Chapter 3
Discrete Dynamical Systems

You are probably acquainted with exponential growth or decay of a single quantity: organisms in a stable environment (e.g., dollars in a bank) growing — or decaying — by a fixed percentage every year, day or hour. Say, you are looking at a gob of algae which grows by 14% every hour. Starting with x_0 grams of biomass, you will have $x_1 = 1.14x_0$ grams after the first hour, $x_2 = 1.14x_1 = (1.14)^2 x_0$ grams after the second, ..., $x_n = (1.14)^n x_0$ grams after the n-th hour.

Now imagine you throw in y_0 grams of algae-eating amoebas, each gram of which destroys 0.12 grams of algae an hour. You then have $x_1 = 1.14x_0 - 0.12y_0$ grams of algae an hour later. Suppose that, without food, the amoebas would die at the rate of 14% per hour, but they increase by 0.08 grams an hour for every gram of algae present. After one hour, you then have $y_1 = 0.08x_0 + 0.86y_0$ grams of them.

The transition from one hour to the next can therefore be described as a matrix multiplication $X_1 = AX_0$, i.e.,

$$\begin{bmatrix} x_1 \\ \\ y_1 \end{bmatrix} = \begin{bmatrix} 1.14 & -0.12 \\ \\ 0.08 & 0.86 \end{bmatrix} \begin{bmatrix} x_0 \\ \\ y_0 \end{bmatrix}.$$

Likewise $X_2 = AX_1 = A^2X_0$ and $X_3 = AX_2 = A^3X_0$, and so on up to $X_n = A^nX_0$ after n hours. In the Cartesian plane, this leaves a trail of points $X_0, X_1, \ldots, X_n, \ldots$ Different initial positions X_0 produce different trails. Together all these trails form a "discrete dynamical system" depicting the multiplicative action of A. The eigenstuff from Chapter 2 emerges as the key to the study and classification of such systems.

§1. Powers and Orbits

A similarity relation $A = MBM^{-1}$ is easily extended to matrix powers, that is to say

$$A^n = MB^nM^{-1}, \tag{3.1}$$

for any integer n. If n is positive, this follows at once from the obvious equation $(MBM^{-1})(MCM^{-1}) = MBCM^{-1}$, by putting $C = B, B^2, \dots$ and so on. For $n = -1$, the same equation with $C = B^{-1}$ will do the trick, and then it extends to any negative n as above.

As shown in the last chapter, M can always be chosen so as to conjugate a given A into a B having one of the following three standard forms:

$$\text{(a)}\ \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \quad \text{(b)}\ \begin{bmatrix} r & 1 \\ 0 & r \end{bmatrix} \quad \text{(c)}\ \begin{bmatrix} r & -s \\ s & r \end{bmatrix},$$

depending on the nature of its characteristic equation. It is clear how to take the n-th power of the first of these. For the second, note that $(B-rI)^2 = 0$, whence $B^n = \big(rI + (B - rI)\big)^n = r^nI + nr^{n-1}(B - rI)$. The third one requires a dip into geometry: writing $B = \rho R(\theta)$ as in Section 2.3, we get $B^n = \rho^n R(n\theta)$, because $R(\theta)^n$ is the n-fold iteration of a rotation through the angle θ and thus equals $R(n\theta)$.

Here is the corresponding line-up of n-th powers B^n:

$$\text{(a)}\ \begin{bmatrix} \lambda_1^n & 0 \\ 0 & \lambda_2^n \end{bmatrix} \quad \text{(b)}\ \begin{bmatrix} r^n & nr^{n-1} \\ 0 & r^n \end{bmatrix} \quad \text{(c)}\ \rho^n \begin{bmatrix} \cos n\theta & -\sin n\theta \\ \sin n\theta & \cos n\theta \end{bmatrix}.$$

If $B = M^{-1}AM$ is in standard form, the appropriate B^n can be used to compute A^n as MB^nM^{-1}. For illustration, let us go back to the matrix A we met in the preamble of this chapter. You should check that it is diagonalizable with eigenvalues 1.1 and 0.9. The respective eigenvectors $[3, 1]$ and $[1, 2]$ can be used as columns of a matrix M, and the powers of A can then be obtained as

$$A^n = \begin{bmatrix} 1.14 & -0.12 \\ 0.08 & 0.86 \end{bmatrix}^n = \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1.1^n & 0 \\ 0 & 0.9^n \end{bmatrix} \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}^{-1}.$$

For $n = 15$ and $n = 35$ — using the approximate values $(1.1)^{15} \approx 4.18$, $(0.9)^{15} \approx 0.206$ and $(1.1)^{35} \approx 28.1$, $(0.9)^{35} \approx 0.025$ — you should check that it gives the following results:

$$A^{15} \approx \begin{bmatrix} 4.97 & -2.38 \\ 1.59 & -0.59 \end{bmatrix} \quad \text{and} \quad A^{35} \approx \begin{bmatrix} 33.72 & -16.85 \\ 11.23 & -5.59 \end{bmatrix}. \tag{3.2}$$

Many applications involve a point or vector being subjected to the same "transition matrix" A again and again. More precisely, you start with an initial point X_0, and then generate a sequence $X_1, X_2, \ldots X_n, \ldots$, called the A-*orbit* of X_0, by the recipe

$$X_n = AX_{n-1}\,, \qquad \text{i.e.,} \quad X_n = A^n X_0.$$

In discussing orbits, we shall consider only non-singular matrices A, and we shall enlarge our notion of orbit to include $X_{-1}, X_{-2}, \ldots$, in other words, all $X_n = A^n X_0$ for $n < 0$ as well. If the matrix happens to be a rotation, X_n will of course go round and round, literally staying in orbit as shown by the following diagram.

FIGURE. 3.1. Part of the B-orbit of $X_0 = (1, 0)$, where $B = R(4^\circ)$.

