-exist. In the long run, the species corresponding to y will die out, and the other will flourish, tending towards a population of $\frac{c}{b}$ (that is, its carrying capacity).

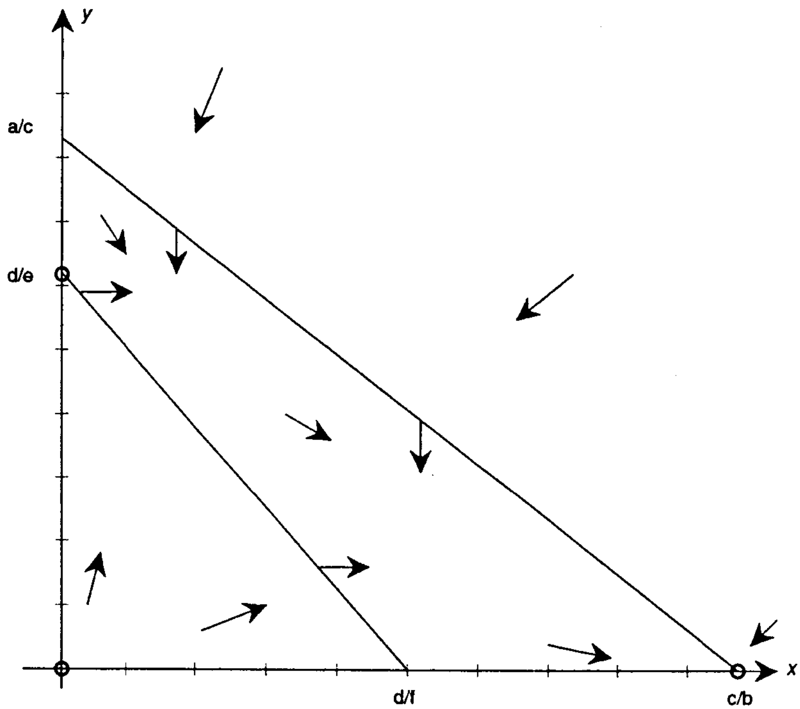

Figure 12. Another possibility for the vector field defined by equation (3)

Other Models of Competing Species

We stress that models represented by equation (3) are just one way of taking into account the interactions between two competing species. We could also take the interaction into account by assuming that the carrying capacity of the island for each species is reduced by the presence of the other. We might imagine, for instance, that if the number of chompers were close to their carrying capacity, then the carrying capacity of the guzzlers would approach zero and that if the number of chompers were small, then the carrying capacity of the guzzlers would not be affected too much. That is, the carrying capacity for the guzzlers is some fraction (between 0 and 1) times 10000 which aproaches zero as the number of chompers approaches 12500 and 1 as the number of chompers approaches zero. The simplest such fraction is $1 - \frac{y}{12500}$. Thus, we might assume that the carrying capacity for the guzzlers when there are y chompers on the island is

$$10000\left(1 - \frac{y}{12500}\right).$$

Similarly, the carrying capacity for the chompers when there are x guzzlers could plausibly be taken as

$$12500\left(1 - \frac{x}{10000}\right).$$

We arrive at the following model describing the the populations of guzzlers and chompers:

$$\begin{aligned} x' &= .014x\left(1 - \frac{x}{10000(1 - \frac{y}{12500})}\right) \\ y' &= .042y\left(1 - \frac{y}{12500(1 - \frac{x}{10000})}\right). \end{aligned} \tag{5}$$

These equations do not make biological sense if either $x \geq 10000$ or $y \geq 12500$ (why not?), so we assume that the state space is the set of all (x, y) such that $0 \leq x < 10000$ and $0 \leq y < 12500$.

For the simple models we have considered (those represented by equations of the form (3)), we see that two competing species always settle down to steady state populations. This is because the corresponding vector field is either like that sketched in Figure 11 or like that in Figure 12. All populations, not at an equilibrium state, tend to a unique eqilibrium state. Depending on the value of the parameters, the species can co-exist (that is, both steady state populations are greater than zero as in Figure 11) or one will die out (so, one of the steady state populations is zero as in Figure 12). These general conclusions can be shown to hold for more general models of two competing species. For more details, see the (very readable) paper by A. Recigno and I. Richardson or J. Maynard Smith's book cited in the bibliography. A more mathematical treatment is given in Chapter 12 of Hirsch and Smale's book.

Exercises on Competing Species.

Exercise 1. Write a program that graphs x against t if x' is given by the equation

$$x' = \frac{.014x}{1 - .0001x}.$$

You will need to choose a value of x when $t = 0$ – you should try different values. Explain why this is not likely to be a good model of population growth. (Hint: what happens when x gets near 10000?)

Exercise 2. If both species tend to a steady population, find the value for the equilibruium point in terms of a, b, c, d, e and f.

Exercise 3. Invent, and justify as plausible, a set of values for a, b, c, d, e and f under which one species would become extinct. (You can use chompers and guzzlers with different growth rates and carrying capacities, or you can imagine a different pair of species.)

Exercise 4. Write and run a computer program which follows the trajectories of a number of specific states for the model you constructed in the last exercise. Explain how the results fit in with sketch of the vector field displayed above.

Exercise 5. Analyze what could happen if $\frac{a}{c} = \frac{d}{f}$ in system (3) above.

Exercise 6. Write a computer program to sketch trajectories of the dynamical system given by equation (5) and run it for several different choices of initial conditions. What happens? Justify your conclusions by sketching the vector field. How does the behavior of this system differ from that of system (4)?

Exercise 7. What sort of behavior can you expect in general for models of the type exemplified by equation (5).

Exercise 8. Write a program to compute the trajectory of system (4) beginning at (1000, 2000), but do not use double precision. Discuss how your results differ from those we obtained above. This is an instance where very different (and spurious) conclusions can result from the numerical errors that result when a computer stores a number by just retaining a given number of digits and dropping the rest. (This is a phenomenon that one must keep in mind when using a computer to track the behavior of variables over time. Many newer computer languages, such as Mathematica, carry as many decimal places as needed to avoid round-off errors – however, older languages like Fortran and Basic truncate digits to a given number of decimal places. If you are making decisions based on a model of some situation, you should make sure not only that you can live with the assumptions made in making the model, but also that the conclusions drawn from computer analysis have not been vitiated by round-off errors. If a lot depends on the analysis, is well worth using a language that allows one to carry arbitrarily many decimal places and checking carefully to see how sensitive the answers are to truncation by running the program several times carrying different numbers of decimal places. If human life depends on the analysis, make sure someone has performed a full-scale error analysis (a topic outside te scope of this monograph).)

Species in Symbiosis

It is certainly not always the case that two species compete. There are many examples in nature of situations where two species are in *symbiosis*. This means that both species can survive quite well without the other, but that the growth rate of one or both is enhanced by the presence of the other.

An example of symbiosis is the crocodile and Egyptian plover bird. The plover bird gets meals by picking insects and leeches from the crocodile's teeth, while the crocodile benefits by getting its teeth cleaned and its mouth rid of harmful pests.

Let us assume that we have two species in symbiosis and that x is the amount of the first species and y the amount of the second species. As in the case of competing species, we assume that each species grows logistically in the absence of the other. However, to say that the species are in symbiosis can be reasonably interpreted as saying that the growth rate of one species will be increased by interactions with the other. Since the number of interactions can, to a first approximation, be thought of as proportional to the product of x

and y, we might write $x' = ax - bx^2 + cxy$ and $y' = dy - ey^2 + fxy$, where a, b, c, d, e, f are positive. (A better approximation to the number of interactions would require more knowledge about the specific species and might involve higher powers of xy.) We could allow $c = 0$ or $f = 0$ if one species enhances the other's growth, but not conversely. Putting these equations together gives a model for species in symbiosis:

$$\begin{aligned} x' &= ax - bx^2 + cxy \\ y' &= dy - ey^2 + fxy. \end{aligned} \tag{6}$$

Here all coefficients are positive (except that we allow either c or f to be zero). Note that this differs from the model represented by equation (4) for competing species only by the sign in front of the term xy.

In the case of symbiosis, we would certainly not expect one species to die out. To get some idea of what happens in equation (6), let's try a specific example. In fact, let's take a, b, c, d, e, f to be the same as in our guzzler-chomper example:

$$\begin{aligned} x' &= .014x - .0000014x^2 + .0000008xy \\ y' &= .042y - .0000034y^2 + .0000021xy. \end{aligned} \tag{7}$$

We can again write a program to see what happens. Let's try the initial values $x = 100$ and $y = 200$. A first run shows that the window (0, 0)-(1200, 12500) is too small, because the y values quickly exceed 12500. A little experimentation shows that replacing 12500 by 30000 (or anything higher) works well.

```
DEFDBL A-Z
SCREEN 12
WINDOW (0, 0)-(1200, 30000)
LET x = 100
LET y = 200
FOR N = 1 TO 12000
```

```
  xprime= (.014-.0000014*x+.0000008*y)*x
  x = x + .1 * xprime
  PSET (N, x)
  yprime= (.042-.0000034*y+.0000021*x)*y
  y = y + .1*yprime
  PSET (N, y)
  NEXT N
PRINT x, y
```

Running the program gives the results shown in Figure 13.

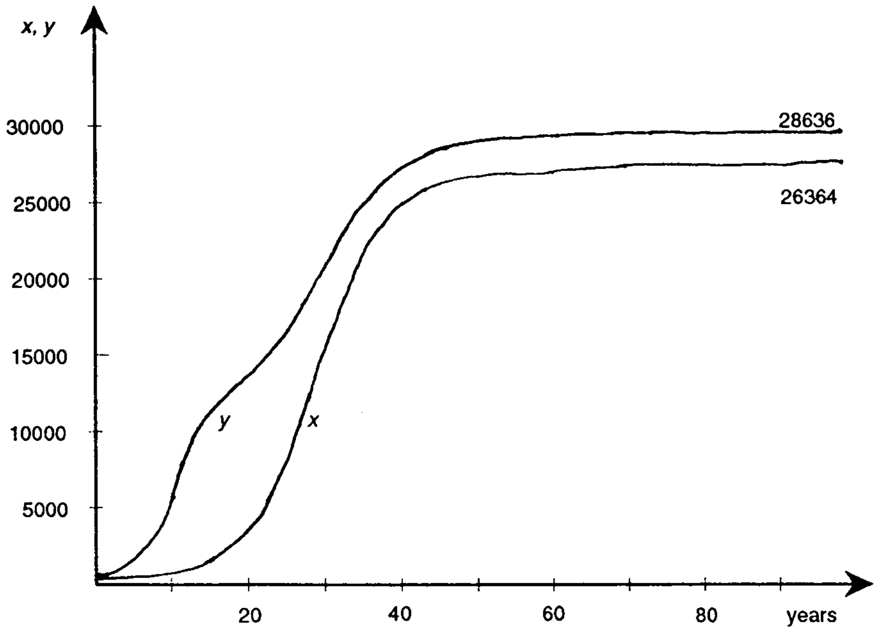

Figure 13. x and y as functions of time in the symbiotic system (7)

Note that x and y tend to the steady populations 26364 and 28636, respectively. The effect of the symbiosis is to allow far more individuals of *each* species to live on the island than would be possible if either species were alone on the island. In fact, the equilibrium population of either species is more than double the carrying capacity of the environment for that species!

In view of our experience with competing species, we might suspect that all trajectories in the state space (the first quadrant of $\mathbf{R}^2$), except those on the axes, tend to the point (26364, 28636). To check, we might try to see what happens for several different starting values of (x, y). As we did in the case of competing species, we suppose that $y(0) = 200$ and have sketched what happens to the state (x, y) (in the course of 100 years) when $x(0) = 2000, 4000, 6000$ and 8000, respectively. The following program is a simple modification of the one for competing species.

```
SCREEN 12
WINDOW (0, 0)-(30000,30000)
FOR K = 1 TO 4
 LET x = 2000 * K
 LET y = 200
  FOR N = 1 TO 12000
  xprime=(.014-.0000014*x+.0000008*y)*x
  x = x + .1 * xprime
  yprime=(.042-.0000034*y+.0000021*x)* y
  y = y + .1 * yprime
  PSET (x, y)
 NEXT N
NEXT K
```

Running the program produces the results shown in Figure 14.

These strongly support our suspicion that all trajectories tend to the point (26364, 28636). To check that this is indeed the case, we roughly sketch the vector field. As in the case of competing species, we verify that the arrows are vertical on the lines $x = 0$ and $14 - .0014x + .0008y = 0$ and horizontal on the lines $y = 0$ and $42 - .0034y - .0021x = 0$. Working out the signs produces the results shown in Figure 15. It shows clearly that, no matter where we start (other than on the axis), the trajectory does indeed tend to the equilibrium point (26364, 28636).

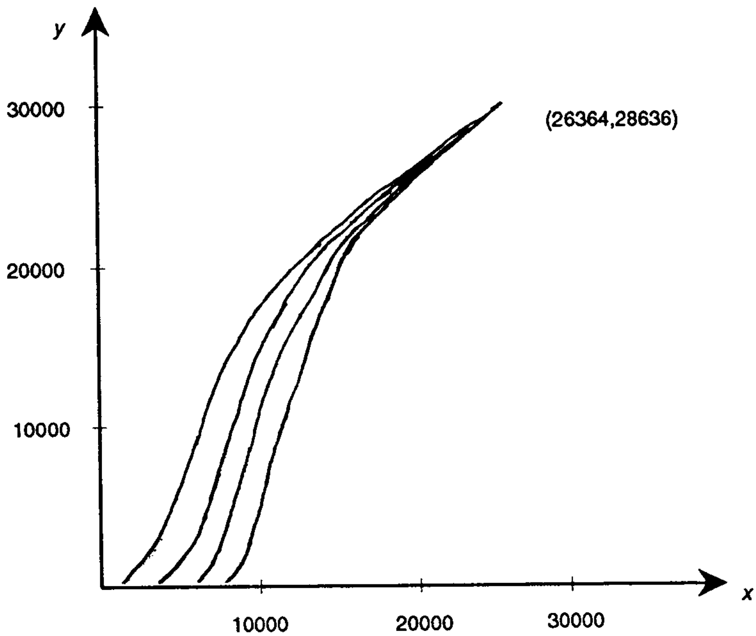

Figure 14. x and y for $x(0) = 200, 400, 600, 800$ and fixed $y(0) = 200$

Exercises on Species in Symbiosis

Exercise 1. Verify the details shown in Figure 13 of the vector field defined by equation (7). In particular, work out the intercepts and all equilibrium points. Show that the arrows do indeed point in the directions indicated.

Exercise 2. Analyze the possibilities for the vector fields determined by the equations having the form (6). (Hint: As in the case of competing species, consider the possible dispositions of the lines on which $x' = 0$ and $y' = 0$.

Exercise 3. From the last exercise you know that for certain parameter values, the x or y coordinates along the trajectories of

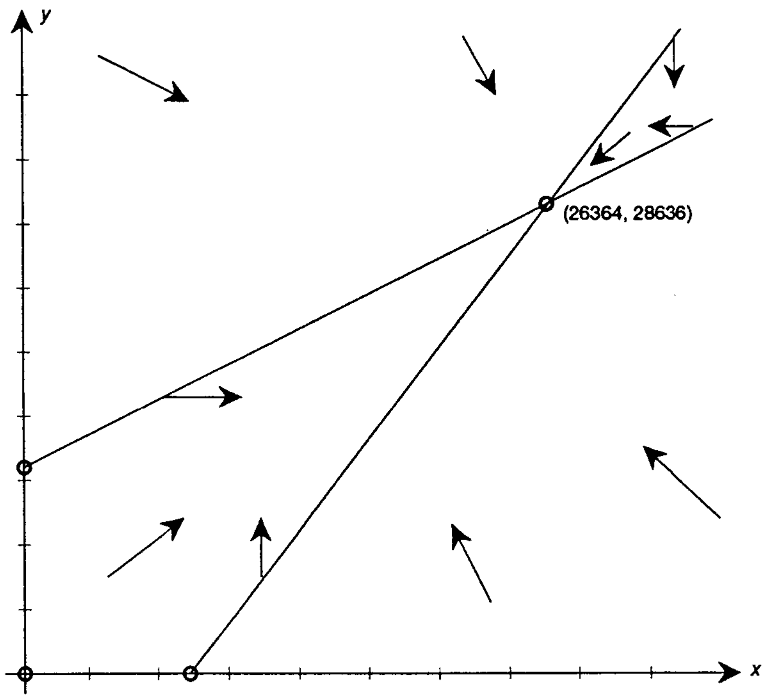

Figure 15. The vector field defined by the equation (7)

equation (6) can tend to infinity. Discuss why this is impossible on biological grounds. (Challenge: Show that this does not happen if one subtracts from each equation. One can think of this as a "second-order interaction effect" which damps down the positive effect of interactions on the growth rate when either x or y is large.)

Exercise 4. We could have tried to model symbiosis starting with the assumption that one species increases the carrying capacity of the environment for the other without affecting the growth rate. Suppose we have two species which grow logistically in the absence of the other. Suppose further that the growth rate of one is .01 with carrying capacity 10000, and the growth rate of the other is .03

and the carrying capacity 25000. Write down a dynamical system which models a symbiotic relationship between the species which does not change the growth rate of either species, but which doubles the carrying capacity of the environment for each. How do these equations compare with equations of the form (7)?

Predator-Prey Models

There is still another way in which we can imagine two species interacting. One species can prey on another: that is, one species is the other's food. This is the case with the brown snake and any bird species on Guam. Let us return to our island and suppose that, instead of snakes and birds, it is populated with rabbits and foxes. Let us imagine further that the rabbits live on the abundant native berries and that the rabbits are the sole food supply of the foxes. (This is decidedly *not* the case for the brown snake – it has a very varied diet.) We again ask what happens over the course of time. Will the number of foxes and rabbits reach a steady state? Might one or both species become extinct?

Let x be the number of rabbits (or, if you prefer, the total mass of the rabbits measured in units of average rabbit size) and y be the number of foxes (or, alternatively, the total mass of the foxes measured in units of average fox size). To determine how x and y change with time, we try to write equations for x' and y'.

We make the following assumptions.

- In the absence of foxes, the rabbits grow logistically.
- The population of rabbits declines at a rate proportional to the product xy. This is reasonable in that the number of encounters between rabbits and foxes will, to a first approximation, be proportional to the product of x and y.
- In the absence of rabbits, the foxes die off, at a rate proportional to the number present.
- The fox population increases proportional to the number of encounters between rabbits and foxes. Again, to a first approximation, this says that there is an increase in the fox population proportional to xy.

We can express the above as equations for x' and y' as follows

$$\begin{aligned} x' &= a\,x - b\,x^2 - c\,xy \\ y' &= -d\,y + e\,xy\,, \end{aligned} \tag{8}$$

where a, b, c, d, and e are positive constants which would have to be determined through field observations. These equations are called the **Lotka-Volterra equations with bounded growth.**

To see how the system behaves, we make up some values for the coefficients. Suppose

$$\begin{aligned} a &= .1 \text{ rabbits per month per rabbit} \\ b &= .00001 \text{ rabbits per month per rabbit} \\ c &= .004 \text{ rabbits per month per rabbit-fox} \\ d &= .04 \text{ foxes per month per fox} \\ e &= .00002 \text{ foxes per month per rabbit-fox} \end{aligned}$$

With these choices of parameters, we obtain the following system

$$\begin{aligned} x' &= .1x - .00001x^2 - .004xy \\ y' &= -.04y + .00002xy. \end{aligned} \tag{9}$$

To see what happens, suppose that we start off with 1000 rabbits and 10 foxes. The following program is a modification of the program for competing species and plots the values of x and y (actually $50 \cdot y$ – see below) over a hundred years (1200 months). We have taken $\Delta t = .1$ month.

```
DEFDBL A-Z
SCREEN 12
WINDOW (0, 0)-(1200, 10000)
LET x = 1000
LET y = 10
FOR N = 1 TO 12000
```

```
    xprime = .1 * x -.00001 * x * x -.004*x * y
    x = x + .1 * xprime
    PSET (t, x)
    yprime = -.04 * y + .00002 * x * y
    y = y + .1*yprime
    PSET (t, 50*y)
    NEXT N
  PRINT x, y
```

In order to plot the foxes and rabbits on the same graph, note that we plot $50 \cdot y$ instead of y. (There will clearly be far fewer foxes than rabbits. If we assume that the average fox needs a rabbit a week to stay healthy, then a fox will eat about 50 rabbits a year, so 50 times the number of foxes should be the same order of magnitude as the number of rabbits). Running the program gives the graphs in the Figure 16.

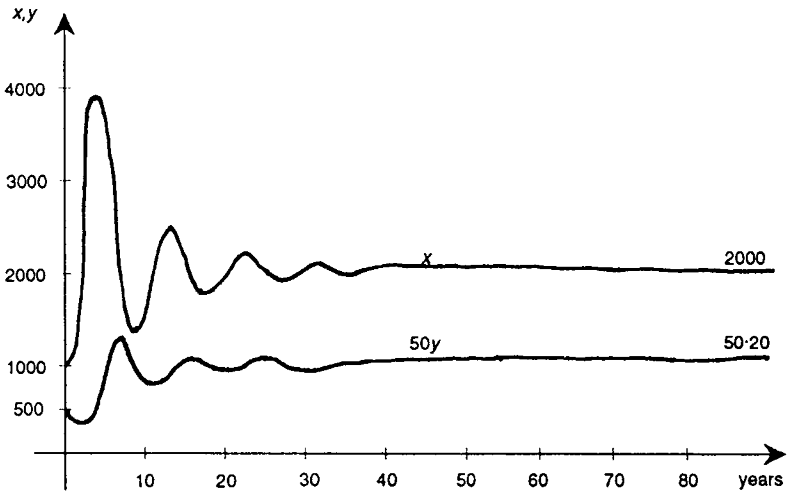

Figure 16. Rabbit and fox populations as functions of time

For these initial values, the number of rabbits and foxes tends to a steady state, with the number of rabbits approaching 2000 and the number of foxes 20. The number of rabbits gets nowhere near the carrying capacity (which is 10000, for the values we chose): the maximum number of rabbits is about 3900 and occurs after 40 months.

Notice that there is a new feature here. The numbers of rabbits and foxes oscillate — their numbers rising and falling in a regular manner — the oscillations gradually dying out towards the steady state. This is particularly noticeable if we follow the trajectory in the state space. The following program produces a picture of the trajectory (we have adjusted the window size to take account of the fact that the rabbit and fox populations never exceeded 5000 and 50, respectively.)

```
DEFDBL A-Z
SCREEN 12
WINDOW (0, 0)-(5000, 50)
LET x = 1000
LET y = 10
FOR N = 1 TO 12000
  xprime = .1 * x -.00001 * x * x -.004* x * y
  x = x + .1 * xprime
  yprime = -.04 * y + .00002 * x * y
  y = y + .1*yprime
  PSET (x, y)
  NEXT N
PRINT x, y
```

Running the program gives the picture in Figure 17, which makes it clear that x and y are oscillating about a fixed point, with the amplitudes of the oscillations slowly dying out.

To sketch the trajectories beginning at the values (2000, 10), (4000, 10), (6000, 10) and (8000, 10), we run the program following.

```
DEFDBL A-Z
SCREEN 12
```

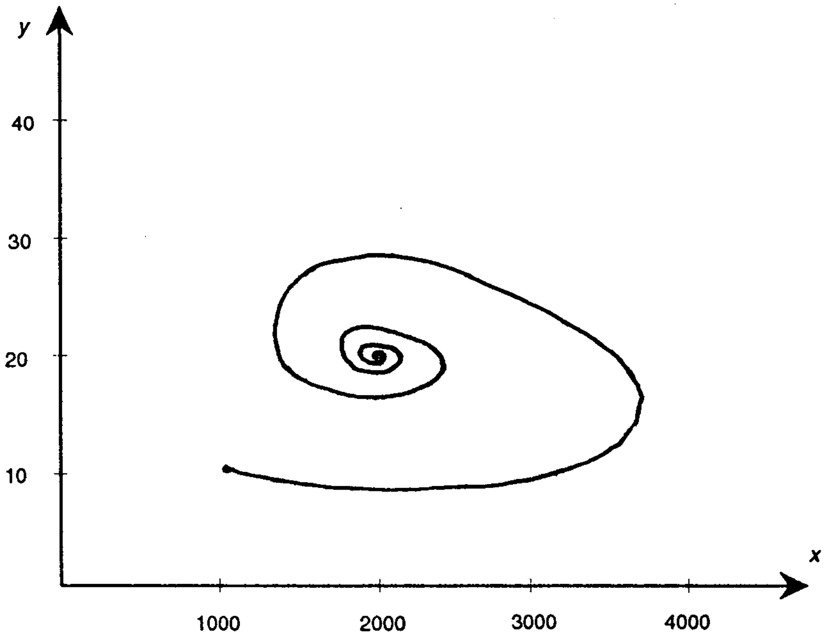

Figure 17. The trajectory of equation (9) beginning at (1000, 10)

```
WINDOW (0, 0)-(10000, 50)
FOR K = 1 TO 4
 LET x = 2000 * K
 LET y = 10
 FOR N = 1 TO 12000
  xprime = (.1 - .00001 * x - .004 * y)*x
  x = x + .1 * xprime
  yprime = -.04 * y + .00002 * x * y
  y = y + .1 * yprime
  PSET (x, y)
 NEXT N
NEXT K
```

The resulting picture, Figure 18, strongly suggests that all trajectories in the first quadrant, except those on the axes, tend to the point (2000,20).

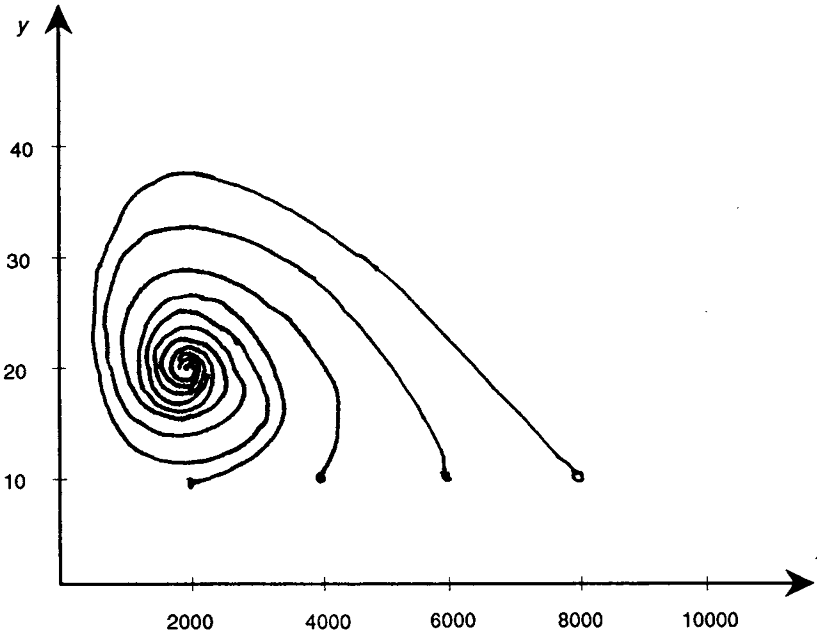

Figure 18. The Trajectories of equation(9) beginning at (2000,10), (4000,10), (6000,10) and (8000,10)

To sketch the vector field, note that the arrows are vertical on the lines $x = 0$ and $.1 - .00001x - .004y = 0$ and horizontal on the lines $y = 0$ and $-.04 + .00002x = 0$. Simplifying, the arrows are vertical on the y-axis and the line $y = -.0025x + 25$ and horizontal on the x-axis and the line $x = 2000$. We have sketched this, together with some representative arrows (hugely magnified) in Figure 19.

The Lotka-Volterra model

Historically, a model of great importance was the so-called *Lotka-Volterra model.* It was developed independently by the Italian mathematical physicist Vito Volterra in 1925-26, and by the mathematical ecologist and demographer Alfred James Lotka a few years earlier. Lotka had observed that the populations of several

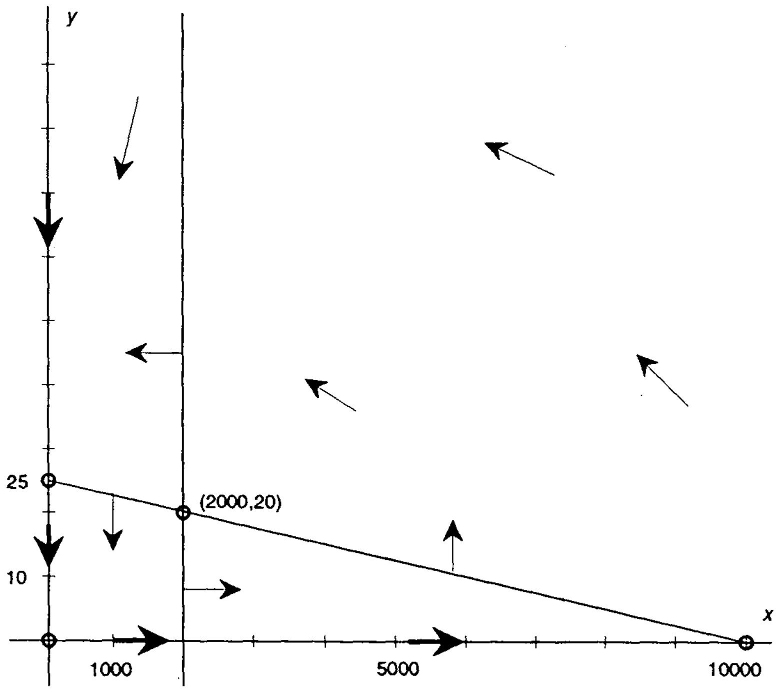

Figure 19. The vector field defined by equation (9)

species of fish in the upper Adriatic oscillated in a regular way. The model we introduced above predicts some oscillations which die out as the populations approach a steady state. The Lotka-Volterra model predicts oscillations which do not die out. Although this model is clearly flawed (as we shall see), it is still of great importance in ecological modeling.

To describe the model, we stick with rabbits and foxes. As above, we let x be the number of rabbits and y the number of foxes. We make the following assumptions regarding the rates of change x' and y' of x and y.

- The birth rate of rabbits is proportional to the number of rabbits. (This is clearly unrealistic if there are no foxes around and the number of rabbits is large – if there are no foxes, it predicts that the number of rabbits grows without bound.)

- The death rate of rabbits is proportional to the number of rabbit-fox interactions. (Rabbits never die of "natural causes"; they just get eaten.)
- The birth rate of foxes is proportional to the number of rabbit-fox interactions. (The better fed the mother fox is, the more viable offspring she will produce. In lean times, predators simply don't reproduce.)
- The death rate of foxes is proportional to the number of foxes.

The following system of equations incorporates these assumptions

$$\begin{aligned} x' &= ax - bxy \\ y' &= -cy + dxy \end{aligned} \tag{10}$$

where a, b, c, and d are positive constants.

Suppose we measure the time t in months, and suppose that

$a =$ .1 rabbits per month per rabbit
$b =$ .004 rabbits per month per rabbit-fox
$c =$ .00002 foxes per month per rabbit-fox
$d =$ .04 foxes per month per fox

This leads to the model

$$\begin{aligned} x' &= .1x - .004xy \\ y' &= -.04y + .00002xy. \end{aligned} \tag{11}$$

To see what happens, we start off with 2000 rabbits and 10 foxes, then with 4000 rabbits and 10 foxes, then 6000 and 10, then 8000 and 10. The following program plots the corresponding trajectories (over a period of 100 years).

```
DEFDBL A-Z
SCREEN 12
WINDOW (0, 0)-(10000, 100)
```

```
LET t = 0
FOR K = 1 TO 4
  LET x = 2000 * K
  LET y = 10
  FOR N = 1 TO 12000
  xprime = .1 * x - .004 * x * y
  x = x + .1 * xprime
  yprime = -.04 * y + .00002 * x * y
  y = y + .1 * yprime
  PSET (x, y)
  NEXT N
NEXT K
```

Running the program gives the results shown in Figure 20.

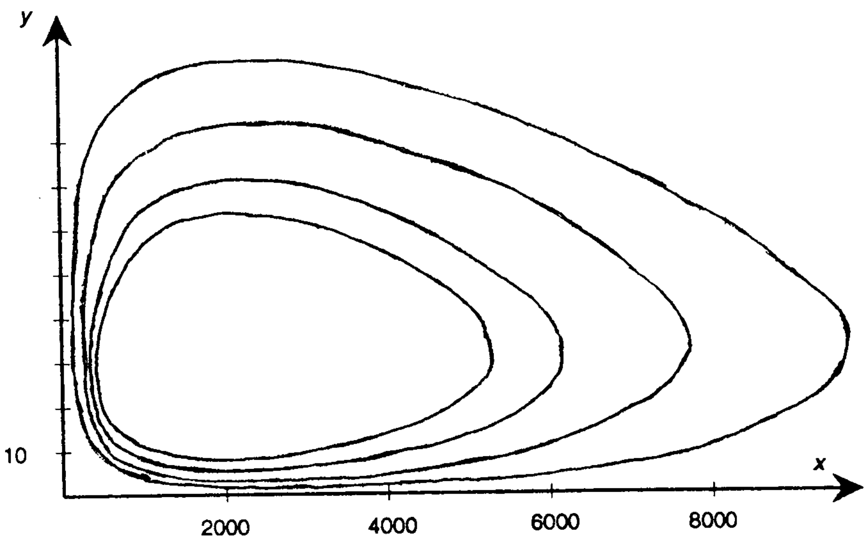

Figure 20. The trajectories of equation (11) through (2000,10), (4000,10), (6000,10) and (8000,10)

We see that the trajectories are closed curves, indicating that the populations of the rabbits and the foxes exhibit *cyclical behavior* — their numbers rise and fall in a regular manner, returning periodically to their original values.

The Lotka-Volterra equations give rise to cyclical behavior and, at first blush, seem to give some indication of how cyclical changes in population might result. However, there are a number of difficulties with the equations. First, killing a few foxes creates huge variations in the maximum number of rabbits. In data taken from actual environments, wild swings in population size are not always observed. Secondly, in the absence of foxes, the equations imply that the number of rabbits will keep growing forever. This is patently absurd.

Structural Stability

The most serious difficulty with the Lotka-Volterra equations involves the subtle, but extremely important, notion of structural stability.

We say that a system of equations is *structurally stable* if arbitrarily small changes to the equations do not change the nature of the solutions.

To explain this, consider again the specific example (11) of the Lotka-Volterra system studied above. Changing one or more of the coefficients .1, -.004, -.00002, .04 slightly, by amounts within $\pm.000001$, say, does not change the nature of the solutions too much – they are still closed curves. (Verify this by a writing a program to sketch solutions of

$$x' = .0999999x - .004001xy$$
$$y' = -.04y + .000021xy,$$

say, beginning at (2000,10), (4000,10), (6000,10) and (8000,10).) However, it would be a mistake to assume that this means that system (11) is structurally stable: .1, -.004, -.00002 and .04 are not the only coefficients that should be considered. We could have written (11) as

$$x' = .1x - .004xy + 0x^2 + 0x^2y + \ldots$$
$$y' = -.04y + .00002xy + 0y^2 + 0x^2y + \ldots, \tag{11'}$$

which makes it clear that 0 occurs a lot of times as a coefficient. If we change the coefficient 0 of x^2 even by the smallest amount, say to -.000001, then we get the system

$$x' = .1x - .004xy - .000001x^2(+0x^2y + \ldots)$$
$$y' = -.04y + .00002xy(+0y^2 + 0x^2y + \ldots),$$

which has solutions of an entirely different nature. In fact, the behavior is similar to that described in equation (8) – after a few oscillations, the populations settle down to fixed numbers. Thus the trajectories of the system change radically by changing the equations slightly. They go from being a set of nested closed curves to a set of curves all of which spiral into one point. (Verify this by writing a program to check it.) The same would be true if we took the coefficient of x^2 equal to -.00000001 or -.0000000001 or *any* negative number, no matter how small.

Had we taken the coefficient of x^2 to be any positive number, no matter how small, the trajectories would all spiral outward from a single point leading to populations of rabbits and foxes which grow without bound.

This means the Lotka-Volterra system (11) is not structurally stable – the same reasoning can be used to establish that no systems of the form (10) are structurally stable (changing the coefficient of x^2 will change the nature of the solutions.

There is a wrinkle worth pointing here. Namely, no term involving pure powers of y (including constants) can occur in the equation for x' (and no term involving pure powers of x can occur in the equation for y'. Thus it would not make sense to say that we were going to test for structural stability of the system (11) by changing the coefficient 0 of y in the equation

$$x' = .1x + 0y - .004xy + 0x^2 + 0x^2y + \ldots$$

to a nonzero number. The reason is that such an equation would not satisfy the biological constraint that $x' = 0$ when $x = 0$ (if there are no rabbits, they can't reproduce, and their rate of change must be 0). However, there is no biological reason why the coefficient of the term x^2 should be 0. In fact, there are good reasons why it should not be. If the coefficient of x^2 (and every other larger power of x) is 0, it can be shown that these equations say that if there are no foxes, the population of rabbits will increase indefinitely, But this cannot happen (either food or space will eventually run out).

It can be shown that all the other systems of equations we have studied (the logistic equation (2) of Chapter 1 and the competing species system (3), the symbiotic system (6) and the Lotka-Volterra system with bounded growth (8) of this chapter) are structurally stable (in the set of systems for which $x' = 0$ when $x = 0$ and $y' = 0$ when $y = 0$). The techniques involved are far outside the scope of this manuscript

Generally speaking, biological systems should always be modelled by structurally stable systems. The reason is that no two biological systems can be exactly the same, even from day to day. Thus systems of equations exactly describing two similar biological systems, or the same system at different times would have to have slightly different coefficients. Yet the behavior in similar systems is similar. This suggests that small changes in the equations do not change the nature of the solutions. That is, the system of equations should be structurally stable.

The same rule of thumb holds for modelling situations in other areas in which one observes regularities in behavior, but in which there are likely to be large numbers of confounding factors which are difficult to quantify. The one notable exception is in the physical sciences – here one often knows the equations exactly precisely because they apply to situations which are simple compared with biological systems.

There is another good reason to insist on structural stability of the systems of equations being used. The purpose of modelling is to get a good idea of what will happen to a system over the long term: will a certain species die out? what would happen if the fox population on an island were halved? The process of modelling is not used to generate specific numerical predictions, but to explore a complicated system and to see what the possibilities are. In such cases, one has very little idea of anything other than the magnitude of coefficents – some times not even that. Any conclusions that one draws cannot be dependent on accidents in the way the coefficients have been chosen.

Other Predator-Prey Models

The following predator-prey model, due to R.M. May, is an attempt to incorporate more natural assumptions regarding the effects of encounters between predators (foxes) and their prey (rabbits). As we will see, it predicts oscillatory behavior, but that behavior is rather different than that predicted by the Lotka-Volterra model. the sense that we have defined above.