However, such closed orbits are rare: for most matrices of Type (c) — i.e., without eigenlines — an orbit will form a kind of spiral. In Cases (a) and (b) the typical "orbit" looks more like a dotted curve which is mysteriously attracted to an eigenline. We shall survey these various possibilities in the next section. For now let us stick to the diagonalizable matrix A introduced in the preamble, whose 15-th and 35-th power we computed above. Figure 3.2 shows some of its orbits.

If the initial point X_0 lies on an eigenline, so does the whole orbit: in fact, if λ is the corresponding eigenvalue, $X_n = \lambda^n X_0$ is just a scalar multiple of X_0 running toward the origin for $\lambda < 1$, away from it for $\lambda > 1$. In our numerical example, the eigenlines are $y = 2x$ (with $\lambda = 0.9$) and $3y = x$ (with $\lambda = 1.1$). An orbit which does not creep along within the first of these will eventually be bullied into alignment with the second.

FIGURE. 3.2. Various orbits for the transition matrix $A = \begin{bmatrix} 1.14 & -0.12 \\ 0.08 & 0.86 \end{bmatrix}$.

It is not hard to see why this must happen. Consider a matrix A with eigenvectors V and W whose corresponding eigenvalues are of the form $\lambda > 0$ and $\mu = s\lambda$ with $|s| < 1$, in other words, any diagonalizable (non-scalar) matrix whose trace is positive. Picking $X_0 = V + W$ for our initial point, we get

$$A^n X_0 = \lambda^n V + \mu^n W = \lambda^n \big(V + s^n W\big) \tag{3.3}$$

for the point X_n, which therefore lies in the direction indicated by the vector $V + s^n W$. As n grows and grows, $s^n W$ shrinks and shrinks — remember $|s| < 1$ — so that the direction in question becomes more and more that of V itself. Thus, the orbit eventually lines up with the eigenline of λ, the larger one (in absolute terms) of the eigenvalues. Slogan:

A typical orbit is attracted to the dominant eigenline.

But is our initial point X_0, the sum of two eigenvectors, really so typical? Yes, since $M = [V, W]$ is invertible, *any* point X_0 can be written in the form $MY_0 = u_0 V + v_0 W$. This is again the sum of two eigenvectors — unless u_0 or v_0 is zero, i.e., X_0 already lies *on* an eigenline. Hence the only orbits which are *not* typical in the sense of the slogan are those running entirely inside one of the eigenlines.

The effect of this attraction can also be observed in the results of Eq. (3.2). After all, the columns of A^n trace out the A-orbits $A^n I^{(1)}$ and $A^n I^{(2)}$ beginning at $(1, 0)$ and $(0, 1)$, respectively. Figure 3.2 suggests that

these would approach the line $(3t, t)$ from below and above, respectively — as is borne out by the numerical evidence. Hence the columns of A^n tend, in the long run, to be "approximate eigenvectors". This phenomenon persists for larger than 2×2 matrices, and is the basis for several techniques of computing eigenvectors.

What about the bioblob of the preamble? The main result is that, unless the ratio of algae to amoebas is *exactly* 1:2, the system tends to evolve toward that ratio becoming 3:1 (the dominant eigenline) and both component populations growing at 10% per hour forever after. Our model predicts this kind of evolution if the initial state was below the line $(t, 2t)$. If, however, the amoebas were initially too numerous, the orbit will crash into the y-axis — and the algae will be wiped out. All this makes sense.

However, absurdity lurks near the edges. When the orbit crosses the y-axis, and continues to strive for the magic 3:1 ratio somewhere in the third quadrant, it has to leave the algae and amoebas behind. Moral: a dynamical system which constantly applies one and the same multiplier cannot be expected to model reality beyond certain limits. More satisfying models for the present "predator-prey" scenario are available, but they are *non-linear*. Linear systems with a better fit to reality — but requiring lengthier explanations — will be described in Section 3 below.

§2. Possible Patterns

We wish to survey the possible shapes of orbits produced by the repeated action of an invertible 2×2 matrix A on points of the plane. Implicitly, we are always dealing with A and A^{-1} at the same time: the reverse transitions from X_n to X_{n-1} are effected by A^{-1}, and of course X_{-n} equals $A^{-n}X_0$. So remember:

A and A^{-1} drive the same orbits but move in opposite directions.

In this context it also helps to keep in mind that A and A^{-1} have the same eigenlines but reciprocal eigenvalues.

As well as going backwards and forwards, we can do double steps by using A^2. Each A-orbit A^nX_0 consists of two A^2-orbits, namely $A^{2m}X_0$ and $A^{2m}X_1$, which it visits alternately. Even if A has a negative eigenvalue, A^2 certainly will not. Hence we need no special treatise on *negative eigenvalues. From the following survey they will be excluded.*

Case (a). In the diagonalizable case, most dynamical systems fall into one of two fundamentally different patterns. The first of these occurs when the two eigenvalues are on different sides of 1, say $\lambda > 1$ and $\mu < 1$, as in the example of the last section. This pattern is known as a *saddle* — because its orbits resemble a topographical map of the ridge (called a saddle) between

two mountain peaks. We have seen that a typical orbit $\{X_n\}$ eventually (i.e., for large $n > 0$) lines up with the eigenline for λ, but we have neglected to say what happnes for $X_{-n} = A^{-n}X_0$. That is easy: A^{-1} has the same eigenlines, but its eigenvalues are $\lambda^{-1} < 1$ and $\mu^{-1} > 1$, the latter now being dominant. Hence X_{-n} lines up with the eigenline for μ.