In order to work with quantities which are roughly the same size, we let x be the number of *hundreds* of the rabbits (alternativley, the bimass of the rabbits in units of *one hundred times* the average rabbit size) and y be the number of foxes (or the biomass of the foxes in units of average fox size). Thus, $x = 5$ means that there are 500 rabbits. We make the following assumptions about the rates of change of the rabbit and fox populations.

- In the absence of foxes, the rabbits grow logistically.
- The death rate of the rabbits is proportional to x and y when x is small, but only to y when x is large. (This is intended to model a situation in which a small number of rabbits will be more widely scattered and, hence, able to hide more easily, whereas a larger number of rabbits will mean that the rabbits will be easily caught by the foxes.)
- The rate of change of the fox population is also logistic, but the carrying capacity is proportional to the number of rabbits.

These assumptions lead to the following equations for x' and y'

$$\begin{aligned} x' &= ax(1 - \frac{x}{b}) - \frac{cxy}{x+d}, \\ y' &= ey(1 - \frac{y}{fx}), \end{aligned} \tag{13}$$

where a, b, c, d, e and f are positive constants.

$$\frac{c}{x+d} = \frac{c}{d(\frac{x}{d}+1)} = \frac{c}{d}\frac{1}{1+\frac{x}{d}},$$

so that when x is small (and, in particular, smaller than d), x/d is small and

$$\frac{c}{x+d} \approx \frac{c}{d}.$$

Thus the death rate $\dfrac{cxy}{x+d}$ of the rabbits is approximately equal to

$$\frac{c}{d}xy,$$

hence proportional to xy. In order to see what happens here, take $a = .6, b = 10, c = .5, d = 1, e = .1$ and $f = 2$ and start with with 2000 rabbits (so $x = 20$) and 10 foxes. We modify one of our programs above. After a little experimentation to get the window size right, we obtain the following.

```
DEFDBL A-Z
SCREEN 12
WINDOW (0, 0)-(20, 20)
LET x = 20
LET y = 10
FOR N = 1 TO 1200
  xprime = .6*x*(1-x/10)-(.5*x*y/(x+1))
  x = x + .1 * xprime
  yprime = -.1 * y *(1-(y/(2*x))
```

```
y = y + .1*yprime
PSET (x, y)
NEXT N
```

Running this program gives the following output. We have also sketched the trajectory starting with 400 rabbits and 4 foxes.

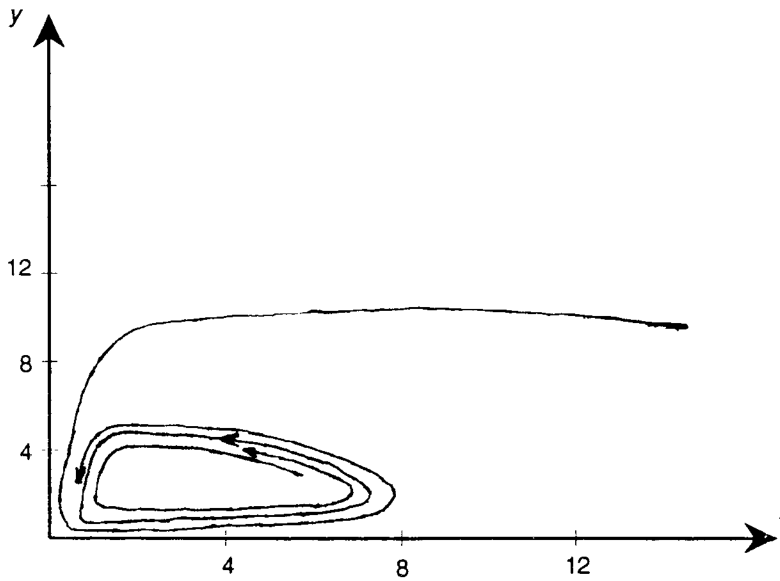

Figure 21. The trajectories of system (13) starting at (20,10) and (4,4)

After a remarkably short period the populations settle down to cyclical behavior. Moreover, this behavior is the same no matter at which point we start. Note that the trajectories wrap themselves onto a single 'closed' trajectory. The word 'closed' refers to the fact that there is no beginning or end of the trajectory – it is like a misshapen circle. Such a trajectory is called a *limit cycle*. This is a new type of equilibrium behaviour – instead of settling down to a fixed population of predators and preys, the populations settle down to a cyclical behavior which is the same for all initial states in some region of the state space. There is an equilibrium here, but it is one that is in motion.

Exercises for Predator-Prey Systems

Exercise 1. Sketch the vector field corresponding to the general system (8). What general conclusions can you draw?

Exercise 2. Invent an argument to show that the orders of magnitude of the parameters we chose in equation (8) to get equation (9) are reasonable. For example, rabbits are known to be tremendously prolific. Show that if you suppose that, without the foxes around, and with a small population, the average rabbit lives 40 months and the average female produces one litter of 3 rabbits a year, you get the value of a we chose above. Ask yourself what sort of carrying capacity for the rabbits is reasonable, how many rabbits per month a fox is likely to need to eat (try 4 to start), and so on.

Exercise 3. Sketch x and y as functions of time along each of the four trajectories of equation (11) beginning at (2000,10), (4000,10), (6000,10) and (8000,10) above (you will have to modify the program to plot x and y as functions of t).

Exercise 4. How long does it take for the populations to return to their initial values for each of the four initial values in the previous exercise?

Exercise 5. Sketch the vector field corresponding to the system represented by equation (11). What will it look like for an arbitrary system of equations of the form given by equation (10)? Note how difficult it is to tell from the vector field whether the trajectories spiral in to an equilibrium point or whether they close up. Can you think of an argument to distinguish the two cases?

Exercise 6 Sketch some of the trajectories of the system

$$x' = .1x - .004xy - .000001x^2(+0x^2y + \dots)$$
$$y' = -.04y + .00002xy(+0y^2 + 0x^2y + \dots).$$

Explain why this shows that system (11) is not structurally stable.

Exercise 7. Show that the system

$$x' = -y$$
$$y' = x$$

is not structurally stable (Hint: first, sketch the solution curves of the system.)

The Brown Tree Snake again

The assumptions of the predator-prey models do not apply to the brown tree snake. The reason is that the brown tree snake has an extremely varied diet and does not limit itself to one species. As we mentioned, the snake has almost completely wiped out the bird populations on Guam. Why then, has the snake population not dropped? There are at least two reasons. The first is that while the snake can consume enormous quantities of prey at one go (one snake was found inside a bird cage with four bumps – it ate all the birds in the cage in one night), it can also live for a long time on almost nothing. The second is that while the snake seems to clearly prefer birds, it also feeds on lizards, such as geckos and skinks, and small mammals. Julie Savidge showed that there was a noticeable drop in the small mammal population on Guam. The fact that the snake can live on skinks and geckos is bad news – these animals have an enormous reproductive rate and quite likely will allow the snake population to maintain its bloated size. It seems likely that because of the absence of large prey, not as many snakes will grow to their full eight foot size. However, there will be just as many four foot snkes.

There is mounting concern about the snake in Hawaii. There are no snakes native to the Hawaiian Islands. However, there have been at least six documented brown snake sightings. One was found crawling in the Pan Am customs area of Honolulu International Airport (and killed by a Pan Am employee). A second was found dead in 1986 near an aircraft hanger at the Barbers Point Naval Air Station on the other side of Oahu. Another two were found in 1989 at Hickham Airforce Base (also on Oahu), one clinging to a fence guy wire near a stream (an Army sergeant beat it to death with a piece of metal), the other dead beneath a cargo plane. Two more were found dead near the runway at Honolulu International Airport in 1991. In all these cases, the snake had apparently stowed away on a flight from Guam to Oahu. A number of other sightings have been reported, but no snakes were found.

There are two commercial flights a day from Guam to Hawaii and the military averages 10 flights a week. In addition, military families are regularly transferred from Guam to Hawaii along with their personal goods, cars, furniture, etc. The possibility of a snake getting through undetected among personal effects is extremely worrisome. Once established on Hawaii, it is widely agreed that the snake would have a field day. Hawaii has a fauna that is composed of a lot of introduced species (for example, rats, mice, lizards, doves). Moreover, some of these species have huge populations – there are little Chinese doves everywhere. The brown snake could reach population levels that make Guam look sickly.

The Hawaii Electric Company has funded a number of programs to make people aware of the problem. In addition, the Hawaiian Department of Land and Natural Resources has set up seven SWAT (Snake Watch and Alert Teams) teams, two on Oahu and one each on neighbor islands, for the purpose of tracking down and eliminating introduced snakes.

Ecosystems with more than Two Species

Real ecosystems involve far more than two species. In such cases there will be more than two quantities in which we will be interested and we can no longer think of a state as a point on the plane.

Fortunately, the cartesian correspondence between real numbers and points on the line, and between pairs of real numbers and points on the plane can be extended.

To specify a point in space, we choose an origin and three lines (or *axes*) intersecting at the origin. Then every point can be given as a triple of numbers and conversely. We do not have direct sensory experience of spaces of dimension higher than three. However, we can define 4-space to be the geometrical object corresponding to quadruples of numbers and think of each quadruple as a point in 4-space. There is no reason to stop with four: we define 5-space to be the geometrical object whose points are quintuples of real numbers, 6-space to be the geometrical space whose points are sextuples of numbers, and so on. If n is any positive integer, we define n-dimensional space to be the set of n-tuples of real numbers and denote it as $\mathbf{R}^n$.

Now, if we were trying to model an ecological system with 10 species, say, we would need 10 variables $x_1, x_2, \ldots, x_{10}$, where x_1 is the number of members of species 1, x_2 the number of members of species 2, and so on. By what we've just said we could think of the 10-tuple of numbers $(x_1, \ldots, x_{10})$ as a point in $\mathbf{R}^{10}$. We refer to a point of $\mathbf{R}^{10}$ as a *state* of the system. Of course we can't visualize 10-dimensional space directly, but the geometric intuition we bring from dimensions one to three often allows us to think geometrically about systems of equations in ten variables.

In analogy with the case of two species, we would then try to find equations giving the rates of change $x_1', \ldots, x_{10}'$ in terms of $x_1, \ldots, x_{10}$. As in the case of two variables, we could think of the 10-tuple $(x_1', \ldots, x_{10}')$ as the rate of change of the point $(x_1, \ldots, x_{10})$ and imagine it as an arrow beginning at the latter. Thus we could view the ten equations giving $x_1', \ldots, x_{10}'$ in terms of $x_1, \ldots, x_{10}$ as associating to each point $(x_1, \ldots, x_{10})$ a rate of change $(x_1', \ldots, x_{10}')$; that is, as defining a vector field on the state space $\mathbf{R}^{10}$. A trajectory (or solution) of these equations starting at a point $(x_1(0), \ldots, x_{10}(0))$ is the set of $(x_1, \ldots, x_{10})$ in $\mathbf{R}^{10}$ at subsequent times. This is a "curve" in $\mathbf{R}^{10}$.

We should point out that the number of variables need not

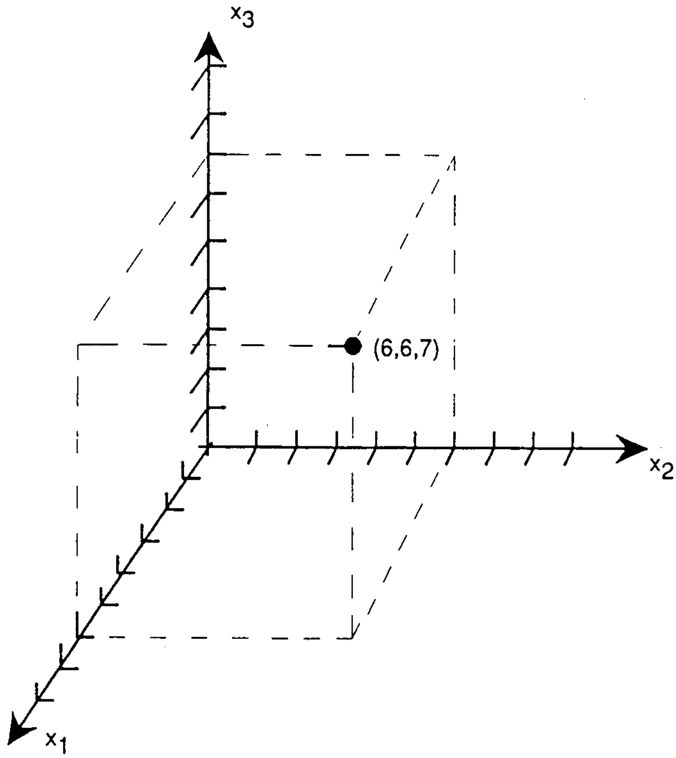

Figure 22. The correspondence between points in 3-space and triples of numbers

correspond to the number of species in an ecosystem. In more complicated models, we might want to consider a single species as composed of three (or more different) types of individuals: perhaps the very young, mature adults, and the very old. Here, three different variables would correpond to a single species, and it is not difficult to imagine that the rates of change of each of the subgroups would be different. Such models are called models with *age structure*.

Summary

We did four things in this chapter.

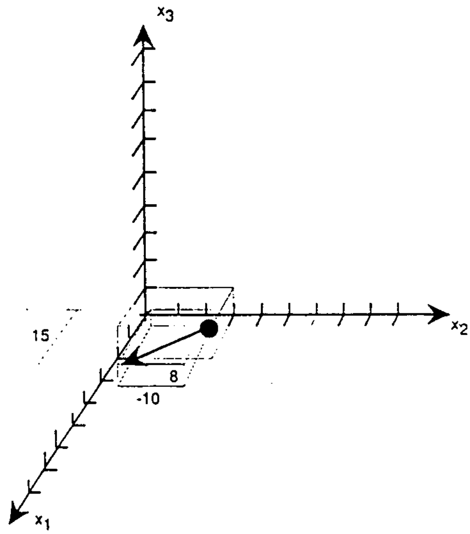

Figure 23. A point (210,315,405) with velocity (15,-10,8)

- We identified two quantities in which we were interested.
- We wrote down equations for the rates of change of these quantities.
- We used these equations for the rates of change to describe how the quantities behave as time passes.
- We introduced a geometric interpretation of pairs of quantities, and rate equations.

In all cases, the quantities x, y in which we were interested were the numbers of members of a species. Depending on different assumptions placed on the interaction between x and y, we exhibited a number of different pairs of equations for x' and y'. As in the case of one species, once we knew the value of x and y at a fixed time, we were able to use the equations for the rates of change x' and y' to figure out what x and y were at any subsequent time.

Geometrically, we think of the pair (x, y) as a point in the plane and call it a state. The two equations for x' and y' are then thought of as assigning a pair (x', y') of numbers to each point in the plane. We think of the pairs (x', y') as the rate of change of the point (x, y) and view it as an arrow beginning at (x, y). With these definitions, we can restate the principle we ennunciated in the last chapter.

- **If you know a state at some time and know the rate of change of every state (or at all times), then you can determine the state at any time.**

CHAPTER 3 – Dynamical Systems

Introduction

We have already mentioned that the process of mathematical modelling is the attempt to cast a piece of reality or situation in mathematics, with a view to using mathematical reasoning to understand the situation being modelled. In this chapter, we try to elaborate on this idea by building on some of the notions that we have seen in the previous two chapters. This chapter is considerably more mathematical than the other chapters, so that you might want to read it lightly first, referring back to it as needed.

The Notion of a State Space

The process of casting a situation into mathematics is that of identifying the salient features of the situation and attaching mathematical objects to them. When most of us think of mathematics, we think of numbers. And, indeed, when studying a situation mathematically, the first thing one does is to try to associate a number or collection of numbers with the situation which characterizes it.

When describing snakes on Guam, we used one number: the snake population. To describe two species, we used a pair of numbers, one equal to the number of individuals of one species and the second equal to the number of individuals of the second species. Likewise, to describe how far one is from New York on Interstate 80 takes one number, but to describe a single place in the United States takes two numbers, your latitude and longitude. To describe the weather at a single point and given time in the United States requires at least five numbers: the temperature, the wind speed, the wind direction (measured, say, as an angle clockwise from the North), the relative humidity (the amount of moisture in the air), and the air pressure. This still does not capture everything: we would probably want to know the rate of precipitation, whether or not there is a thunderstorm around, the rate of change of the air pressure, and so on. To describe the weather in Massachusetts at a given time, we might decide to sample these five different numbers at one thousand different points in the state. This would encode Massachusetts's weather as 5000 numbers.

In general, we prefer to think geometrically, thinking of a single number as a point in a line $\mathbf{R}$, a pair of numbers as a point in the plane $\mathbf{R}^2$, a triple of numbers as a point in 3-space $\mathbf{R}^3$, and 5000 numbers as a point in 5000-dimensional space $\mathbf{R}^{5000}$.

Confronted with a real situation, we would try to isolate enough properties to distinguish among all situations of interest. In this case, each conceivable situation corresponds to a point in $\mathbf{R}^n$ (or, perhaps, to some subset of $\mathbf{R}^n$ – it might happen that some or all of the numbers must satisfy some constraints. For example, when considering two species models , the corresponding points (x, y) in $\mathbf{R}^2$ had to satisfy $x \geq 0$ and $y \geq 0$. The set of all points in $\mathbf{R}^n$ corresponding to a conceivable situation is called the *state space.*

Here are some examples of state spaces.

1. Suppose we are interested in some ecological system consisting of three species. Let x, y, z be the populations (or, alternatively, the biomass) of each species. Then we model the "state" of the system as a point (x, y, z) in $\mathbf{R}^3$ (more precisely, as a point in the first octant of $\mathbf{R}^3$).

2. Consider a binary star system, consisting of two stars orbiting one another. If we are only interested in the motion of the stars, we can think of the stars as points and model the states of the system as follows: fix an origin and coordinate axes in space (this enables us to label all points in space). Then the position of each star can be given by specifying three numbers — so we need six to specify the positions of the two stars. Furthermore, we need to give three numbers to specify the velocity of each star: the velocity in the x-direction, the velocity in the y-direction, and the velocity in the z-direction — thus, six to specify both velocities. We don't need to specify the acceleration of each, because this is given by Newton's

law of gravitation once we know the mass of each star. Thus, the state of the system can be modeled as a point in $\mathbf{R}^{12}$.

3. Consider a model of an economy in which we are interested in the numbers of different goods produced and the numberconsumed. Then we model the state of the system by associating one variable to each type of good. A state of the economy would then be a point in $\mathbf{R}^N$ where N is the number of different goods.

Time and Change

The most enduring feature of our world and the situations that we encounter is change. Any attempt, therefore, to understand some aspect of reality entails understanding the changes that take place over time. Indeed, the ultimate goal of much mathematical modelling is to predict change and, where possible, to suggest strategies that allow one to direct that change.

From the point of view of mathematical modelling, taking account of time means that the numbers with which we characterize a situation will be allowed to change as time changes. For example, if we are driving along Interstate 80, we take account of the change in position by giving our distance west of New York at each time. Phrased somewhat differently, we view the number that characterizes our position as changing. At each time it has a definite value; that is, it is a function of time.

To be a little more formal, recall that a *function* is a mathematical object which assigns, to each element of one set an element of another. To take account of changes over time, we think of the states as functions of time. In the case of the auto on I-80, the actual distance west of New York on I-80 as a function of time. This is just a way of saying that at each time our auto is some particular distance, measured along I-80, from New York.