The second major pattern — depicted on the left of Fig. 3.3 — is called a (proper) *node* and occurs when the two eigenvalues are on the same side of 1. If $1 < \mu < \lambda$, for instance, we still have $\mu = s\lambda$ with $s < 1$, and the argument given in the last section still leads to the same conclusion: orbits will bend in the direction of the eigenline for λ. This time, however, they do not actually approach it, since μ^n just grows more slowly than λ^n without actually shrinking. If $\mu < \lambda < 1$, on the other hand, μ^n shrinks faster than λ^n, so that orbits outside of eigenlines slide into the origin tangent to the eigenline for λ. Figure 3.3 shows both the outward bend and the inward slide at either extreme of the typical orbit: in one direction it is driven by A and in the other by A^{-1} — with eigenvalues on opposite sides of 1.

FIGURE. 3.3. Proper and improper nodes

The transition matrices driving these two dynamical systems are

$$\begin{bmatrix} 1.17 & -0.06 \\ 0.04 & 1.03 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1.35 & -0.25 \\ 0.25 & 0.85 \end{bmatrix}$$

respectively. The first of them has the same eigenlines $(3t, t)$ and $(t, 2t)$ as the example of Section 1, but with eigenvalues $\lambda = 1.15$ and $\mu = 1.05$ both > 1 — hence with outward moving orbits. The second one has the single eigenline (t, t) with eigenvalue 1.1. It is therefore of Type (b) — to be dicussed presently. But first a few words about the disposition of the eigenlines in Case (a).

Eigenslopes. It is easy to write down a formula for a 2×2 matrix A with given eigenvalues $\lambda > \mu$ and eigenlines having slopes s and t, respectively: $A = MD(\lambda, \mu)M^{-1}$ where the slopes appear in the second row of M — explicitly:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \frac{1}{t-s} \begin{bmatrix} 1 & 1 \\ s & t \end{bmatrix} \begin{bmatrix} \lambda & 0 \\ 0 & \mu \end{bmatrix} \begin{bmatrix} t & -1 \\ -s & 1 \end{bmatrix}.$$

Working out the product on the right, you get a remarkable result, namely

$$b = \frac{\lambda - \mu}{s - t} \qquad \text{and} \qquad c = -st \cdot b$$

for the off-diagonal entries of A. It shows that b and c have the same sign if and only if s and t have opposite signs. Moreover, b is positive if and only if the larger slope goes with the larger eigenvalue. These observations are particularly useful when b and c do have the same sign — which can, of course, happen only in Case (a).

Case (b). The second diagram of Fig.3.3 is an *improper node* generated by a matrix with a single eigenline. However, you would be quite right in wondering if it might not just be a proper node with the two eigenlines too close together to be distinguished by the naked eye. In fact, all properties of the improper node could be derived from Case (a) along these lines — but a tidier argument is obtained by starting from scratch.

Consider a 2×2 matrix A with a single eigenline and eigenvalue $\lambda = r = (1/2)\,\mathrm{tr}\,A$. In Section 2.2, we saw that the standard form B can be obtained as $M^{-1}AM$, where $M = [V, W]$ — using any W outside the eigenline and setting $V = (A - rI)W$. To explore the orbit $X_n = A^n X_0$, we might as well take $W = X_0$. Reading the second column of the similarity relation $A^n M = MB^n$ and recalling the formula for B^n in Section 1, we get

$$A^n W = nr^{n-1}V + r^n W = nr^{n-1}\big(V + \frac{r}{n}W\big) \tag{3.4}$$

for the point X_n. It therefore lies in the direction indicated by the vector $V + (r/n)W$. As n grows and grows, $(r/n)W$ shrinks and shrinks — so that this direction becomes more and more that of V itself, i.e., that of the eigenline. Of course, there is no strictly larger eigenvalue this time, but with a less literal meaning of the word "dominant", the slogan of Section 1 may be allowed to stand.

For $r < 1$, the summand $r^n W$ will again shrink toward 0 faster than $nr^{n-1}V$, and make the orbit slide into the origin along the eigenline. In the opposite case $r > 1$, we again get the outward bend — just as for the proper node.

Conspicuously absent from this discussion has been the possibility of *stationary points*, where $AX = X$ for some $X \neq 0$, i.e., where 1 is an eigenvalue. It was not excluded from our corroborations of the "slogan" — which therefore survives — but it was omitted from our diagrams and their descriptions. Figure 3.4 shows the kind of patterns which occur in Cases (a) and (b), respectively.

FIGURE. 3.4. Dynamical systems of Types (a) and (b) with stationary points.

The transition matrices driving these two strange looking systems are

$$\begin{bmatrix} 1.12 & -0.06 \\ 0.04 & 0.98 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1.25 & -0.25 \\ 0.25 & 0.75 \end{bmatrix}$$

respectively. The first of them has the same eigenlines $(3t, t)$ and $(t, 2t)$ as the proper node in Fig.3.3, but with eigenvalues $\lambda = 1.1$ and $\mu = 1$, hence with orbits parallel to the λ-eigenline moving away from the line of stationary points. The second one has the single eigenline (t, t) with the eigenvalue 1. Its matrix is essentially a shear parallel to the eigenline. Each orbit therefore runs with steps of equal size — smaller ones close to the stationary line, bigger ones farther away from it.

Despite their peculiar appearance, the patterns of Fig.3.4 can be regarded as the ancestors of all the saddles and nodes considered previously. In fact, any A of Type (a) or (b) can be scaled to an $A^\circ = (1/\lambda)A$ which has 1 for an eigenvalue. The orbits of A can then be inferred from those of A° in much the same way as is done below for matrices with no eigenlines.

Case (c). To understand what happens here, we go back to the scenario of Chapter 2, where an Assistant observes our orbits as encoded by a matrix L: for every X, he sees a $Y = LX$. In terms of the inverse $M = L^{-1}$, this means that an MY is substituted for every X.

FIGURE. 3.5. Systems of Type (c): spirals and closed orbits.