We remark that this usage of the word "function" is very different from one of its uses in ordinary language, where it often connotes an unspecified dependence relation or a hazy causal link between two or more quantities. For example, one hears that "the number of housing sales is a function of the interest rate" or that one's "health is a function of diet" meaning that the interest rate affects the number of housing sales or that one's diet influences one's health. We shall *never* use the word "function" in this sense. From our point of view, these phrases mean something else entirely. The first would mean that to each interest rate, there corresponds a definite number of housing sales (perhaps a 7.1% rate corresponds to 10317 housing sales) – this is certainly not the case: the number of housing sales depends on a great many other factors besides interest rate, so there is not a definite number of them which we can associate with a given interest rate. (We might, however, construct a *model* in which we assume, for simplicity, that the number of housing starts is a function of the interest rate.) The second phrase "health is a function of diet" is meaningless unless we first specify what we mean by the set of diets and what we mean by the set of healths. One way of doing this is to construct a mathematical model in which a given set of mathematical objects corresponds to the set of diets and a given set of objects to healths. In summary, with our specialized meaning of the woed "function", neither the phrase "the number of housing sales is a function of the interest rate" nor the phrase "health is a function of diet" has meaning outside the context of a mathematical model.

Functions

Since we model changes in time by functions and since you will probably have encountered the notion of a function in high school, it is worth making a few additional remarks about them.

The type of functions most familiar to you will be the functions which associate to each real number another real number by means of a formula. (We often say that a function which associates to each real number another real number is a function from R to R). For example, we might have

$$
\begin{aligned}
f(t) &= 3t - 2 \\
g(t) &= t^2 + 5t - 2 \\
h(t) &= \frac{t^3 - 1}{1 + t^2}
\end{aligned}
$$

There is no doubt that a formula is a convenient way of giving a function, but we don't always have them available. Observing the temperature over some period of time at some particular spot in the outdoors , defines temperature (at that spot) as a function of time, but there will clearly be no formula. To figure out the value of this function, the time had better be past and we had better have a recording of the temperature. This type of function is what's known as a *data function*. Functions can also be given by recipes which do not easily lend themselves to formulas. We shall see later that, specifying rates of change of a quantity at every time together with a value at some given time uniquely determines that quantity as a function of time. This will be the most common way in which we will encounter functions.

One other difference betweeen the type of functions we deal with and the functions usually encountered in high school algebra classes, is that we will usually deal with functions that associate with each time a state: that is, which take the set $\mathbf{R}$ to the set $\mathbf{R}^n, n \geq 1$. Thus, to each time, such a function associates n numbers. An example of a function from $\mathbf{R}$ to $\mathbf{R}^3$ is

$$\mathbf{f}(t) = (t^2, t - 3, 1 - t^2).$$

Note that this function can be thought of as a collection of three functions from $\mathbf{R}$ to $\mathbf{R}$, namely $f_1(t) = t^2$, $f_2(t) = t - 3$ and $f_3(t) = 1 - t^2$. In the same way, a function from $\mathbf{R}$ to $\mathbf{R}^n$ can be thought of as a collection of n functions from $\mathbf{R}$ to $\mathbf{R}$. There is nothing to prevent some of these functions being data functions and others being given by formulas or other recipes. A little later, we shall talk about functions which associate an arrow to each state – we shall see that these can be thought of as functions from $\mathbf{R}^n$ to $\mathbf{R}^n$.

Notation

One of the reasons for introducing geometric concepts such as state space and points in state spaces (as opposed to collections of numbers) is that they provide a way of visualizing systems and, hence, a mental shorthand which allows us to think about complex models. A complement to the mental shorthand is provided by good notation. We pause here to lay out our notational conventions: we will remind you about them frequently and you should review them until they are second nature to you.

We have seen that a point in $\mathbf{R}^n$ corresponds to n numbers. If we don't want to specify the numbers, we write $(x_1, x_2, \ldots, x_n)$ (or, sometimes, (x, y), (x, y, z), or (w, x, y, z) if n is 2, 3, or 4, respectively). When we are thinking of the n numbers as a point in $\mathbf{R}^n$, we use a single boldface letter

$$\mathbf{x} = (x_1, x_, \ldots, x_n) \in \mathbf{R}^n$$

to denote it. We name functions with letters. If we want to talk about a function f, say, that associates to each point of a set A a point of B, we use the notation $f : A \to B$. If B is $\mathbf{R}^n$ we use a boldface letter to denote the function. Thus we might write

$$\mathbf{a} : \mathbf{R} \to \mathbf{R}^n.$$

The value of the function at some particular time t is denoted by $\mathbf{a}(t)$. If we want to emphasize that fact that $\mathbf{a}(t)$ consists of n numbers, we will write $\mathbf{a}(t) = (a_1(t), \ldots, a_n(t))$, (so $\mathbf{a} : \mathbf{R} \to \mathbf{R}^n$ consists of the n functions $a_1 : \mathbf{R} \to \mathbf{R}$, $a_2 : \mathbf{R} \to \mathbf{R}$, …, $a_n : \mathbf{R} \to \mathbf{R}$.)

Curves in Phase Space

To take account of the the change in weather in Western Massachusetts, each of the 5000 quantities which we think of as describing the weather must be allowed to vary. Rather than thinking of the behavior of the weather over time as 5000 functions of time, we think of it as a single function of time which associates to each time a point in $\mathbf{R}^{5000}$. A point in $\mathbf{R}^{5000}$ is a "snapshot" of the weather. As time changes, the "snapshot" moves – at each time the function specifies a definite point in $\mathbf{R}^{5000}$ corresponding to the weather at that time. As the point moves, it traces out a curve in $\mathbf{R}^{5000}$, which

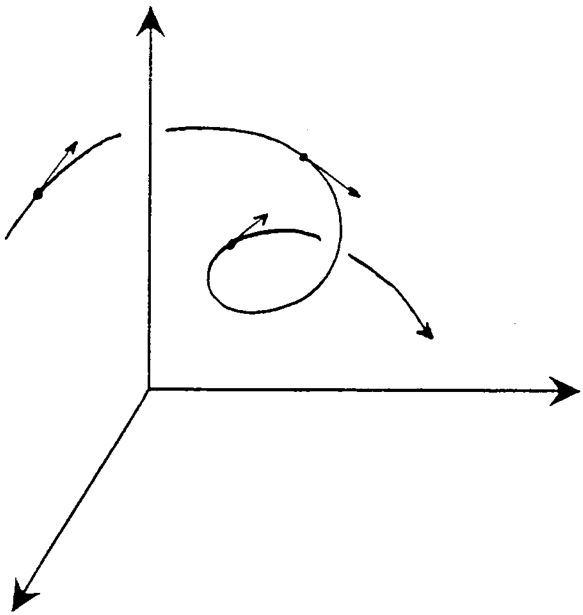

Figure 1 The motion of a State in $\mathbf{R}^3$ traces out a curve

represents the change in the weather over time. In this way, our snapshot becomes a "movie".

In modelling a situation, we begin by setting up a dictionary which associates to every conceivable real situation a point in $\mathbf{R}^n$ where n is the number of quantities we need to characterize the situation. The actual situation at a given time is a definite point in $\mathbf{R}^n$. As time changes, we get different points, corresponding to the changed situations. This defines a function from the set of times $\mathbf{R}$ to the set of states $\mathbf{R}^n$. As time increases, we get a curve in $\mathbf{R}^n$ by plotting the different points. We think of the state moving along the curve.

Rate Equations and Dynamical Systems

In chapters 1 and 2 we saw repeated examples in which knowing the rates of change of points on the line or plane allowed one to determine what happens to a state at any future time. If you know the state at which you start and the rate of change of that state, then you can figure out the state a short time later. But, by assumption you know the rate of change of this latter state, so you can figure out the state a short time later still. Continuing in this manner allows you to figure out the state at any future time.

This observation leads directly to the general notion of a dynamical system. Namely, a *dynamical system* is a set of equations expressing the rate of change of a state in terms of the state and time. Symbolically, if $(x_1, \ldots, x_n)$ is the state, then a dynamical system is a set of equations of the form

$$\begin{aligned} x_1' &= f_1(x_1, \ldots, x_n, t) \\ &\vdots \\ x_n' &= f_n(x_1, \ldots, x_n, t). \end{aligned}$$

More compactly, a continuous dynamical system is an equation of the form

$$\mathbf{x}' = \mathbf{F}(\mathbf{x}, t)$$

where $\mathbf{F} : \mathrm{R}^{n+1} \to \mathrm{R}^n$. In words, a dynamical system is a set of equations which express the rates of change of a set of variables in terms of the *same* set of variables and, possibly, time.

The word "system" in the phrase "dynamical system" refers to the system of equations explicitly giving the rates of change of all the state variables. The whole point of the definition is that if we know the values of the variables characterizing a state at some time, then we can compute the rates of change of these variables. Once we can compute these rates of change we can compute the values of the variables at subsequent times.

The following sets of equations are examples of dynamical systems.

a. $\begin{aligned} x' &= 3x + y \\ y' &= 2y \end{aligned}$ b. $\begin{aligned} x' &= 3xy \\ y' &= x^2 - t \end{aligned}$ c. $\begin{aligned} w' &= 3t \\ x' &= 3x - y \\ y' &= xywt. \end{aligned}$

The following sets of equations are not dynamical systems.

$$\text{d.}\quad \begin{aligned} x' &= 3x + y \\ y' &= 2yzt \end{aligned} \qquad \text{e.}\quad \begin{aligned} x' &= 3x' - y \\ y' &= x' - t \end{aligned} \qquad \text{f.}\quad \begin{aligned} w' &= 3zt \\ x' &= 3x - y \\ y' &= xywt. \end{aligned}$$

In the sets d and f we need extra information to compute the rates of change. In d, we can't compute x' and y' without knowing z. If we gave z as a function of x, y, t or added an equation giving z' as a function of x, y, z, t, then we would get a dynamical system. Likewise, in f we don't have an expression for z or z'. In e, we would have to solve for x'. While it is not difficult to do this, we reserve the phrase "dynamical system" for systems in which the rates of change are given *explicitly* in terms of the underlying quantities.

Vector Fields and Autonomous Dynamical Systems

A special role is played by dynamical systems, such as a above, which do not explicitly involve time. Such systems are called *autonomous*. They have the form

$$\mathbf{x}' = \mathbf{F}(\mathbf{x})$$

where $\mathrm{F} : \mathrm{R}^n \to \mathrm{R}^n$. In terms of components

$$\begin{aligned} x_1' &= f_1(x_1, \ldots, x_n) \\ &\vdots \\ x_n' &= f_n(x_1, \ldots, x_n). \end{aligned}$$

As we indicated in the last chapter, there is a nice geometrical way to picture these systems. We imagine the system associating to each point $\mathbf{x}$ the arrow $\mathrm{F}(\mathbf{x})$. So, for example, the system a above associates to each point (x, y) of the plane, the arrow $(x', y') = (3x+y, 2y)$. So, to the point (1,3) it associates the arrow $(3\cdot 1+3, 2\cdot 3)$ = (6, 6). At the point (1, 1), we have the arrow (4,2); at the point (-1,0) the arrow (-3,0); at the point (0,-1), the arrow (-1, -2). We

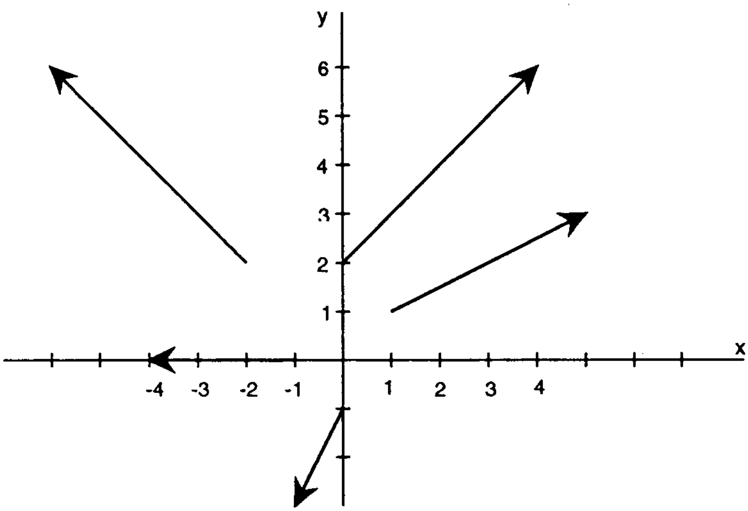

Figure 2. The vector field $(x', y') = (3x + y, 2y)$ at selected points

have sketched the arrows at these points in Figure 2. Every point has an arrow attached to it which represents the rate of change at that point. (Of course, we can't draw infinitely many arrows any more than we can list all the points in $\mathbf{R}^2$). We can think of all the points as moving with the velocities given by the arrows.

A set of arrows attached to every point of $\mathbf{R}^n$ is often called a *vector field*. Thus, an autonomous dynamical system is a vector field. If we are at some point of $\mathbf{R}^n$, then the point moves along a curve in such a way that its rate of change is the arrow assigned to it by the dynamical system. These arrows are necessarily tangent to the curve along which the point moves. Curves that satisfy the property that they are tangent to the arrows of a dynamical system are called *integral curves* or *trajectories* of the dynamical system. Given an autonomous differential equation, the geometric way of describing our method of finding what happens to a point is to start at the point, follow the arrow at that point for a short period of time, stop, follow the arrow at the new point for a short period of time,

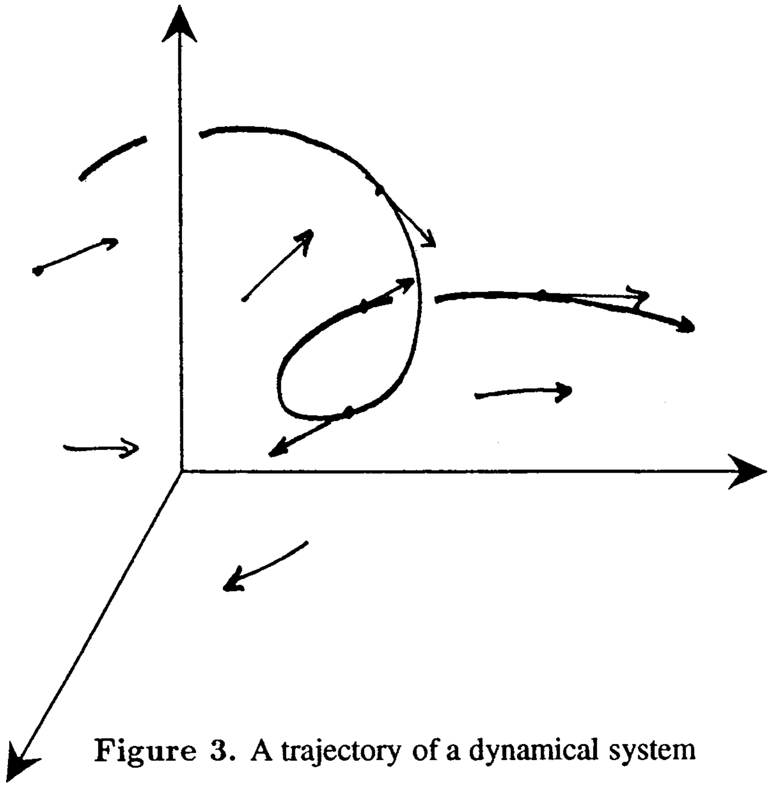

Figure 3. A trajectory of a dynamical system

stop, and so on.

Nonautonomous Dynamical Systems

Finally, any nonautonomous dynamical system can be considered to be an autonomous dynamical system by introducing another variable. For example, consider the nonautonomous system b above:

$$x' = 3xy$$
$$y' = x - t.$$

Simply introduce a new variable z and write $z = t$. The rate of change z' of z with respect to t is just one unit per unit time. (Since $z = t$, in a single unit of time, z changes by one unit.) So we can rewrite the system as the autonomous system

$$
\begin{aligned}
x' &= 3xy \\
y' &= x - z \\
z' &= 1
\end{aligned}
$$

in which we have treated time as an extra variable. Thus, we have two different ways of viewing a nonautonomous system: as a vector field on a space with one larger dimension or as a vector field on the state space in which the arrows change with time.

The Process of Modelling

With the discussion above in mind, and the examples in the last two chapters in hand, we can now be much clearer about what mathematical modelling is. For us, the construction of a mathematical model consists of the following.

1) Setting up a correspondence between possible situations and points in $\mathbf{R}^n$. In particular, this means identifying a set of n properties to which numerical values can be assigned and which uniquely characterize the situation. In this case $\mathbf{R}^n$ (or some subset of it) becomes the state space.

2) Finding equations which express the rates of change $x_1', \ldots,$ x_n' in terms of the variables $x_1, \ldots, x_n$ identified in step 1) and, possibly, time. In other words, finding a dynamical system which expresses the rates of change of the relevant variables.

Informally, the dynamical system is often referred to as the model. We will avoid this usage, because we wish to emphasize that step 1) is an essential part of any model. It is also worth pointing out that the above is a rather narrow view of what mathematical modelling is – it refers only to modelling by dynamical systems. Dynamical systems are not the only, and in some cases not the most desirable, way to model phenomena.

We hasten to make two comments. Any model is necessarily a caricature in the sense that no mathematical model can be a complete description. Just as a historian seeking understanding of, say, a certain event in World War I sifts and selects facts, suppressing some and according others great significance, so, too, an individual modelling a situation selects certain phenomena and characteristics of those phenomena and ignores others in search of insight into the situation under study. Modelling is a dynamic process – greater insight frequently leads to new concepts which, in turn, lead to a reconsideration of what facts and phenomena are salient. Indeed, commonly accepted understandings of historical events or physical or biological phenomena can vary dramatically from one time to another.

The second comment is that the identification of a model with a dynamical system is common, and is certainly the case we are concentrating on in this monograph. More generally, however, a mathematical model of a situation or context consists of a correspondence between certain aspects of the situation and a class of mathematical objects, together with a rule which expresses how these mathematical objects change. The mathematical objects one uses need not be numbers or collections of numbers – they might be functions, or geometric objects such as shapes, or algebraic objects, or even logical propositions.

The tools of a historian, or a social scientist, are his or her language, its words or grammar, and the disciplinary concepts forged out of that language. For the individual involved in mathematical modelling, the tools are those of mathematics, the basic mathematical objects, their grammar (that is, the rules for operating on them), as well as more specialized mathematical concepts. The historian needs to check carefully that the language he or she uses is appropriate to the situation being studied. So, too, the mathematical modeller needs to check carefully that the mathematical concepts he or she uses provide a good fit with the reality being studied. It is important to make sure that not too much violence is done in trying to graft a dynamical system onto reality. Some aspects of real situations just are not quantifiable in any reasonable sense.

Equilibrium points, limit cycles, and attractors

In the two-species models we presented above, the state could be given by specifying two numbers. This meant that we were dealing with vector fields on the plane $\mathbf{R}^2$. We have already pointed out that these models are very simplified – they do not take into account the age structure of the population, for instance. Nevertheless, we saw that even such simple models gave us useful insights. They also give us insight into what to expect in much more general dynamical systems.

In the models we used in the last chapter, we were usually led ineluctably to ask what happens to a trajectory over a long time. Typically trajectories do not go off further and further out on the plane (i.e. they do not extend to infinity). A trajectory which does not stay in some bounded region corresponds to the value of at least one of the numbers characterizing the state going to infinity, and this usually does not happen in real systems if the number which tends to infinity refers to position, or number of individuals, or mass, or some other physical quantity.

If a trajectory does not "go off to infinity", it is reasonable to ask where it goes. In the plane, there is a complete answer. To explain it, we will need to introduce some terminology. The first singles out those points to which trajectories can tend.