Thus, $X_n = AX_{n-1}$ turns into $MY_n = AMY_{n-1}$. Left multiplication by M^{-1} then produces $Y_n = (M^{-1}AM)Y_{n-1} = BY_{n-1}$, whence $Y_n = B^nY_0$ for $B = M^{-1}AM$. Moral:

The A-orbit $X_n = A^nX_0$ is the M-image of the B-orbit $Y_n = B^nY_0$.

For clarity, it is best to imagine the points $X_n = (x_n, y_n)$ and $Y_n = (u_n, v_n)$ plotted in two different planes, the x, y-plane and the u, v-plane. With B chosen conveniently, we first examine the B-orbits Y_n in the Assistant's u, v-plane and later transfer them back to the x, y-plane via M . Incidentally — it is instructive to look back at Cases (a) and (b) through the eyes of the Assistant and confirm what we already know.

But what about Case (c)? Remember that $B = \rho R(\theta)$, so that $Y_n = \rho^n R(n\theta)Y_0$. If $\rho = 1$, this moves around the origin on a circle, as shown in Fig.3.1. If $\rho > 1$, it moreover expands by the factor ρ each time it turns through the angle θ, thus moving on an expanding spiral in the u, v-plane. Analogously, $\rho < 1$ gives a contracting spiral. Finally, M transforms the circles of the u, v-plane into ellipses of the x, y-plane and deforms the spirals accordingly. Therefore, if $\det A = \rho^2$ equals 1, the orbit X_n moves on some kind of ellipse. If $\det A > 1$, it spirals outward; if $\det A < 1$, inward.

Figure 3.5 shows orbits belonging, respectively, to the matrices

$$A = \begin{bmatrix} 1.07 & -0.13 \\ 0.23 & 0.87 \end{bmatrix} \qquad \text{and} \qquad A^\circ = \begin{bmatrix} 1.09 & -0.13 \\ 0.23 & 0.89 \end{bmatrix}$$

with $\rho \approx 0.98$ and $\rho^\circ = 1$. Moreover, A is approximately equal to ρA°.

Without eigenlines for guidance, orbits of Type (c) can be tricky to sketch. For a first glimpse, it is advisable to look at the columns of the transition matrix A, and remember that they represent the "next step" in the orbits starting at $(1,0)$ and $(0,1)$, respectively. Other easy initial points, with x_0 and y_0 taken from 0 and ± 1, may also be considered. By the way — this technique is helpful in Cases (a) and (b), too.

Next, we represent the elliptical orbits of $A^\circ = (1/\rho)A$ by drawing the M-image of the unit circle. This can be done elegantly from the "polar decomposition" of M coming up in Chapter 5. For now, we do the best we can by plotting M-images of suitable points on the unit circle, again starting with the coordinates 0, ± 1. With this basic ellipse in place, the shape of the spiral orbits of A may be inferred from the sizes of ρ and θ.

To allay the scruples of sceptics, we hasten to admit that the word "ellipse" has so far been used informally. In Chapter 5 its technical meaning will be justified as well.

§3. Linear Difference Equations

Difference equations are the country cousins of differential equations. If you are willing and able to study the latter, you should go right to the next chapter and glance back at the present one only as the need arises.

A difference equation $X_{n+1} - X_n = F(X_n)$ tells you how to get to the "next state" X_{n+1} from the "present state" X_n in terms of an explicit formula $F(X_n)$. If the latter is linear, i.e., $F(X) = PX$ for some matrix P, you are looking at a *linear* difference equation. In that case, you might as well read it as $X_{n+1} = (I + P)X_n$ and view the whole thing as a discrete dynamical system with the transition matrix $A = I + P$.

Mathematically the two notions are identical. The main reason for the double terminology is the formal resemblance between differences and differentials (cf. Chapter 4). In this vein, it is also customary to write ΔX_n instead of $X_{n+1} - X_n$. Thus $\Delta X_n = PX_n$ means the same as $X_{n+1} = AX_n$.

Usually P is a smallish matrix, so that the step from X_n to X_{n+1} is not too large. As long as all the entries of P are $< 1/2$ in absolute value, it is easy to check that both trace and determinant of $A = I + P$ are positive, whence A is invertible and has no negative eigenvalue.

In the sampling of applications which makes up the bulk of this section, we shall come across some "inhomogeneous" equations of the form

$$X_{n+1} = AX_n + C \qquad \text{or} \qquad \Delta X_n = PX_n + C \tag{3.5}$$

involving a constant vector C. If $C = -PK$ for some other constant K (which means $K = -P^{-1}C$ for invertible P) we get $PX + C = P(X - K)$. Equation (3.5) can therefore be rewritten as

$$\Delta \tilde{X}_n = P\tilde{X}_n \qquad \text{with} \quad \tilde{X} = X - K \tag{3.6}$$

and treated as before. The point $X = K$ (which means $\tilde{X} = 0$) is an "equilibrium", to which the orbits relate as they normally do to the origin.

We shall also encounter systems masquerading as "second order" *scalar* equations

$$x_{n+1} = ax_n + fx_{n-1} + g \qquad \text{with} \quad f \neq 0, \tag{3.7}$$

producing a sequence $x_0, x_1, \ldots$ of numbers. If we define $y_{n+1} = cx_n$ for some $c \neq 0$, this says exactly the same as the matrix equation

$$\begin{bmatrix} x_{n+1} \\ \\ y_{n+1} \end{bmatrix} = \begin{bmatrix} a & b \\ \\ c & 0 \end{bmatrix} \begin{bmatrix} x_n \\ \\ y_n \end{bmatrix} + \begin{bmatrix} g \\ \\ 0 \end{bmatrix}, \quad \text{with} \quad bc = f\,, \tag{3.8}$$

which is just a special case of Eq.(3.5) or — in its explicit version — Eq.(3.9) below. If $a + f \neq 1$, the corresponding $P = A - I$ is invertible, and the equilibrium value of x is easily calculated as $g/(1 - a - f)$.

Some Typical Applications

Discrete dynamical systems are most common in the social sciences and biology — where the observables tend to obey the pulse of natural rythms or seasonal cycles instead of varying smoothly.

In the following examples, the letters a, b, c, d, g, h will always refer to the parameters of the standard equation

$$\begin{bmatrix} x_{n+1} \\ \\ y_{n+1} \end{bmatrix} = \begin{bmatrix} a & b \\ \\ c & d \end{bmatrix} \begin{bmatrix} x_n \\ \\ y_n \end{bmatrix} + \begin{bmatrix} g \\ \\ h \end{bmatrix}, \tag{3.9}$$

while A, P as well as C, K are to be read as in the equivalent matrix formulas Eq.(3.5) and Eq.(3.6).

In the first five applications, A will be diagonalizable. Apart from the remarks about "eigenslopes" in Section 2, a couple of pointers about the eigenvalues λ and μ will be useful: they are both positive whenever both $\det A = \lambda\mu$ and $\operatorname{tr} A = \lambda + \mu$ are positive; they lie on opposite sides of 1, if and only if $\det P = \det(A - I) = (\lambda - 1)(\mu - 1)$ is negative. In all but one of the examples (the fourth), it will be assumed that P is invertible.

1. Symbiosis. All four entries of A are positive but less than 1. Moreover $\det A > 0$, i.e., a, d are collectively greater than b, c. Positivity of the interaction coefficients b, c means that the quantities x, y mutually enhance each other — like fungi and cyanobacteria living together in lichens — whence the term "symbiosis". The fact that the diagonal entries a, d are smaller than 1 means that x or y, each left to itself, would decay exponentially. It also implies $\operatorname{tr} P < 0$.

To begin with, let us stay with the "homogeneous" equation, where $g = h = 0$. We distinguish two cases.

(i) $\det P < 0$. The eigenvalues of A form the pattern $0 < \mu < 1 < \lambda$. Since b and c are positive, the eigenslopes have opposite signs with that for λ being positive. Thus we get a saddle pattern: all orbits seem to come from near the μ-eigenline, move toward the origin, and then veer off to follow the lure of the λ-eigenline, as can be seen on the left of Fig.3.6. The biggest difference between this and Fig.3.2 is that the μ-eigenline does not belong to the "real world"of the first quadrant. From anywhere in the latter, an orbit will eventually approximate the ratio $y : x$ typical for λ-eigenvectors, with exponential growth to the tune of λ.

(ii) $\det P > 0$. Since $\operatorname{tr} P < 0$, the eigenvalues of P must both be negative, so we have $0 < \mu < \lambda < 1$. The position of the eigenlines is as above (for the same reasons) but now we get a proper node, with orbits in the first quadrant vaulting toward the λ-eigenline and sliding to oblivion as dictated by λ.

What about an inhomogeneous equation with $C \neq 0$ and equilibrium $K = -P^{-1}C$? Since all entries of the adjoint $P^\star$ are negative, those of $-P^{-1}$ are positive if and only if $\det P > 0$. Hence in Case (ii) with non-negative components of C, we have K in the first quadrant with orbits wandering toward it in node fashion, as shown on the right of Fig.3.6. In Case (i), the position of K is of minor interest: wherever it is, all orbits — except those which step exactly along μ-eigenvectors — will eventually be swept away in the direction of the dominant eigenline.

This kind of model was used by the British physicist and Quaker L.F.Richardson in several works about the nature of arms races and the causes of war. His mutually enhancing quantities x and y are the military budgets of two opposing alliances. The off-diagonal entries b, c represent the "defensive" terms (i.e., keeping up with the enemy), the diagonal a, d the tendency toward "fatigue" (i.e., pressure for budget cuts), and the constants g, h the requirements of maintaining the military.

2. Conflict. Most things remain as in (1.) above, except that now $b = -\beta$ and $c = -\gamma$ are both negative — meaning that the quantities x, y inhibit each other, like two populations competing for the same habitat.

The easiest way to describe A is to say that its adjoint $A^\star$ looks like a transition matrix for "symbiosis". It therefore has the same trace and determinant — hence poduces the same eigenvalues: (i) $0 < \mu < 1 < \lambda$ if $\det P < 0$, and (ii) $0 < \mu < \lambda < 1$ if $\det P > 0$. The main difference between the present scenario and the previous one is the *location* of the eigenlines. This time (since $b < 0$) it is the one for μ that crosses the "real world", i.e., the first quadrant.

The saddle orbits of Case (i) are similar to the ones in the preceding model — except that they run toward the origin instead of away from it. The nodal orbits of Case (ii), on the other hand, appear like reflections in the y-axis of their counterparts in the homogeneous version of (1.). Either way, unless they carefully step along the μ-eigenline, all orbits from the first

FIGURE 3.6. A homogeneous saddle and an inhomogeneous node, both restricted to the first quadrant.

quadrant exit the real world through one of the coordinate axes: one of the competing quantities x or y will be annihilated.

We have tacitly assumed $g = h = 0$. In fact, a general statement about an equilibrium $K \neq 0$ would be difficult to make, because $P^\star$ now has positive as well as negative entries. If a and d are very close to 1, the entries of $-P^{-1}$ are overwhelmingly positive, and a C with positive components can be expected to produce a K in the first quadrant. However, this also means $\det P < 0$, and the equilibrium (in a saddle context) would be of no practical interest.

This model, too, has military applications — the best known being due to the British engineer F.W.Lanchester, who used variants of it in his sundry attempts to understand modern warfare. Here x, y represent the numerical strengths of two armies, a, d the effects of self-inflicted "operational losses", β, γ the respective enemy's "fire power", and g, h (if present) regular supplies of fresh recruits.