Definition. A point $\mathbf{x}_0$ of the plane $\mathbf{R}^2$ (more generally, of $\mathbf{R}^n$) will be called an *equilibrium point* or *rest point* of a dynamical system $\mathbf{F} : \mathbf{R}^2 \longrightarrow \mathbf{R}^2$ (more generally, of $\mathbf{F} : \mathbf{R}^n \longrightarrow \mathbf{R}^n$) if $\mathbf{F}(\mathbf{x}_0) = \mathbf{0}$. That is, if the arrow the dynamical system assigns to the point $\mathbf{x}_0$ is the zero vector.

More generally we say that a subset of $\mathbf{R}^n$ is a *limit set* if every point in it occurs as the limit of some set of trajectories, either as $t \to \infty$ or as $-t \to \infty$. So equilibrium points are examples of limit sets. A second type of limit set is the so-called limit cycle, which we encountered in Robert May's predator-prey model in Chapter 2.

Definition. A trajectory of a dynamical system is *closed* if any point of the trajectory comes back to itself after finite time under the motion of the dynamical system. Closed trajectories which are the limit of some (nonempty) set of trajectories, either as $t \to \infty$ or as $-t \to \infty$, are called *limit cycles.*

Closed trajectories will lie in some bounded region of the plane (or, more generally, $\mathbf{R}^n$) and contain no rest points. Closed trajectories on the the plane, divide the plane into two pieces: an inside and an outside.

A set is said to be *compact* if it is bounded and if it contains any point which is a limit of points in the set (that is, if the set contains all its boundary points). Perhaps the most important theorem regarding dynamical systems in the plane is the Poincaré-Bendixson theorem.

Theorem (Poincaré-Bendixson). *Any non-empty compact limit set of a dynamical system* $\mathbf{x}' = \mathbf{F}(\mathbf{x})$, $\mathbf{F} : \mathbf{R}^2 \longrightarrow \mathbf{R}^2$ *on the plane, which contains no equilibrium point, is a closed orbit.*

The Poincaré-Bendixson theorem is highly non-trivial. For example, if we have a dynamical system on the plane and we find a region (with finite area) whose boundary has the property that the arrows at each point of the boundary point into the region, then it easily follows from the Poincaré-Bendixson theorem that the region contains either an equilibrium point or a limit cycle.

Another consequence of the Poincaré-Bendixson theorem is that in the long run, states of dynamical systems in the plane either settle down to a point (in which case, both numbers characterizing the state tend to fixed values) or to a steady cyclical behavior (in which the values of the quantities oscillate periodically about some fixed values.)

There is another observation that we can make based on the systems we have already examined. Namely, the limit sets of most interest are those that are such that trajectories through all nearby points tend to the limit set as $t \to \infty$. Such limit sets are called *attractors.* (Limit sets with the property that all nearby points tend away from the set as $t \to \infty$ are called *repellors*). For example, in the Lotka-Volterra system described by equation (12), there was is an equilibrium point (0,25), but it is not of physical interest. The slightest change in either x or y and the system sweeps off to one of the other equilibria points. If we didn't have the equations in front of us, we wouldn't even suspect that there was a rest point – it is unobservable for all practical purposes. A second observation is that a point in phase space moves fairly quickly towards an attractor. The moral here is that the part of a dynamical system which we are most likely to observe will reflect the parts of the system close to attractors. This is one of the departure points for René Thom's theory of models and is a key insight for the understanding of "catastrophe theory".

Appendix to Chapter 3

The following material is, strictly speaking, an aside. However, the reader may find it useful when reading other books on modelling.

Discrete Dynamical Systems

There is another type of dynamical system, intimately related to the ones we have defined above. To define them, we suppose that we measure time in *discrete* units. These time units might be hours, seconds, tenths of a second, years, centuries, whatever, but once fixed, we only allow ourselves to consider integer multiples of them. As before, we represent the state of a system as a point in $\mathbf{x} \in \mathbf{R}^n$. We denote the state of the system at time $t = 0$ by $\mathbf{x}[0]$ and the state of the system at the k^{th} time step as $\mathbf{x}[k]$. We use square brackets to remind ourselves that we are talking about numbers of time intervals and use k to denote time instead of t to remind ourselves that we are only allowing integer values of time. If, for example, the units of time were years, the value of $\mathbf{x}$ one decade after the inital time would be denoted by $\mathbf{x}[10]$ and the value two centuries previous to the inital time by $\mathbf{x}[-200]$. It would not make any sense to speak of the value of $\mathbf{x}$ after a half a year: that is, we do not allow ourselves to write $\mathbf{x}[.5]$.

A *(nonautonomous) discrete dynamical system* is a system in which the state at any time depends *only* on the state at the time before. That is, it is a system which can be written in the form:

$$\mathbf{x}[k+1] = \mathbf{F}_k(\mathbf{x}[k])$$

where $\mathbf{F}_k$ is a function from $\mathbf{R}^n$ to $\mathbf{R}^n$ for each $k \in \mathbf{Z}$. An *autonomous discrete dynamical system* is an equation of the form:

$$\mathbf{x}[k+1] = \mathbf{F}(\mathbf{x}[k])$$

where $\mathbf{F}$ is a function from $\mathbf{R}^n$ to $\mathbf{R}^n$.

Note that there is no talk of rates of change in a discrete dynamical system: we get from the state at one time unit to that of the next by some rule applied to the former. Any function $\mathbf{F} : \mathbf{R}^n \longrightarrow \mathbf{R}^n$ defines a (autonomous) discrete dynamical system. You merely set $\mathbf{x}[k+1] = \mathbf{F}(\mathbf{x}[k])$. That is, you take any point and repeatedly apply the function.

Note that a discrete dynamical system, like a continuous dynamcial system, is completely deterministic in the sense that if you know the initial state, you can know what that state will be at any time thereafter.

Discrete dynamical systems arise from continuous dynamical systems by assuming that rates of change are constant over some small time interval. The moment we choose a Δt, we convert a continous dynamical system

$$\mathbf{x}' = \mathbf{F}(\mathbf{x})$$

into the discrete dynamical system

$$\mathbf{x}[k+1] = \mathbf{x}[k] + (\Delta t)\mathbf{F}(\mathbf{x}[k])$$

with time unit equal to Δt. Notice that the latter equation is nothing but another way to write equation (1):

$$\mathbf{x}_{\text{new}} = \mathbf{x}_{\text{old}} + (\Delta t)\mathbf{x}'_{\text{old}} \tag{1}$$

For more on the relation between continuous and discrete dynamical systems see Hirsch and Smale's book. For information on discrete dynamical systems in biology, see the books of May, Maynard Smith and Murray cited in the bibliography.

Exercise 1. Consider the function $\mathbf{F} : \mathbf{R}^2 \longrightarrow \mathbf{R}^2$ given by $\mathbf{F}(x_1, x_2) = (.5x_1 - .25x_2, -.5x_1)$. If $\mathbf{x}[0] = (4, 6)$, find $\mathbf{x}[3]$ and $\mathbf{x}[4]$. Write a computer program to find $\mathbf{x}[10]$ and $\mathbf{x}[100]$. What do you think the limit of the $\mathbf{x}[k]$ are as k gets bigger an bigger? Does this limit depend on the choice of $\mathbf{x}[0]$? values of $\mathbf{x}[0]$

Exercise 2. Let $\mathbf{F} : \mathbf{R}^2 \longrightarrow \mathbf{R}^2$ be given by $\mathbf{F}(x_1, x_2) = (.5x_1x_2, -.7x_1 + x_2)$ and consider the (continuous) dynamical system $\mathbf{x}' = \mathbf{F}(\mathbf{x})$. If $\Delta t = .1$, explicitly write out the corresponding discrete dynamical system.

Dynamical Systems and Calculus

Historically, the development of dynamical systems was inextricably linked to calculus — we'll say more about this in Chapter 4. Before the development of the computer, to determine a future state given a dynamical system and an initial condition, one either had to do all the computations by hand (which, while feasible in theory, was not usually a practical option – it would take too long) or use calculus, which provided closed form answers in certain simple cases, and which could be used to help approximate future states in more complicated systems.

Even if you choose not to study calculus, you should be aware of some of the commonly used terms relating to dynamical systems which are inspired by calculus. First, the rate of change $\mathbf{x}'$ of a state is often referred to as the *derivative* of $\mathbf{x}$ with respect to time (and often denoted as $\dot{\mathbf{x}}$ or as $\frac{d\mathbf{x}}{dt}$). What we are calling a continuous dynamical system is often called a *differentiable* dynamical system.

If we are given an initial state and want to find the state at some time in the future, we say that we are *integrating* or *solving* the dynamical system. The process of integrating a system on a computer is often called *numerical integration*. The very simple process we used in the examples we chose in this chapter, namely that of choosing a Δt and replacing the continuous system $\mathbf{x}' = \mathbf{F}(\mathbf{x})$ by the discrete dynamical system $\mathbf{x}[k+1] = \mathbf{x}[k] + (\Delta t)\mathbf{F}(\mathbf{x}[k])$ is called *Euler's method*. (Euler's method could also be described as the process of integrating the system $\mathbf{x}' = \mathbf{F}(\mathbf{x})$ by iterating relation (1)). The process of passing from a continuous to a discrete system by choosing a fixed time interval Δt is called *discretizing* the system.

Chapter 4 — The Greatest Mathematical Model Ever

Introduction

Most of the examples of dynamical systems we have seen up to now have been drawn from contexts in the life sciences. We have done this partly because the variables in the systems we examined are quite familiar to us — we read in the papers every day about the population of some group or another. Our presentation has, however, reversed the order in which dynamical systems were discovered and used.

The purpose of this section is to present some of the historical background which underpins our current understanding of dynamical systems and the fundamental role they play in science. The process of constructing a set of differential equations to describe some aspect of reality is so embedded in modern consciousness that it is easy to forget that the equations are not absolute truths – merely constructs which help us understand certain phenomena. Moreover, dynamical systems have been so successful in modelling physical phenomena that it is sometimes easy to forget that they are not the only way to model phenomena outside the physical sciences. How did dynamical systems come to be the pre-eminent tool that they are today?

Kepler

Our story begins on what loosely might be described as the dawn of modern science: January 1, 1660. On that day, Johannes Kepler hitched a ride from the provincial capital of Linz to the imperial capital of Prague. Kepler had already distinguished himself by writing a much admired treatise *Mysterium Cosmographicum* detailing why the planets were where they were — his theory was based on the way the five Platonic solids (the tetrahedron, the cube, the octahedron, the icosahedron, and the dodecahedron) could be inscribed in one another. Kepler was one of the few individuals who accepted the Copernican theory that the planets revolved around the sun – he did so on purely theological grounds. To improve his theory, Kepler needed access to astronomical data which would allow him to compute the positions of the six planets then known. The best available data (indeed the only reliable data) were in the hands of the greatest astronomer of the era, Tycho de Brahe, a colorful Dane of noble lineage, who had recently been named to the position of imperial mathematicus (at a fantastic salary) and set up by the emperor, Rudolph II, in a huge observatory near Prague. Brahe had been the first individual to make systematic, daily observations of the positions of the planets. He did this over a period of twenty years from his former observatory, Uraniburg, on an island off the Danish coast.

Kepler had accepted a position as Brahe's assistant, hoping to see Brahe's data and use it to further his own theory. Brahe, on the other hand, hoped to enlist Kepler's formidable mathematical skills to advance his own theory that the sun revolved around the earth and that the other planets revolved around the sun. In February the two men finally met. Here is how Koestler describes the meeting in his biography of Kepler entitled *The Sleepwalkers.*

> At last, then, on February 4, 1600, Tycho de Brahe and Johannes Keplerus, cofounders of a new universe, met face to face, silver nose to scabby cheek. Tycho was fifty-three, Kepler twenty-nine. Tycho was an aristocrat, Kepler a plebeian; Tycho a Croesus, Kepler a church mouse; Tycho a Great Dane, Kepler a mangy mongrel. There were opposites in every respect but one: the irritable, choleric disposition which they shared.

What follows is one of the more incredible episodes in the history of human thought. Kepler was put to work on computing the details of the orbit of Mars, an orbit which had stubbornly resisted analysis. Kepler had to wait two years until Brahe's death to gain full access to the data from the preceding twenty years. After eight years of Herculean labor, in which Kepler painstakingly threw out one long-accepted idea after another, Kepler discovered the first two of his famous three laws. The first was that the orbits of the planets were elliptical, not circular (it took him four years to abandon the idea that the sun was at the center of the orbit, another two to realize that the orbits were not circular, and another two to identify the shape as that of an ellipse – he had tried fitting all sorts of ovals.) The second was just as radical – the speed of a planet in its orbit was not constant, but varied, moving faster when it was nearer the sun than when it was further away. Kepler's third law was published eight years later still and gave a precise relationship between the planet's average distance from the sun, and the length of time it took to make one complete orbit around the sun.

Kepler's laws were the first instances of precise quantitative statements regarding the planets which allowed predictions. The three laws can be regarded as a mathematical model, although not a dynamical system. They allow us to make many useful inferences regarding planetary and lunar motion. We add that Kepler's three laws are not labelled and are buried in his writings admidst an incredibly lush array of metaphysical and theological musings. It is a testimony to Newton's genius that he was able to immediately identify them as the central verifiable statements in Kepler's work.

Exercise 1. An *ellipse* is defined to be the set of all points the sum of whose distances from two fixed points (called *foci*) is a constant. The semi-major axis, usually denoted a is the longest distance between a point on an ellipse and the center and the semi-minor axis b is the shortest distance. The size and shape of an ellipse are determined by specifying the values of any two of the following quantities (see Figure 1).

a : the semi major axis

b : the semi minor axis

c : the distance of the center from one focus

e : the eccentricity (which is defined as c/a)

r_p : the closest distance from a focus to the ellipse

r_a : the farthest distance from a focus to the ellipse.

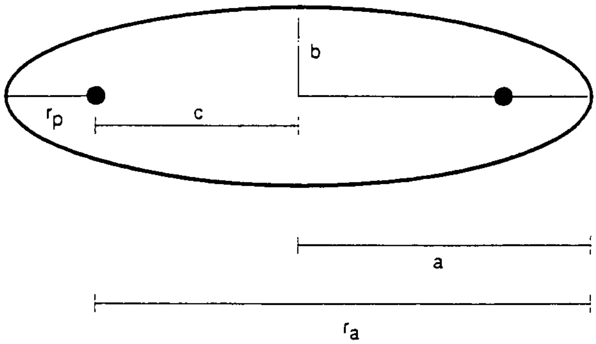

Figure 1. An ellipse.

Show that this is the case by establishing the following relationships among these quantities:

$$a^2 = b^2 + c^2$$
$$r_p = a - c = a(1 - e)$$
$$r_a = a + c = a(1 + e)$$

Exercise 2. Kepler's Third Law states that the ratio of the squares of the periods of any two planets is equal to the ratio of the cubes of the semimajor axes of their respective orbits. This law is also valid for a circular orbit ($e = 1, r_p = r_a = a$). Use this law, and the fact that the moon rotates about the earth once every 27 days, to find the distance from the earth's center, as a fraction of the distance to the moon, that a satellite must be to rotate in such a way that it remains in fixed position with respect to a point P on the earth's surface (such orbits are called *geostationary*). [Hint: to stay in a fixed position over the Earth's surface, a satellite must be in a circular orbit in the plane of the Earth's equator and its period of revolution must be equal to that of the Earth on its axis.] The most useful communications satellites are in geostationary orbits – why do you suppose this is? If the mean radius of the moon's orbit is 239,000 miles, what is the radius of a geostationary orbit?

Galileo

Our story turns now to Galileo, a contemporary of Kepler. The details of Galileo's life are probably familiar enough to you not to bear recounting. Galileo was the first person to turn the newly discovered telescope on the planets. He was the first to see the moons of Jupiter and a firm believer in the Copernican theory. Galileo's most enduring contribution to scientific thought, however, was his theory of motion. Galileo's observation of the heavens allowed him to formulate a fundamental principle which is now known as *Galileo's Principle of Inertia.* It says that *an object moving in a straight line with a given speed will continue to move in that straight line with that same speed for all time, unless acted on by an external force.*

At the time, this principle was highly non-obvious. Our experience with motion on earth would tend to lead us to believe that if we set an object in motion in a given direction with a given speed, say by pushing it, it will tend to slow down (unless we push it downhill or drop it, in which case it will speed up). And, indeed, in those days the word inertia referred to an object's tendency to resist motion. Nowadays, every schoolchild learns that it is friction that slows objects down. We are familiar with the notion that an object thrown in space will continue with its same speed in straight line, unless under influence from the sun or a planet. We learn these things early enough so that Galileo's principle seems intuitively obvious to us.

Galileo's principle of inertia has a very important corollary. It says that you don't have to explain why something, such as a planet, is moving. What you have to explain instead is why it is *not* moving in a straight line and why it is *not* moving with constant speed. In other words, what you have to explain is change in velocity, not change in position. Here, it is worth emphasizing that we are again using the word *velocity* to refer to both the speed and direction in which a body moves. Kepler had made an attempt at explaining the motion of the planets (that is, of deriving his three laws from more fundamental principles). However, he thought he had to explain why the planets didn't slow down and so he was looking in the wrong direction for an explanation. (His explanation of planetary motion is interesting. He thought that the planets were always trying to slow down: what kept them in motion was that the sun turned and dragged the planets along with it – the effect of the sun diminished with distance, he thought, so that the further a planet was from the sun, the slower it would go. He visualized the sun as a circular broom with long flexible radial spokes attached that swept the planets around with the sun and which bent backwards the further out from the sun they extended.)

Newton

To Newton is attributed the statement following.

> If I have seen a little farther than others, it is because I have stood on the shoulders of giants.

And towards the end of his long life, Newton offers us the following assessment of himself.

> I do not know what I may appear to the world; but to myself I seem to have only been like a boy playing on the seashore, and diverting myself in now and then finding a smoother pebble or a prettier shell than ordinary, whilst the great ocean of truth lay all undiscovered before me.

Two of the giants on whose shoulders Newton stood were Galileo and Kepler and two of the pebbles which he seized were Kepler's three laws and Galileo's principle of inertia. Using these and his own development of the calculus, Newton changed utterly the way in which we view our world and laid the groundwork for the technological advances and modes of thought which characterize our era.

Newton realized that Galileo's principle of inertia meant that what had to be explained was a change in the direction or speed of an object. So any time an object changed velocity, something had to cause that change. Newton called the cause of this change the *force* on the object. The simplest possible assumption you could make is that the force is *equal* to the rate of change in velocity. But experience tells us that the more massive an object, the more difficult it is to change its speed or direction. (If you are standing midway down a hill, you will find it more difficult to stop a fully loaded tractor trailer that is rolling towards you than a marble.) Thus, the next simplest assumption you could make is that the force is the mass times the rate of change of the velocity. Newton took this as his definition of force. We write this symbolically as $\mathbf{F} = m\mathbf{v}'$, where $\mathbf{F}$ denotes the force, m the mass which is assumed constant, $\mathbf{v}$ the velocity, and $\mathbf{v}'$ the rate of change of the velocity. The rate of change of the velocity is often called the *acceleration*. Since we visualize the rate of change of a collection of numbers (and the velocity $\mathbf{v}$ is a collection of three numbers: the velocities in the x, y and z directions) as an arrow, the acceleration can be thought of as an arrow and, hence, so can the force.