His simplest version has no recruits and no operational losses, so that $a = d = 1$. This borderline case was omitted above for uniformity of presentation: it works without a hitch. The eigenvalues of A are $1 \pm \sqrt{\beta\gamma}$, and the μ-eigenline is characterized by the ratio $x : y = \sqrt{\beta} : \sqrt{\gamma}$. To avoid the premature annihilation of one of the armies, the *squares* of their numbers must therefore be in the same ratio as the fire powers of their respective enemies. This is called "Lanchester's Law of Squares".

3. Predation. With reference to the set-up in (1.), the present one could be described by saying that the first row of P was multiplied by -1. Thus we have $a > 1$, $b = -\beta$ with $0 < \beta < 1$, and no changes in the second row of A. Now $\det A = ad + \beta c$ is automatically positive, but we add a new requirement, namely $\det P < 0$.

This condition excludes the possibility that P (and hence A) might fall under Cases (b) or (c) — where determinants are never negative. The only possible pattern is therefore a saddle, with the dominant eigenline having the smaller slope since $b < 0$. As $c > 0$, the eigenlines will not lie in separate quadrants. But could their slopes be negative? No: if you go back to the remark on eigenslopes in Section 2 and continue the computation, you will find

$$(a-1) = \frac{(\lambda-1)t + (1-\mu)s}{(t-s)}\,.$$

Since all the bracketed terms are positive, so are the slopes s and t.

In short, the system looks very much like the algae-amoeba story which began this chapter and whose orbits are depicted in Fig.3.2. We have already admitted the weaknesses of this model, and shall endeavour to produce a better (non-linear) one in the upcoming appendix. However, if you wish to emulate Lanchester and Richardson, you might contemplate its adaptation to a martial setting — say, a population occupied by an oppressive army.

4. Alternation. In many ways this appears like a special case of (1.) with $g = h = 0$ and $a + c = b + d = 1$. However, $a - 1 = -c$ and $d - 1 = -b$ implies $\det P = 0$. Consequently P has the eigenvalue 0, whence A has the eigenvalues $\lambda = 1$ and $\mu = a + d - 1$, with $-1 < \mu < 1$. Note: $\det A = \mu = d - c = a - b$ is not necessarily positive here.

One easily checks that the eigenlines pass through the points (b, c) and $(-1, 1)$ respectively — the former consisting entirely of stationary points. Orbits therefore take the strange degenerate form shown on the left of Fig.3.4. If $\mu > 1$, they march straight toward the stationary line in single file; if $\mu = 0$, they jump onto it immediately; if $\mu < 1$, each of them splits into the A^2-orbits A^2X_0 and A^2X_1, which advance toward it from opposite directions. Equation (3.3) with $\lambda = 1$ and $s = \mu$ corroborates these statements.

For a story-line, go back to the soft-drinkers of Exercise 10 in Chapter 2. If their brand loyalty changed every year with the same regularity, it could be modelled by such a system. Whatever their initial market shares may have been, Coke and Pepsi would in the long run approach the eigenratio 3:2. This is the very simplest instance of a "finite Markov Chain", so named after the Russian mathematician A.A.Markov. More generally, such a chain would involve m "states" (in our case $m = 2$) and given probabilities p_{ij} of switching from State i to State j, in other words, an $m \times m$ matrix with non-negative entries and column-sums = 1.

5. Reproduction. In many biological species, the young have to wait out one reproductive season before taking part in the mating game. Thus a population will consist of x_n adults and y_n juveniles in the n-th season. The adults are made up of the surviving fractions ax_{n-1} and by_{n-1} of last season's oldsters and youngsters. The number of juveniles is proportional

to the previous number of adults, say $y_n = cx_{n-1}$. Altogether this yields Eq.(3.8) with $g = 0$ and a, b, c positive.

Since $d = 0$, the eigenvalues have the form $\lambda = \frac{1}{2}\big(a + \sqrt{a^2 + 4bc}\big)$ and $\mu = a - \lambda$. Hence $-\lambda < \mu < 0 < \lambda$, and μ has absolute value smaller than λ. Eigenvectors are easy to find: we have

$$A - \lambda I = \begin{bmatrix} \mu & b \\ c & -\lambda \end{bmatrix}$$

and a similar formula for $A - \mu I$ with λ and μ interchanged — whence $[\lambda, c]$ and $[\mu, c]$ are eigenvectors for λ and μ, respectively. Only the λ-eigenline runs through the first quadrant.

All A-orbits in the first quadrant are attracted to this line, hence the quotients $x_n/y_n = x_n/cx_{n-1}$ will eventually approximate λ/c. This is particularly striking when viewed as a property of the sequence $x_0, x_1, x_2, \ldots$ produced by the second order equation $x_{n+1} = ax_n + bcx_{n-1}$ à la Eq.(3.7): the quotients x_n/x_{n-1} of successive terms converge to λ. The first example of this phenomenon was found 800 years ago by the Italian mathematician Fibonacci. Using the equation $x_{n+1} = x_n + x_{n-1}$ and the initial values $x_0 = 0$, $x_1 = 1$, he obtained the famous sequence

$$0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, \ldots$$

whose successive quotients approximate $\lambda = \frac{1}{2}(1 + \sqrt{5})$. His motivation fits well into our present framework: the sequence, he says, counts the total number of adult female rabbits in a certain imaginary breeding scheme.

Coming back to the two-dimensional representation, we have to deal with the question of growth ($\lambda > 1$?) and the exact shape of the orbits. As $\mu - 1$ is certainly negative, the positivity of $\lambda - 1$ is equivalent to the negativity of $(\lambda - 1)(\mu - 1) = \det P = (1 - a) - bc$. Hence we see growth if and only if $(1 - a) < bc$, i.e., adult mortality is overcome by juvenile survival and adult fertility. Otherwise there is stagnation or decay.

Because of the negativity of μ, the A-orbits shuttle back and forth between the A^2-orbits A^2X_0 and A^2X_1, as they do in (4.). Since A^2 has eigenvalues $\lambda^2 > \mu^2 > 0$, its orbits form a node or saddle pattern depending on whether $\det(A^2 - I) = \det(A - I)\det(A + I) = (1 - a - bc)(1 + a - bc)$ is positive or negative. The latter, shown in Fig.3.7, happens whenever these two factors differ in sign, that is:

$$1 - a < bc < 1 + a\,.$$

If bc lies below or above this interval, we get (respectively) a contracting or expanding node.

6. Spending. In a classical paper, the American economist P. Samuelson suggested that "national income" x be considered as having three sources:

FIGURE 3.7. Fibonacci's system has a negative eigenvalue. Alternate points lie on the saddle orbits of A^2. Successive points have been joined by lines.

consumer (y), investment (z), and government (g) spending, so $x = y+z+g$. These quantities are subjected to three assumptions: in a given quarter, consumers spend a certain fraction c of the previous quarter's income, so $y_n = cx_{n-1}$; investment is proportional to the *difference* between consumer spending this quarter and last, so $z_n = \beta(y_n - y_{n-1}) = \beta c x_{n-1} - \beta y_{n-1}$; government spending is constant. Putting it all together, we get

$$\begin{bmatrix} x_n \\ \\ y_n \end{bmatrix} = \begin{bmatrix} c(1+\beta) & -\beta \\ \\ c & 0 \end{bmatrix} \begin{bmatrix} x_{n-1} \\ \\ y_{n-1} \end{bmatrix} + \begin{bmatrix} g \\ \\ 0 \end{bmatrix},$$

i.e., Eq.(3.8) with $b = -\beta$ and $a = c(1+\beta)$. Since c and β are positive but less than 1, so is $\det A = c\beta$.

Keeping in mind that $c^2(1+\beta)^2 - 4c\beta < c^2(1+\beta)^2 - 4c^2\beta = c^2(1-\beta)^2$, it is easy to see that the eigenvalues (if any) of A lie between $c\beta$ and c. Hence the orbits are spirals or nodes converging to the equilibrium $x = g/(1-c)$, $y = cx$. Spirals occur when consumer spending is sluggish, in the sense that

$$c < \frac{4\beta}{(1+\beta)^2}.$$

Of course, the whole story could be presented as a second order equation for x alone in the style of Eq.(3.7), namely $x_{n+1} = c(1+\beta)x_n - c\beta x_{n-1} + g$.

7. Respiration. For a slightly different example, we look at the ventilation of carbon dioxide in the blood. The simplest model assumes that the volume (y) of each breath is proportional to the blood CO_2-level (x) during the previous breath, so that $y_n = cx_{n-1}$. Exhalation reduces this level by an amount βy_{n-1} but metabolic processes augment it again by a constant

FIGURE 3.8. The linear model of CO_2 ventilation is almost periodic with period 6. Successive points have been joined by lines to show the pattern.

g. For the current CO_2-level, we therefore get $x_n = x_{n-1} - \beta y_{n-1} + g$. Altogether, we have

$$\begin{bmatrix} x_n \\ \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & -\beta \\ \\ c & 0 \end{bmatrix} \begin{bmatrix} x_{n-1} \\ \\ y_{n-1} \end{bmatrix} + \begin{bmatrix} g \\ \\ 0 \end{bmatrix},$$

i.e., Eq.(3.8) with $b = -\beta$ and $a = 1$. Since c and β are positive, so is $\det A = \beta c$. Eigenvalues (if any) are positive and add up to $\operatorname{tr} A = 1$. This eliminates saddles and expanding nodes, leaving spirals and contracting nodes as the only possibilities. The equilibrium point $x = g/\beta c$, $y = cx$ is always in the first quadrant.

The main difference between this example and the last one is that expanding spirals are now possible, because $\det A = \beta c$ is not constrained to be less than 1. A minor difference is that x and y are not necessarily measured in the same units, the choice of which influences the numerical values of β and c. On the other hand, βc does have an intrinsic meaning: since $\beta y_n = \beta c x_{n-1}$, it represents the amount of CO_2 exhaled for every unit of CO_2 present in the blood during the previous breath — independently of the units used. For abbreviation, let us write $\beta c = \delta$.

Spirals occur when $1 - 4\delta$ is negative, i.e. when δ is greater than 0.25, they are contracting when $\det A = \delta$ is less than 1. For δ outside the interval

$$0.25 < \delta < 1$$

of contracting spirals, we even get contracting nodes or expanding spirals, depending on whether it lies below 0.25 or above 1.

§4. Examples and Exercises

Worked Examples

The first three examples will again deal with the matrices introduced at the beginning of Section 2.4, namely

$$A_1 = \begin{bmatrix} 6 & -2 \\ 2 & 1 \end{bmatrix} \qquad A_2 = \begin{bmatrix} 5 & -2 \\ 2 & 1 \end{bmatrix} \qquad A_3 = \begin{bmatrix} 5 & -3 \\ 3 & 1 \end{bmatrix} \quad ,$$

but this time, we look at them in reverse order. We shall use some of the information already gleaned about them in Chapter 2. When we write "Example 2.x", we mean Example x from that chapter.