Newton's next insight was that, in many cases, the force could be calculated independently of this equation. Concerning the planets, Newton realized that something must be exerting a force on them because they changed direction and, hence, velocity. Like Kepler, Newton felt that the obvious candidate was the sun. Unlike Kepler, Newton had Galileo's principle of inertia and so he could assume that if the sun weren't there, the planets would just go on moving in a single direction with a constant speed. (Kepler probably would have agreed that they would go in a straight line, but would have thought that they would have ground to a halt). For Newton the simplest reasonable assumption to make was that the sun exerts a force directly towards itself. Since the force is proportional to the rate of change of the velocity, this meant that Newton assumed that an object at rest above the sun would move directly towards the sun with ever increasing velocity Think of what would happen if you shot something directly toward the sun — our experience with objects falling on earth suggests that this is reasonable. Newton went on to further assume that the magnitude of the force exerted by the sun should be proportional to the mass of the planet on which it was acting and that it should get weaker the further out from the sun the planet went. It was not clear how much weaker it should get – should it be proportional to one over the distance of the object from the sun, or one over the distance squared, or one over the distance cubed, or perhaps some higher power? Here there was no reasonable guide, but Newton had done experiments on light and was aware that the illumination was proportional to one over the square of the distance from the source. It seemed reasonable to assume that the same was true of the force exerted by the sun. Moreover, Newton proved that to be consistent with Kepler's third law this *had* to be the case.

To sum up, Newton made two assumptions.

- The force exerted by the sun on a planet is equal to the mass of the planet times the rate of change of its velocity.
- The force exerted by the sun on a planet points directly toward the sun and its magnitude is equal to the mass of the planet times a constant times one over the distance squared from the sun.

In particular, equating the two expressions for the force, we find that at every point of a planet's orbit, the change in velocity points directly toward the sun and has magnitude equal to a constant times one over the distance squared from the sun. This suggests that we have a dynamical system. But given a dynamical system, we know how to find any future state given a current state and a computer! Newton didn't have computers, but he solved the dynamical system anyway, inventing calculus along the way.

The Most Successful Mathematical Model Ever

Let us try to phrase this description as a dynamical system. First, Newton's assumptions imply that a planet will move in the plane in space determined by its position at any two successive times and the sun. This is because moving off this plane would require a change of direction not pointing towards the sun, hence a force not pointing towards the sun and, by assumption, the force always points directly at the sun.

Imagine that we have drawn two axes, an x-axis and an y-axis, on the plane and that the sun is at the origin. Then the position of a planet at any time can be given by specifying its x-coordinate and its y-coordinate. Let v be its velocity in the x-direction and w its velocity in the y-direction. That is, $x' = v$ and $y' = w$. Thus, just as the position of the planet at any time is given by a pair of numbers (x, y), its velocity is given by the pair of numbers (v, w). Similarly, the pair of numbers (v', w') will give the rate of change of its velocity. If we measure x and y in miles and time in seconds, then v and w will have units of miles/sec and v' and w' of miles/sec/sec.

Now, Newton says the force $\mathbf{F}$ the sun exerts on a planet at position (x, y) points directly toward the sun, which is assumed to be at the origin $(0, 0)$, and that its magnitude is proportional to the product of the mass of the planet and the reciprocal of the distance from the sun squared. To express this symbolically, let's take it a piece at a time.

First, how can we say that the arrow corresponding to the force points directly at $(0, 0)$? The key observation, here, is that at any point (x, y) of the plane, the arrow (x, y) points directly *away* from the origin. We have indicated this in Figure 2 below.

Here is one explanation of why the arrow (x, y) at the point (x, y) points directly away from $(0, 0)$ —you may prefer to construct your own explanation. The reason is that the point (x, y) can be thought of as the head of the arrow beginning at $(0, 0)$ and ending at the point (x, y). Thus, the arrow (x, y) at the point (x, y) is the arrow from $(0, 0)$ to (x, y) slid out along the line joining $(0, 0)$ and (x, y) so that it starts at the point (x, y). In sliding it out along this line, it doesn't change direction, and hence continues to point directly away from the origin. Figure 3 attempts to illustrate this.

If (1,2) is an arrow at some point, then the arrow $-1 \cdot (1, 2)$ (which is equal to $(-1, -2)$) points in the exact opposite direction.

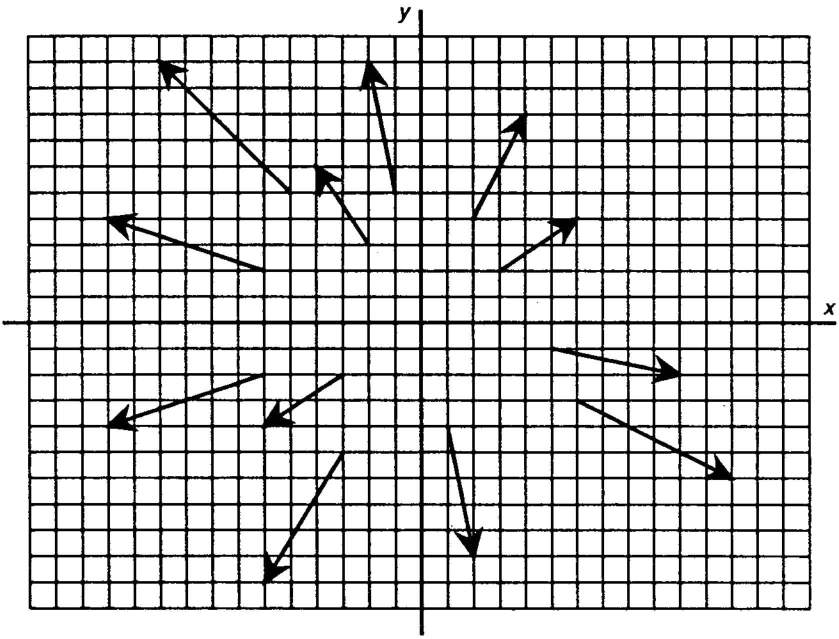

Figure 2. The arrow (x, y) at the point (x, y) points directly away from $(0, 0)$

Thus, since the arrow (x, y) at the point (x, y) points directly away from $(0, 0)$ (that is, the sun), the arrow $-1 \cdot (x, y)$ (or, what is the same thing, $(-x, -y)$) points directly towards $(0, 0)$.

The next part of Newton's prescription says that the magnitude of the force should fall off as one over the distance from the origin squared. Now the magnitude of an arrow is just its length. So we first ask ourselves how to find the length of an arrow. This is easy! Just use the Pythagorean theorem. So, for example, the length of the arrow (3, 1) is $\sqrt{3^2 + 1^2} = \sqrt{10}$. This has the same length as the arrow $(-3, -1)$. In general, the length of the arrow (x, y) is $\sqrt{x^2 + y^2}$. The length of $(-x, -y)$ is $\sqrt{(-x)^2 + (-y)^2} = \sqrt{x^2 + y^2}$, which is the same as the length of (x, y).

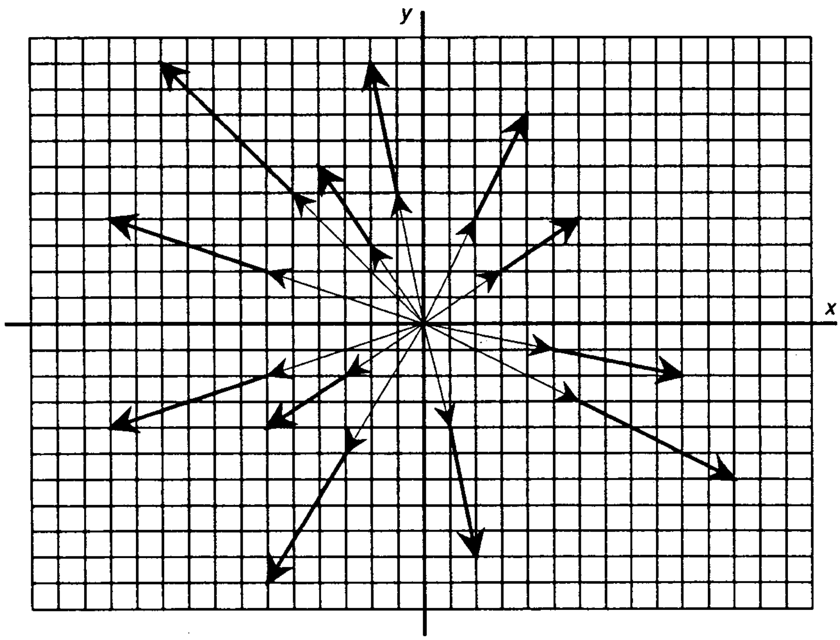

Figure 3. Sliding the arrow joining $(0,0)$ to the point (x,y) out to begin at (x,y)

Exercise 3. (a) In the two-dimensional xy plane, how far from the origin (0,0) is the point (2,-1)? What is the distance of this point from the point (-3,5)?

(b) In the three-dimensional xyz space, how far from the origin (0,0, 0) is the point (2,-1, 6)? What is the distance of this point from the point (-3, 5, 1)?

[Hint (for both parts): sketch the points, construct suitable right triangles, and use the Pythagorean theorem.]

The length of the arrow $(-x,-y)$ gets longer the further the point (x,y) is from the origin. We want it to become shorter. Before figuring out how to do this, let us first ask if we can find a way to keep all the arrows the same length. That is, can we find a set of arrows that point directly towards the origin and which all have length 1, say? Here the key observation is that if s is a positive number, then the arrow $ts \cdot (x, y) = (sx, sy)$ points in the same direction and has length s times the length of (x, y). (The length of (sx, sy) is $\sqrt{(sx)^2 + (sy)^2} = \sqrt{s^2(x^2 + y^2)} = s \cdot \sqrt{x^2 + y^2}$ if $s > 0$.) This means that the way to find an arrow which points in the same direction as a given arrow, but which has length equal to 1 is to multiply the arrow by 1 over its length. Thus, $\frac{1}{\sqrt{10}} \cdot (3, -1) = (\frac{3}{\sqrt{10}}, \frac{-1}{\sqrt{10}})$ points in the same direction as $(3, -1)$ and has length equal to 1. In particular, *the arrow*

$$\frac{1}{\sqrt{x^2 + y^2}} \cdot (-x, -y) = (\frac{-x}{\sqrt{x^2 + y^2}}, \frac{-y}{\sqrt{x^2 + y^2}})$$

at the point (x, y) points directly at the origin and has length 1.

The distance from the origin to the point (x, y) is $\sqrt{x^2 + y^2}$ and we want an arrow at the point (x, y) whose magnitude is equal to the product of the mass of a planet and the reciprocal of the distance from the sun squared. If we let m denote the mass of the planet at (x, y) and K the constant Newton mentions, then the arrow representing the force on a planet must be $m \times K \times (\frac{1}{\sqrt{x^2 + y^2}})^2$ times the arrow above:

$$\begin{aligned} \mathbf{F} &= \frac{mK}{x^2 + y^2} \cdot \frac{1}{\sqrt{x^2 + y^2}} \cdot (-x, -y) \\ &= \left(\frac{-mKx}{(x^2 + y^2)^{3/2}}, \frac{-mKy}{(x^2 + y^2)^{3/2}}\right) \end{aligned}$$

On the other hand, Newton's first assumption tells us that the force at (x, y) is m times the rate of change of the velocity (v', w') at (x, y). So we have

$$\mathbf{F} = (mv', mw').$$

Equating the two expressions for **F** and cancelling m from both sides

$$v' = \frac{-Kx}{(x^2+y^2)^{3/2}}$$
$$w' = \frac{-Ky}{(x^2+y^2)^{3/2}}$$

where K is a positive constant which must be determined experimentally, and which is the same for every planet.

Notice that this is not a dynamical system. The reason is that v' and w' are not expressed as functions of v and w. To make it a dynamical system, we merely have to remember that we defined v and w by setting $x' = v$ and $y' = w$. If we add these two equations to those above, we do get a dynamical system:

$$\begin{aligned} x' &= v \\ y' &= w \\ v' &= \frac{-Kx}{(x^2+y^2)^{3/2}} \\ w' &= \frac{-Ky}{(x^2+y^2)^{3/2}}. \end{aligned} \tag{1}$$

These equations are just the mathematical statements of the two simple assumptions that Newton made.

From what we have said in the last chapters, we know that given (x, y, v, w) at some fixed time t_0 (that is, the position (x, y) and velocity (v, w) of the planet at time t_0), we can determine the state (that is, the position and velocity of the planet) at any future time.

Let's do it. Suppose, for simplicity, that $K = 1$ and that we have the initial condition $x(0) = 6$, $y(0) = 0$, $v(0) = 0$, $w(0) = .1$. The following program will plot the points (x, y) as the time t changes.

```
DEFDBL A-Z
SCREEN 12
```

```
WINDOW (0, 0)-(10, 10)
LET t = 0
LET deltat = .001
LET x = 6
LET y = 0
LET v = 0
LET w= .1
FOR N = 1 TO 12000
xprime = v
yprime = w
vprime = -x/ ((x^ 2+y^ 2)^ (3/2))
wprime = -y/((x^ 2+y^ 2)^ (3/2))
x= x+deltat*xprime
y= y+deltat*yprime
v= v+deltat*vprime
w= w+deltat*wprime
PSET (x, y)
NEXT N
```

One thing we notice right off is that the shape changes quite a lot as we change `deltat`. The answers seem to converge as we take `deltat` smaller and smaller, but one eventually runs out of patience. It can be shown that accuracy is increased if, instead of saying that the the new position is equal to the old plus the velocity at the beginning of the time interval (or at the end of the time interval) times the length `deltat` of the time interval, we say that the new position is equal to the old plus the velocity in the middle of the time interval times the length. That is, $\mathbf{x}(t + \Delta t) = \mathbf{x}(t) + (\Delta t) \cdot \mathbf{v}(t + \frac{\Delta t}{2})$. Similarly, the velocity $\mathbf{v}$ at this halfway point is equal to the velocity at a time Δt before (which is in the middle of the preceeding interval) plus Δt times the rate of change $\mathbf{v}'$ of the velocity. That is, $\mathbf{v}(t + \frac{\Delta t}{2}) = \mathbf{v}(t - \frac{\Delta t}{2}) + (\Delta t) \cdot \mathbf{v}'(t)$. There is one slight problem: what is $\mathbf{v}(\frac{\Delta t}{2})$? We use the special equation $\mathbf{v}(\frac{\Delta t}{2}) = \mathbf{v}(0) + (\frac{\Delta t}{2})\mathbf{v}'(0)$. Making these modifications gives the following program.

```
DEFDBL A-Z
```

```
SCREEN 12
WINDOW (-10,-10)-(10, 10)
LET t = 0
LET deltat = .01
LET x= 8
LET y= 0
LET v= 0+(deltat/2)*(-x/((x^ 2+y^ 2)^ (3/2)))
LET w= .1+(deltat/2)*(-y/((x^ 2+y^2)^ (3/2)))
FOR N = 1 TO 12000
x = x+deltat*v
y= y+deltat*w
r= SQR(x*x+y*y)
vprime = -x/r^ 3
wprime = -y/r^ 3
v= v+deltat*vprime
w=w+deltat*wprime
PSET (x,y)
NEXT N
```

We have plotted the results in Figure 4. Notice that the orbit is elliptical, and one of the foci is (0,0) (the sun)! Just to be sure, we plot the orbits starting at the initial states (6, 0,0,.1),(8,0,0.1), (10,0,0,.1), (12,0,0,.1). These are illustrated in Figure 5. All the orbits are again elliptical!

We are sweeping a rather delicate mathematical point under the rug here. The technique that one uses to numerically integrate a set of rate equations can be quite important. The method we outlined above is a special case of the Runge-Kutta method, which is more accurate than the so-called Euler (or naive) method that we have been using up to now. (This is a standard fact proved in numerical analysis courses – the details are outside the scope of this course.)

Exercise 4. Plot a variety of orbits with different initial conditions. In particular, you should plot some where $v \neq 0$ at time $t = 0$.

Exercise 5. Find the coordinates of the ends of the major and minor axes of the ellipse sketched in Figure 4. (The easiest way to

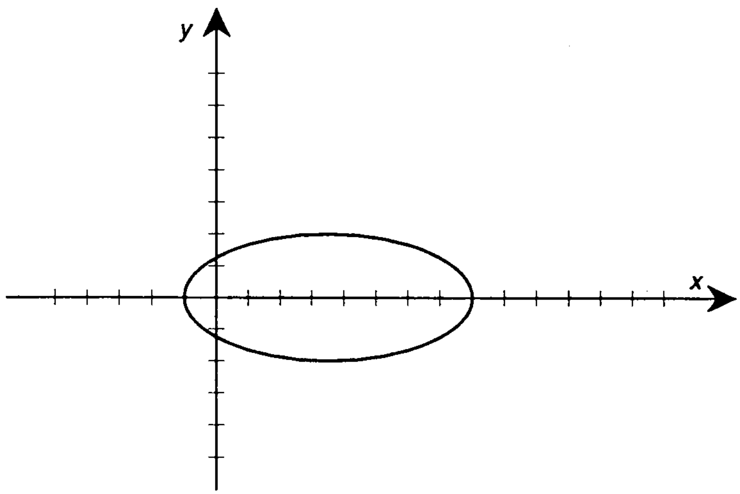

Figure 4. Orbit of a planet with initial state (8,0,0,.1) moving in accordance with (1)

do this is to modify the program to print out x and y values along the orbit.) Using these and exercise 1, prove that the origin is one of the foci.

Newton, of course, did not have access to computers, so that things were not so simple for him. In what is arguably the greatest *tour de force* in the history of science, Newton actually found formulas which gave x, y, v, w as functions of t. He also showed that these solutions satisfied Kepler's three laws.

It is impossible to understate the effect that Newton's achievement had. Few people could follow all the details, but all realized that on the basis of two simple principles, Newton was able to deduce what it had taken Kepler twenty years of painstaking observation and calculation to uncover.

Nor did Newton stop here. He had a penchant for generality and realized that there was no good reason why only the sun should attract

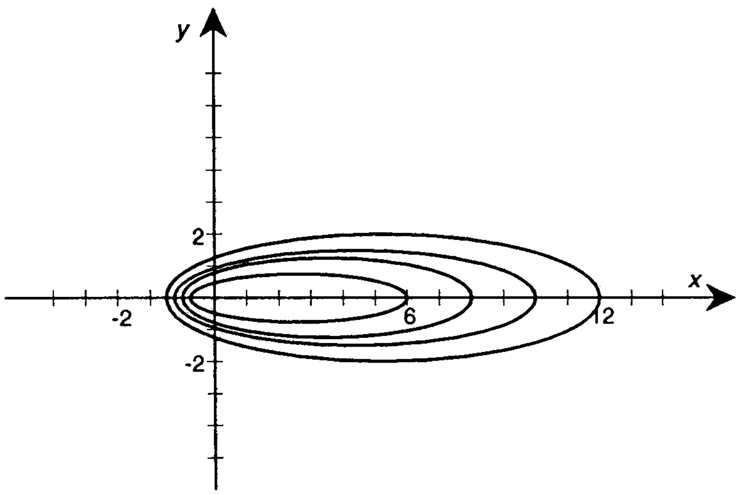

Figure 5. Orbit of a Planet with Varying Initial States

the planets. Doubtless the planets attracted the sun and every other planet: why else should the moon orbit earth? He posited that every object exerts a force on any other object which points directly towards the first object and has magnitude equal to a universal constant (i.e. a constant that is the same for any two objects) times the product of the masses of the two objects times one over the distance between the objects squared. This is called *Newton's Law of Universal Gravitation* and is certainly the most successful model of all time.

Newton's law of universal gravitation allows us to set up a dynamical system which describes the motion of any fixed number of heavenly objects which interact only among themselves. For example, to describe the motion of a binary star with a large planet orbiting one of the stars, we would need six variables to describe the motion of one star, three, call them x_1, y_1 and z_1, to give the coordinates of its position in space and another three, call them u_1, v_1 and w_1, to give its velocity in the x, y and z directions, respectively.

For the second star and the planet, we would need six additional variables each. (We emphasize that the origin is at a fixed point of space – and not at the center of one of the stars or the planet.) Using Newton's law of universal gravitation, we can easily write down a dynamical system involving the eighteen variables and use a computer to determine a state at any future time. The problem of determining the future states associated with this dynamical system is called the **three-body problem.** Despite continuous study in the three hundred years since Newton's time, there are a good many questions that remain unanswered regarding the possible range of behavior of the solutions.

Fortunately, for calculations of the motions of the inner planets in the solar system – say, for example, the orbit of Mars – very good results are obtained by neglecting the forces of attraction of the other planets (since they are so much smaller than that due to the Sun).

Exercise 6. Consider a single planet orbiting the sun whose initial position is (0.5, 0.0, 0.0) and whose initial velocity is (0.0, 1.63, 0.0)). Suppose that the sun attracts the planet with a force that points directly toward the sun, but that the force is proportional to one over the distance of the planet from the sun (instead of one over the distance squared). Suppose, further, that the planet exerts no force on the sun. Compute the orbit of the planet by writing out the position for successive time intervals of 0.001 and plot two circuits about the sun. Do the same for a force which drops off as the cube of the distance. Do either give closed orbits?

The Ubiquity of Dynamical Systems

The success of Newton's Law of Universal Gravitation set dynamical systems on center stage. In the two centuries that followed, it was discovered that almost all of the laws of physics could be phrased as systems of differential equations. The rates of change of variables were often more accessible to observation and conjecture than the variables themselves. As the range of physics broadened, it was discovered that construction of a useful theory usually first involved finding variables whose rates of change were particularly susceptible to analysis. For example, a key step in formulating Newton's Law of Gravitation was realization that the rate of change of velocity admitted a nice description.

The success of dynamical systems in physics suggested that the way to approach other academic disciplines was to set up dynamical systems which allowed one to model the phenomena of interest. In this way it was felt that other sciences could attain the predictive success of physics.

However, the identification of variables and laws governing phenomena outside the purview of physics has been fraught with difficulty. It is easy to think of physical situations which require an enormous number of variables to model as a dynamical system. To model the solar system as a dynamical system would require six variables for each planet, moon, and asteroid. However, physical systems are not as complicated as those that arise outside of physics – at least we know how to write down the equations! Moreover, it is often clear what variables can be neglected in physical situations. For example, we can neglect the gravitational force that asteroids, for example, exert on other planets.

In contradistinction, an ecosystem such as a tropical rain forest supports a huge variety of species which depend on one another in complicated ways. If we let one variable be the number of each species, then we would need a huge number of variables. Here, however, it is far from obvious how to write down the equations. And even if we could, it is difficult to see how we could ever know the values of the variables at any one time. Even ecosystems with far fewer species, such as those on the Canadian tundra, are very poorly understood.

Many biological processes, such as those which describe the respiration of an animal, or those which guide the development of an embryo can be thought of as enormous dynamical systems. What is surprising is that, despite the size of these dynamical systems and despite the fact that the coefficients in the equations making up such a system must vary from species to species and, indeed, individual to individual, there are general qualitative similarities in such processes in different individuals. This suggests that certain qualitative properties of dynamical systems are preserved by changing the equations in various ways. The pursuit of this idea has led to some of the more interesting mathematical research of recent years.

It also points the different ways in which modelling is used in physics and in biology and the social sciences. In physics, we are frequently interested in specific numerical predictions. Where will a given planet be three years and two weeks hence? This sort of precision is out of reach in the biological sciences – we don't know the equations precisely. However, we can use dynamical systems to make qualitative predictions. Moreover, exploration of the different types of behavior that result under different assumptions, and comparison with reality, often gives us insight into the rate equations variables that are most important in understanding biological systems.

Chapter 5 — Dynamical Systems and Determinism

The Art of Modelling

One of the lessons from physics was the importance of identifying the key variables that needed to be tracked. It was the key insight of Galileo that it was not velocity that needed to be explained, but change of velocity. Physics set the standard which the other sciences sought to emulate: identify the relevant variables, find the laws expressing the rates of change of these variables and express them as a dynamical system. Once expressed as a dynamical system, we can use a high speed computer to track states into the future.

Some chinks in this philosophy began to appear within physics itself. With the onset of quantum mechanics, it became increasingly clear that certain physical situations could not be modelled as points in $\mathbf{R}^n$, because the quantities that one wants to measure are not numbers. For example, the uncertainty principle implies that one cannot measure the position and momentum of a particle at the same time to an arbitrary degree of accuracy. This means that the state cannot be thought of as a point in $\mathbf{R}^6$. The solution turned out to be to regard the state not as a point in $\mathbf{R}^n$, but as a function. One could then write an equation for the rate of change of this state (the Schrödinger equation). Build into this new set-up was some fundamental indeterminacy. Nevertheless, quantum effects only appeared at the limits of observation: in the very small, or very short, or very energetic. In the macroscopic world around us, quantum effects were hidden and the physics of Newton applied.

Another problem has been recognized only in recent years. Some very simple dynamical systems turn out to exhibit behavior which is very complicated and extremely sensitive to small changes in initial position. Long term predictions in systems exhibiting this behavior are very problematic.

Identifying variables – Epidemics

In chapter 3, we have looked at a number of population models involving different species. We did this because the variables were easy to identify. In most cases, identifying the appropriate quantities to characterize a state is much more problematic and one of the first steps that must be taken in exploring a situation. Construction of models appropriate to a field should always be undertaken by (or in consultation with) someone thoroughly conversant with the situation under study. At its best, the property of building and validating models is a process in which the relevant variables and concepts are identified and intuitions about the situation are progressively sharpened. In this way, model building is a tool in which the process of building a model leads one to ask questions, make conjectures and test alternate formulations. Model-building, like writing, is a process in which one gains more information in being forced to impose a structure on one's thoughts.

Let us consider a population model in which it is not immediately clear what the variables should be. Let's suppose we are considering a disease with the following characteristcs:

- the disease lasts 10 days
- you catch the disease from contact with someone who already has it
- once you have caught the disease, you can't catch it again for another year

Suppose that we start with 200 people who have the disease. Can we say what will happen in the long run? Will everyone eventually get the disease: might the disease die out? How many people are likely to have the disease at one time?

To answer these questions, we would like to build a dynamical system. The first thing to do is to decide what a state such be. It is clear that there should be at least two categories of people: those who have the disease and those who don't (and, so, we should have the numbers, or mass, of each as a quantity needed to specify the state). However, if we try to write down the rate of change of each of these numbers, it is clear that we need something more.

Let us say that I (for *Infected*) is the number of individuals with the disease and x the number of individuals who do not have the disease. Let's decide to measure time in days. If we try to write down I' we see that it will be the difference between the number of people who get sick each day and the number who get better. We can estimate the number who get better: since the disease lasts ten days, we would expect that the number of people who get better each day is about one tenth of the number who have the disease (this assumes, of course, that everyone who was sick did not get sick on the same day). However, we will have difficulty figuring out the number of people who get sick purely from a knowledge of I and x. The reason is that some of the x people who do not have the disease will be immune and the number of people who come down with the disease will surely be affected by what proportion of the population is immune. Thus, we further subdivide the population. As before, we let I be the number of individuals who are sick; but now we divide the population which does not have the disease into two groups: those who are immune (in this case, those who do not have the disease, but who had it less than one year ago) and those who are susceptible (that is, those who have never had the disease and those who had it more than one year ago). We let R (for *Recovered*) be the number who are immune and S (for *Susceptible*) be the number who are susceptible.

Now we try to write out equations for I', R' and S'. We have already indicated that, since the disease lasts ten days, it is reasonable to suppose that one tenth of those sick get better each day. Can we say how many get sick? Since the disease is spread by contact, we might assume that the number of individuals who get sick is proportional to the number of contacts between susceptible individuals and those who are sick. The number of contacts is in turn, to a first approximation, proportional to the product $I \cdot S$. To get some idea of the magnitude of the constant of proportionality, let's suppose that we are modelling the situation in a medium sized town and that each individual comes in contact with, on average, one thousandth of the town. (This assumption would be way out of line if we were modelling a disease in New York.) Few diseases are so virulent that you necessarily get it if exposed to someone with the disease. Let's suppose that, in our case, it takes on average five contacts with someone diseased to contract the disease. If this is the case, we would expect that the number of individuals who get the disease each day would be $\frac{1}{5} \times \frac{1}{1000} \times IS$. Since the number of individuals who get better is $\frac{1}{10}I$, we have

$$I' = .0002IS - .1I.$$

To get an equation for the rate of change R' of the number of individuals who are immune, note that everyone who recovers from the disease is immune. Since individuals stay immune for one year, each day we would expect that one three hundred and sixty-fifth of those who are immune, lose their immunity. On the other hand, all those who recover each day become immune. Thus,

$$R' = .1I - .0027R$$

(since $\frac{1}{365} = .0027$). We can get an estimate for the rate of change of the number who are susceptible in the same way. This number must be equal to the number who lose their immunity and, hence, become susceptible minus the number who stop being susceptible (by virtue of becoming sick). We have

$$S' = .0027R - .0002IS.$$

Thus, we wind up with the following dynamical system:

$$\begin{aligned} I' &= .0002IS - .1I \\ R' &= .1I - .0027R \\ S' &= .0027R - .0002IS. \end{aligned} \tag{1}$$

We emphasize that we made all kinds of assumptions in the process of getting this system. However, if we accept these assumptions, we can use the system to explore what will happen as time increases.

For instance, let's imagine that we are in a town of 40000 people and that initially 200 people are sick and no one is immune. What can we expect over the course of a couple of years? To determine this, we plot I, R, S as functions of time. Again, we write a simple program. Let's suppose we track the progress of the disease over 2 years and that we choose the time step to be intervals of one tenth of a day. Since there are 730 days in 2 years, we will need to do the computation $10 \cdot 730 = 7300$ times.

```
DEFDBL A-Z
SCREEN 12
WINDOW (0,0)-(730,40000)
I=200
R=0
S=39800
FOR N=1 TO 7300
  PSET (N, I)
  PSET (N, R)
  PSET (N, S)
  Iprime= .0002*I*S - .1*I
  Rprime= .1*I-.0027*R
  Sprime= .0027*R - .0002*I*S
  I = I + .1*Iprime
  R = R + .1*Rprime
  S = S + .1*Sprime
NEXT N
PRINT I, R, S
```

Running the program gives the graphs sketched in Figure 1 and shows that the populations I, R and S tend to the values of 500, 38462 and 1038, respectively . This means that, at any time in our town, most of the 40000 people are immune, about five hundred people will be sick, and a little more than twice that susceptible. Phrased somewhat differently, the trajectory of (1) beginning at the point (200, 0, 39800) tends to the equilibrium point (1038, 38462, 500) (which suggests that the latter is an attractor).

Exercise. It is difficult to convincingly sketch three-dimensional output on the plane, so write (and run) a program to sketch the

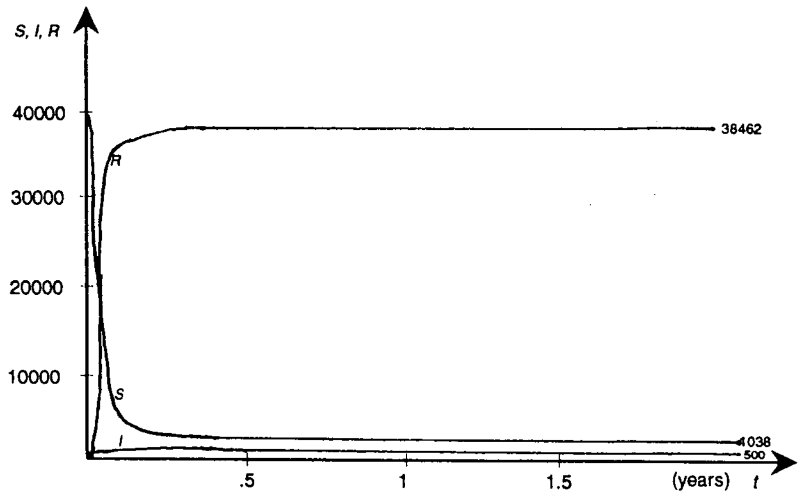

Figure 1. I, R and S as functions of time determined by equation (1)

projections of the trajectory of (1) beginning at (200, 0, 39800) onto the I-R, I-S and R-S planes.

Exercise. Can you find the equilibrium point of (1) analytically? By changing the coefficients in system (1), can you arrange that some of the variables oscillate?

Exercise. How does the conclusion sketched in the text change if half of the population is immune.

Another interesting case occurs if the disease confers permanent immunity (for example, measles or chickenpox as opposed to the common cold). In this case, there is no contribution of the immunes to the susceptible population, so the number R of immunes is always increasing and the number S of susceptibles is always decreasing. We have the system:

$$I' = .0002IS - .1I$$
$$R' = .1I$$
$$S' = -.0002IS.$$

The more general system

$$I' = aIS - bI$$
$$R' = bI$$
$$S' = -aIS$$

is a well-known epidemiological model, called the S-I-R model. S-I-R models have been used to study outbreaks of the plague. In this case, one has to allow for deaths (so that $S + I + R$ is not constant). In one study (G. F. Raggett, Modelling the Eyam Plague, *Bull. Inst. Math. and its Appl.*, 18 (1982) 221-26), a modified S-I-R model is applied to a plague outbreak in the English village of Eyam in 1665-66. Raggett shows how to determine the parameters. The agreement between the actual data and the model is very good. For more on S-I-R models see Murray's book and the references therein.

Transferability of Models — Rumors

It sometimes happens that studying a model built for one purpose can yield insight into another area. Sometimes, the whole model can, with suitable reinterpretation, be used in an entirely different context. In fact, a useful model can serve as a metaphor for other situations.

As an example let us consider rumors. Probably all of us have seen demonstrations of how rumors spread and how they change as they spread. The party game in which one person begins by whispering a message to one other person, together with instructions to pass the message on to another person, who in turn is to pass it on to yet another, and so on, until the message finally comes back to the person who initiated it, shows that there can be a huge difference between what started out and what the last person wound up hearing.

There is, however, a class of rumors which have been studied and which remain remarkably similar over time. These are the so-called urban myths. There are a number of features that characterize them. They are usually exchanged between friends, often in confidence, with the preliminary remark that what is about to be related happened to a unspecified friend (or a friend of a friend). The reader will almost certainly recognize some of them: one concerns the lady who tried to dry off her wet cat and popped it into the microwave oven for a half a minute ... (a more gruesome variant is the babysitter who try to dry a newborn in the same way); another, an about-to-be-married couple who decide to slip downstairs in the altogether for a little fun and frolic, only to walk into a surprise pre-nuptial party.

Such stories spread across the country and it is interesting to ask how long it takes for most people to have heard them. Here, the most useful models divide people into those who have not heard the rumor (the Susceptibles), those who have heard the rumor, believe it and are actively spreading it (the Infecteds), and those for whom the rumor has become stale news or who have grown to disbelieve the rumor (the Recovereds). Depending on the modelling assumptions, one might or might not imagine the recovereds as becoming susceptible again. The reader may enjoy constructing a specific example of such a model as an exercise.

Examples of a similar sort arise in some disease models which treat germs as predators and susceptible individuals as prey.

Complex systems and structural stability

So far we have concentrated on systems with relatively few degrees of freedom. The principles are the same for systems with a large number of variables, although the number of computations required to get anything useful can be immense. For example, since an ounce of matter typically contains over 10^{23} molecules, if we wanted to model the behaviour of an ounce of matter by keeping track of the motion of separate molecules, we would need at least 6×10^{22} variables (three for the position of each molecule in space and three for the velocity in each of the coordinate directions), so that a state of the system would correspond to a point in $\mathbf{R}^{6\times10^{22}}$. Even this is a simplification, for it assumes that we can treat the molecules as points in space – in actual fact, the molecules can have a definite orientation and they can be spinning. Fortunately, we are rarely interested in the position and velocity of each of the constituent molecules in a given amount of matter. For most purposes, we are interested in more macrosopic properties. Thus we can take averages of the variables, arriving at quantities like the average velocities of the molecules (which is closely related to the quantity we call temperature) or the density of the matter.

On the other hand, a gram of matter is comparatively simple. Real ecosystems are much more complex. We would need one variable each for the biomass of each different species. In addition, we might want to divide the species up by age structure and take into account seasonal variations and climactic considerations. Writing down the equations for the rates of change would be largely a matter of guesswork – it is hard to imagine that the coefficients could be known within even an order of magnitude.

Despite this complexity, there are regularities in ecosystems. This is somewhat mysterious because, unlike the situation with a gram of matter, it is not clear what, if anything, averages of the variables mean. What does it mean to compute the average biomass of rabbits, foxes and berries?

An even more striking example is the development of a human embryo. If one tries to model this as a dynamical system, the sheer number of variables is overwhelming: for a start, one needs the concentration of thousands of different organic compounds (lipids, enzymes, hormones, free radicals) in the uterus. Moreover, every woman is different, so there will be individual differences. Nonetheless, the process of embryo development is remarkably similar in all human beings – there are a number of distinct stages (blastulation, gastrulation, etc.) which the embryo goes through. The time frame is remarkably similar for different individuals (quickening occurs at 12 weeks, total gestation period is about 40 weeks). In fact, the existence of the science of medicine is testimony to the similarities between different individuals – if everyone were too different, there would be no way of identifying the same disorder in different individuals or of being relatively certain that a drug that worked for one person would work for another.

In pondering the question of how regularities could appear in very complicated dynamical systems (such as those governing embryo development or ecosystems), René Thom, a French mathematician advanced a very ingenious theory of models. He began with the observation that, given a dynamical system and a state changing in accordance with that dynamical system, the state rapidly moves to an attractor. Thus, when observing a real system, one should (unless there is some compelling reason to do otherwise) assume that the state corresponding to it was near an attractor of a dynamical system. Thom then proposed to think of the dynamical system, itself, as changing. As the dynamical system changed, the attractor would change and hence the observed state would change.

The notion that the states that one observes are at or near attractors jibes well with our experience with models of two species systems. In all instances we saw that the populations settled down very rapidly either to a fixed equilibrium population (corresponding to a point attractor) or to a steady cyclical behavior (corresponding to a limit cycle). In Thom's view, as environmental conditions change the parameters in the model would change – perhaps the growth rate of a certain species would be altered or the carrying capacity. This change would alter the position of the attractor. In certain instances, the attractor might change its nature entirely: splitting into two or more attractors, turning from a point into a limit cycle, or even disappearing altogether. Thom called these sudden changes in the nature of an attractor *catastrophes* (whence the name *catastrophe theory* to refer to this type of modelling).

Thom, following Poincaré, emphasized that the key role played by dynamical systems which were structurally stable in modelling biological and ecological situations. Intuitively a system is *structurally stable* if small changes in the dynamical system results in dynamical systems which have trajectories (and, in particular, attractors) which do not differ too much from those of the original system.

Mathematically, the difficulty is to define what is meant by "small" changes and what is meant by "not differing too much".

But Thom went further. He and his coworkers established that many *catastrophes* were stable in the sense that changes between similar systems gave rise to similar changes in similar attractors and, hence, to changes in form which were recognizably similar. He observed that one could talk of stability not just of dynamical changes, but of *changes* in dynamical systems. The mathematical difficulties in defining what structurally stable change of form means were formidable. Nevertheless, Thom succeeded in many cases. One celebrated result conjectured by Thom (and proved by Mather) was that there were only seven different types of structurally stable changes that could occur between attractors upon varying four or fewer parameters in the class of dynamical systems known as gradient systems. These structurally changes were called the *elementary catastrophes* and the details of the changes were widely studied. (As sometimes happens with new results, the theory was oversold and exaggerated claims were made on its behalf: the popular press even got into the act.)

Gradient dynamical systems enjoy a number of very special properties, For example, all of their attractors are point attractors: they cannot have limit cycles. It was widely recognized at the time that "most" systems were not gradient systems. Nevertheless it was hoped that most systems were like gradient systems in that their attractors were point attractors or, at worst, limit cycles. This hope offered the possibility that most morphological processes could be understood in terms of the elementary catastrophes and a handful of other transitions (such as the much studied change of a point attractor into an limit cycle, the so-called *Hopf bifurcation*, and its higher dimensional analogues).

As well shall see, this hope was too optimistic. However, catastrophe theory and the underlying mathematics (now known as *singularity theory*) underwent intensive development and advanced our understanding of a great number of phenomena, notably in optics. The viewpoint of catastrophe theory, namely that of trying to model a situation as an an attractor of a dynamical system, and then allowing the system to vary in prescribed ways, is exceedingly useful in a wide variety of contexts. More controversial was Thom's use of his models of change (the elementary catastrophes) as investigative tools in areas where it was not clear that the underlying dynamical systems were gradient dynamical systems. The debate on this issue has been fascinating and we encourage the reader to browse through Thom's book and the reviews of it in the lead journals of the mathematical, biological and physical professional societies. A similar sort of debate is currently occurring regarding "chaotic" systems. The underlying conflict seems to revolve around the question of whether one should use purely mathematical constructs as metaphors and speculative devices.

Weather — Chaotic Dynamical Systems

The success of Newton's theory of gravitation and the vision it afforded of a universe in which all objects moved, in possibly complicated ways, in accordance with a set of simple rules betokened the possibility that everything might, at root, be reducible to a set of simple physical laws. In this view, there are a few underlying dynamical systems, knowledge of which, together with knowledge of the values of the variables, would allow one to predict what might happen arbitrarily far into the future. (Remember that knowledge of a dynamical system, allows one to say what will happen to any state over time.) "Such an intelligence", wrote Laplace, speaking of the Prime Mover, "would embrace in the same formula the movements of the greatest bodies of the universe and those of the lightest atom; for it, nothing would be uncertain and the future, as the past, would be present to its eyes." And, indeed, many of the advances in physics seemed to bear this out.

One situation that a great deal of effort has gone into has been weather prediction. The earth's atmosphere is a fluid and the equations that describe a fluid, the *Navier-Stokes equations*, are well-known, albeit far from being thoroughly understood.

Since a fluid is a continuous medium, it is not appropriately modelled as a point in $\mathbf{R}^n$, but rather as a collection of functions (of latitude, longitude, height above the earth, and time). The set of all such collections of functions forms a mathematical space which has an infinite number of dimensions. The equations which describe the rate of change of such objects are not equations of a type we have encountered so far, (they are partial differential equations). Nevertheless, by looking at the atmosphere at a large, but finite number of places on the earth, we can take an approximation of the state space to be $\mathbf{R}^N$ for some large but finite N. The Navier-Stokes equations then define a dynamical system on this space. Given a set of observed values at the chosen observation points, we can then use the dynamical system to project what the values will be at future times, at least in theory.

This is the technique used by the National Metereological Center in the United States and a number of weather forecasting centers in other countries. The best forecasts come out of the European Center for Medium Range Weather Forecasts in Reading, England. This center is funded by the Common Market countries, who decided to pool their resources to provide accurate weather prediction in Western Europe. It is staffed by a rotating crew of young scientists, mathematicians, computer programmers and technicians from the Western European countries and boasts some of the best computer facilities in the world.

The European Center divides the earth's atmosphere into 100 kilometer by 100 kilometer by 1 kilometer boxes and takes temperature, pressure, humidity, and a couple of other readings in each (actually, in some boxes more readings are taken). Each such reading in each such box represents the value of a separate variable. It takes over one hundred thousand such boxes to represent the atmosphere. Since there are at least five readings taken in each box, the atmosphere is represented by a point in $\mathbf{R}^N$ where $N > 500000$. Since each variable requires an equation giving its rate of change, the model requires more than a half a million equations.

The dynamical system is run on a Cray Supercomputer which can accept observations fed in at the rate of 100 million observations per second. The Cray does 400 million calculations per second. At this rate, it takes it three hours to generate a ten day forecast. (To get an idea of how fast this is, note that in following the course of the cold epidemic two years into the future, we had to do about 140000 calculations since each pass through the loop requires 18 calculations and we had to pass through the loop 7700 times. The Cray computer would complete the calculation in one three thousandth of a second.)

Incidentally, the "Cray" in Cray computer refers to Seymour Cray, a reclusive genius who founded the company that makes the computer. Another Cray computer is owned by Lucas Films, which uses it to help create the special effects seen in such films as *Star Wars*.

Mathematical modelling performed on high speed computers has changed weather forecasting into a science. The European Center estimates that several billion dollars are saved a year as a result of its forecasts. Nevertheless, anyone who has watched the weather forecasts on television will realize that the forecasts are often wrong. The long-range forecasts extending beyond a day or two are particularly subject to error. Until very recently, there was great optimism that better computers and better means of gathering data (more satellite coverage and remote control measurement of far off regions) would yield very accurate predictions.

After all, no one seriously questions the validity of the Navier-Stokes equations. They are grounded in Newtonian physics which has been thoroughly tested and are known to accurately describe the behavior of fluids. So it would just seem a matter of making sure that we had enough data and computers fast enough to process them. It is not unreasonable to suppose that given sufficient investment, we could arrange to record readings every cubic kilometer. The increases in computer speed seem likely to continue through the next decade as new technologies are brought on line.

Indeed, the inventor of the digital computer, John von Neumann, confidently predicted that within a century we would not only have the capacity to predict the weather months in advance, but also to alter it at will. This dream was seriously called into question by a discovery Edward Lorenz made in 1962. Lorenz was a metereologist at MIT with a strong mathematical bent. He had constructed a model weather system which mimicked that of the earth, but on a much smaller scale. Lorenz's system had only twelve equations, which were a rough approximation of the Navier-Stokes equations.

Lorenz ran the equations on a computer in his office, looking for regularities and patterns in his model weather system. At that time, computers were much more primitive than they were today – a mass of tubes and circuit boards and wires. Lorenz's computer, a Royal McBee, made 60 calculations per second and was always breaking down. One day in the winter of 1961, Lorenz wanted to re-examine part of a run he had made earlier. Rather than repeat the whole run, he typed back in values that the computer had given as output about midway through the run and left to get a coffee. Upon returning about an hour later, he noticed that the new output had begun to diverge significantly from the output he had obtained earlier. On comparing the two graphs of the output, he noticed that they were very similar at first, but after a while started to bear no resemblance to one another.

Since a dynamical system is completely deterministic, this seemed impossible. In an autonomous dynamical system, if you feed in a set of values on one day and a day later you feed in the same set of values, you will get the same output. The initial conditions determine the behavior of the system for all time. Lorenz carefully re-examined what he had done. It turned out that he had not fed in exactly the same values. His computer gave output to six decimal places: to shortcut typing time, Lorenz had fed back only the first three decimal places, figuring that a difference of less than one thousandth in the initial values was not going to make too much difference. This difference of less than one thousandth was enough to totally change the behavior!

This is in sharp contrast to the two-species models we looked at earlier. If one changes the values of the initial numbers of each species by one thousandth, one winds up with exactly the same values in the long run. The system damps out slight fluctuations in the inputs. Lorenz's system seemed to do the opposite – as time went on it magnified the differences in inputs.

Lorenz realized the staggering consequences of his discovery. There was no contradiction with absolute determinism. If you fed in *exactly* the same values, you would get exactly the same output. But, in the real world, it is impossible to measure the values of physical quantities such as temperature and humidity exactly. Moreover, by measuring them at 100 kilometer intervals, one was ignoring changes in the values that could occur in the 100 kilometer intervals between the places one took the measurements.

Over the next year, Lorenz tracked down where this behavior was coming from. In the end, he boiled his system of twelve equations down to three equations which represented a highly idealized description of motion of a fluid in a horizontal layer which was being heated from below. The fluid on the bottom heats up and rises. As it rises it cools and tends to fall to the bottom, re-approaching the source of the heat, where it heats up again and begins to rise. This motion is called convective motion (or convection). For low values of the heat this sets up a circular motion in the fluid. As the heat is turned higher, wobbles appear in the circular motion. As the heat is turned still higher, the whole pattern breaks down and becomes turbulent. This is because the motion speeds up and some of the cold fluid that has come from the top begins to circulate back up, pushed along by the current before it has warmed up. Instead of making it all the way up and around, it may fall back toward the bottom, reversing the direction of the cycle. Soon, any recognizable pattern breaks down. Lorenz's equations were the following:

$$\begin{aligned} x_1' &= -10x_1 + 10x_2 \\ x_2' &= 28x_1 - x_2 - x_1x_3 \\ x_3' &= -\frac{8}{3}x_3 + x_1x_2. \end{aligned} \tag{2}$$

Here, x_1 measures the intensity of the convective motion and x_2 the difference in temperature between the ascending and descending currents. When x_1 and x_2 have the same sign, it means that warm fluid is rising and cold fluid is descending. The variable x_3 measures the amount by which the change in temperature fails to fall linearly with height. A positive value, means that the temperature falls much more rapidly near the boundary than higher up. The crucial coefficient is the the number 28 multiplying the variable x_1 in the second equation. If this number were less than $\frac{470}{19} \approx 24.74$, then the equations would have point attractors corresponding to steady convection. Above this value, all hell breaks loose.

Lorenz shows that any solution starting within some ball of finite radius about the origin remains bounded (and therefore must tend to some attractor). The point $(0, 0, 0)$ is a rest point (corresponding to no convection), but it is not an attractor – trajectories tend away from it. The points $(6\sqrt{2}, 6\sqrt{2}, 27)$ and $(-6\sqrt{2}, -6\sqrt{2}, 27)$ are also rest points (they correspond to the states of steady convection, but they are also unstable. Upon starting at a point (0,0,0), the solution will loop to the right a few times, then to the left a few (but possibly different number of times), then to the right and so on back and forth from left to right in an irregular manner. In Figure 2 we have sketched a solution for 50 loops. We have also sketched the plane $z = 27$ – it contains equilibrium points which correspond to states of steady convection, but which are not attractors. The parts of the trajectory below this plane are sketched as dotted lines. If we were to start at a different initial point, we would get a similar looking picture, but the number of loopings to the right and the left, and the sequence in which they occur, would be totally different (and equally as unpredicatble).

The figure that is sketched represents an attractor that is neither a point nor a closed curve. Rather it is an infinitely long line which is confined to a region in space. It seems to lie on a surface: closer examination reveals this to be the case, but the surface is like a set of very thin strips which interleave in a very complicated fashion. For more details, we refer the reader to Lorenz's original (and very readable) paper "Deterministic Nonperiodic Flow" in the *Journal of Atmospheric Sciences*, 20 (1963), pages130-41.

Attractors which lie in a bounded region of space and which are neither points nor closed curves have been dubbed *strange attractors*. Such attractors exhibit the phenomenon called *sensitive dependence on initial conditions*, which means that states initially close together evolve in radically different ways over longer and longer time intervals. The sensitive dependence on initial conditions means that, even though the system is deterministic, long term

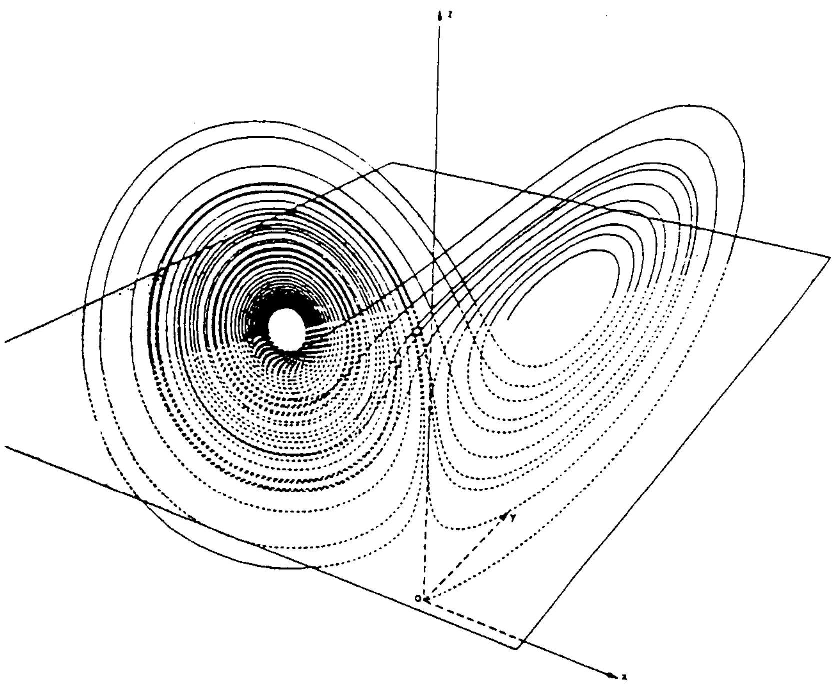

Figure 3. The Lorenz attractor.
Oscar Lanford made this picture of a trajectory
of system (2) beginning near (0,0,0)

prediction is, in practice, impossible. More colloquially, this phenomenon has come to be known as the *butterfly effect*, referring to the possibility that the draft caused by the movement of the wings of a butterfly in Tokyo could translate one month later into a snow storm in New York City. The image seems to have been inspired by the resemblance of Lorenz's attractor to a butterfly.

The reader is encouraged to write and run a computer program to verify that the behavior sketched above does indeed occur – there is nothing like verifying such things for yourself. A step size of .005 works fine.

Because Lorenz's paper first came out in a metereological journal, it wasn't until the mid 1970's that it came to the attention of mathematicians. In the meantime, mathematicians had come to realize that attractors of structurally stable systems could be more complicated than they had first thought and they had examples of dynamical systems which seemed to provide instances of this. However, the mathematicians' examples struck other scientists as rather contrived. Lorenz's example was immediately accepted as natural.

At the turn of the century, Henri Poincaré, had discovered that there were some extremely complicated solutions to Newton's equations describing the motions of just three bodies moving in accordance with the law of gravity. He had observed that slight changes in the initial conditions resulted in huge changes in the subsequent motion and despaired of ever fully understanding them. In *Science and Method*, he wrote

> A very small cause which escapes our notice determines a considerable effect which we cannot fail to see, and then we say that the effect is due to chance. If we knew exactly the laws of nature, and the situation of the universe at the initial moment, we could predict exactly the situation of that same universe at a succeeding moment. But even if it were the case that the natural laws had no longer any secret for us, we could still know the situation approximately. If that enabled us to predict the succeeding situation with the same approximation, that is all we require, and we should say that the phenomenon had been predicted, that it is governed by the laws. But it is not always so; it may happen that small differences in the initial conditions produce very great ones in the final phenomena. A small error in the former will produce an enormous error in the latter. Prediction becomes impossible ... .

However, the equations Poincaré had examined were not structurally stable (because the nature of the solutions changes if one adds a friction term to the system). It was widely felt that if you added friction, so as to get something structurally stable, the friction would damp out the large disturbances (and the whole system would tend toward an equilibrium point).

When Lorenz's paper was "discovered" it caused an enormous stir. Here was a set of simple equations which actually arose in physics, which was structurally stable, and which embodied what Poincaré had feared. Before Lorenz's paper, almost no one would have guessed that such a simple system could exhibit such behavior. Indeed Lorenz's equations are, at first sight, no more difficult looking than the equations we wrote down describing the spread of a common cold.

Since then even simpler equations have been observed to exhibit sensitive dependence on initial conditions. Such behavior is termed "chaotic" and the corresponding dynamical systems *chaotic dynamical systems*. At present, there is an enormous amount of work being devoted to investigating such systems.

The Last Word

The discoveries detailed in the last section have resulted in a very different approach to many problems. Before 1970, variations from regularity and periodic behavior were seen as the result of external influences. The slight, but unpredictable variations in ecological cycles had been seen as the result of "random" environmental influences. Nowadays, it is much more common to look for the cause of these seemingly random variations within the dynamical system itself. Radical swings in the stock market are much less likely to be considered as the result of a catastrophe changing the dynamical system, than as a reflection of the possible existence of a strange attractor in the underlying dynamical system.

Scientists are much more cautious about the validity of long term predictions based on dynamical systems. At the European Center for Medium Range Weather Forecasts, all long range weather forecasts are run several times with slightly different initial conditions. If the forecasts agree, then greater confidence is placed on the forecast. If they disagree, then it is assumed that the initial state is in an area where it is being attracted by a strange attractor, and the forecast is taken to be unreliable.

This concludes our story. Dynamical systems, coupled with high speed computing, will continue, at least for the foreseeable future, to be one of the principal tools for modelling and understanding reality. It is well to remember that: 1) they are not the sole way of modelling our world and 2) even if one knows the equations of a dynamical system, one's problems may be just beginning. Nonetheless, dynamical systems can be a powerful speculative tool. They are the chiel tool for prediction in surprisingly many sciences and social sciences. Moreover, the effort to model a situation as a dynamical system, much as the effort to phrase a thought in written English, is likely to be instructive.

Further Reading

P. Cvitanovic, *Universality in Chaos*, (Second Edition) , Bristol, Adam Hilger, 1989.

Five College Calculus Project, *Calculus in Context I & II*, Five Colleges, Amherst 1991.

T. Fritts. The Brown Tree Snake, *Boiga irregularis*, a Threat to Pacific Islands, U .S . Fish Wildl. Serv., Biol. Rep., **88(31)** (1988) 36pp.

J. Gleick, *Chaos: Making of a New Science*, New York, Penguin, 1987.

M. Hirsch and S. Smale, *Differential Equations, Dynamical Systems and Linear Algebra*, Academic Press, New York, 1974.

A. Koestler, *The Watershed*, Doubleday, New York, 1963.

E. Lorenz, Deterministic Nonperiodic Flow, *Journal of Atmospheric Sciences*, **20** (1963), pages130-41. (Reprinted in Cvitanović, pp. 367-378)

E. Lorenz, On the Prevalence of Aperiodicity in Simple Systems, in *Global Analysis*, eds. Mgrmela and J. Marsden (New York, Springer Verlag, 1979), pp. 53-75.

R.M. May, *Stability and Complexity in Model Ecosystems*, 2nd ed., Princeton, Princeton Univ. Pr., 1975.

J. Maynard Smith, *Models in Ecology*, Cambridge, Cambridge Univ. Pr., 1982.

J.D. Murray, *Mathematical Biology*, New York, Springer-Verlag, 1989.

A. Recigno and I. Richardson, The struggle for life; I, Two species, *Bull. Math. Biophysics*, **29** (1967), 377-388.

J.A. Savidge, Extinction of an Island Forest Avifauna by an Introduced Snake, *Ecology*, **68(3)** (1987), 660-668.

R. Thom, *Structural Stability and Morphogenesis: An Outline of a General Theory of Models*, Reading, W. A. Benjamin, 1987.

P. Waltman, *Competition Models in Population Biology*, Philadelphia, SIAM, 1983.