1. For $A = A_3$, compute A^8.

We have $A = MBM^{-1}$ and therefore $A^8 = MB^8M^{-1}$, where M and B are as in Example 2.3, that is::

$$M = \begin{bmatrix} 2 & -\sqrt{5} \\ 3 & 0 \end{bmatrix} \qquad \text{and} \qquad B = \begin{bmatrix} 3 & -\sqrt{5} \\ \sqrt{5} & 3 \end{bmatrix} \quad .$$

To raise B to a power, we first write it in the form of a scaled rotation matrix $\rho \cdot R(\theta)$. Converting the first column of B to polar coordinates, we get $(3, \sqrt{5}) = \sqrt{14} \cdot (\cos\theta, \sin\theta)$ with $\theta = \arctan(\sqrt{5}/3) = 36.7°$. Therefore $B = \sqrt{14} \cdot R(\theta)$, and $B^8 = (14)^4 \cdot R(293.6°)$. The cosine and sine of $293.6°$ give 0.400 and -0.916, respectively. Multiplied by $(14)^4 = 38416$, these become the entries of the first column of B^8, namely 15366 and -35189 approximately. Finally we get (again approximately):

$$A^8 = M \begin{bmatrix} 15366 & 35189 \\ -35189 & 15366 \end{bmatrix} M^{-1} = \begin{bmatrix} -16108 & 47211 \\ -47211 & 46840 \end{bmatrix} \quad .$$

This calculation is most easily done by writing $B^8 = \alpha I + \beta J$, where $\alpha = 15366$, $\beta = -35189$, and $J = R(90°)$. Then $MB^8M^{-1} = \alpha I + \beta MJM^{-1}$, so the only serious matrix multiplication is the easy MJM^{-1}. Incidentally, the exact entries of A^8 are: -16112, ± 47232, 46864.

Note: A^8 could have been computed by successive squaring as $\big((A^2)^2\big)^2$, with an assist from Cayley-Hamilton. You should try this as a check on our result. The advantage of the present method is that it requires no more work for the 88-th (or higher) power than it does for the 8-th.

2. Find a matrix H such that $\operatorname{tr} H > 0$ and $H^2 = A_2$.

Again we convert to standard form: $A_2 = MBM^{-1}$, with M and B as in Example 2.2. In fact, $B = 3 \cdot G_{12}(1/3) = K^2$, where $K = \sqrt{3} \cdot G_{12}(1/6)$,

because $G_{12}(a)^2 = G_{12}(2a)$ for any a. Therefore, if we set $H = MKM^{-1}$, we get $H^2 = A_2$. In detail this means

$$H = \sqrt{3}\begin{bmatrix} 2 & 1 \\ 2 & 0 \end{bmatrix}\begin{bmatrix} 1 & 1/6 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 0 & 1/2 \\ 1 & -1 \end{bmatrix} = \frac{1}{\sqrt{3}}\begin{bmatrix} 4 & -1 \\ 1 & 2 \end{bmatrix} \quad .$$

Note: if we had chosen the negative square root of 3, we would have come out with $-H$; of course, the square of $-H$ is also A_2, but the trace would then be negative. Incidentally, in Cases (b) and (c), the square root (if any) of a matrix is unique up to sign, while in the diagonalizable case there are typically four possibilities. Can you explain this?

3. For every real number z, define a matrix $A(z)$ in such a way that $A(n) = A_1^n$ for all integer n.

As shown in Example 2.1, the matrix A_1 is similar to the diagonal matrix $B = D(5, 2)$. For the latter it is easy to define $B(z) = D(5^z, 2^z)$ and get $B(n) = B^n$ for integer n. Again we use similarity: since $A_1^n = MB^nM^{-1}$, we get what we want by setting $A(z) = MB(z)M^{-1}$. Explicitly, we then have $A(z) =$

$$\frac{1}{3}\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} 5^z & 0 \\ 0 & 2^z \end{bmatrix}\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 4f^z - t^z & 2t^z - 2f^z \\ 2f^z - 2t^z & 4t^z - f^z \end{bmatrix}$$

writing $5 = f$ and $2 = t$ to avoid confusion.

Note: As long as we stay away from matrices with negative eigenvalues, we can always construct such a "continuous interpolation" $A(z) = MB(z)M^{-1}$. In Cases (b) and (c), we need to use not only α^z (for $\alpha > 0$) but also $G_{12}(za)$ and $R(z\theta)$. Thus, each of the discrete orbits A^nX_0 described in Section 2 above lies on the continuous curve $A(z)X_0$.

4. In Case (c), every spiral orbit of a matrix A expands or contracts by a certain constant c_A between any two crossings of the x-axis, the y-axis, or any other line through 0. Find c_A for

$$A = \begin{bmatrix} 1.07 & -0.13 \\ 0.23 & 0.87 \end{bmatrix} \quad .$$

First we find a positive number ρ and an angle θ such that $B = \rho R(\theta)$ is similar to A. Since $\det B = \rho^2$ and $\operatorname{tr} B = 2\rho\cos\theta$, and since similar matrices have the same determinant and trace, the former equals 0.9608 and the latter 1.94. This makes $\rho = 0.9802$ and $\theta = 8.273°$ approximately.

Now, every B-orbit contracts by the factor ρ at every step, and there are $180/8.273 \approx 22$ steps in every half-turn. Therefore B^{22} is approximately

$-\rho^{22}I$ *and so is* $A^{22} = MB^{22}M^{-1}$. *Hence* $c_A = \rho^{22} = (0.9802)^{22} = 0.644$. *Take a ruler and check this on Fig.3.5.*

5. For the matrix A given above, find the direction in which orbits cross (a) the positive x-axis, (b) the negative y-axis; determine the half-line which is crossed (c) horizontally from left to right, (d) vertically upward.

An orbit passing through X *obviously evolves in the direction* $AX - X = (A - I)X$. *Note that*

$$A - I = \frac{1}{100}\begin{bmatrix} 7 & -13 \\ 23 & -13 \end{bmatrix} \quad \textit{and} \quad (A-I)^{-1} = \frac{125}{2600}\begin{bmatrix} -13 & 13 \\ -23 & 7 \end{bmatrix} \quad .$$

Since we are only concerned with the direction (not the magnitude) of the steps, we can ignore the positive scalars adorning these matrices.

For the positive x*-axis, we may take* $X = (1, 0)$ *and (a) get the direction* $(7, 23)$. *For the negative* y*-axis, take* $X = (0, -1)$ *and (b) get* $(1, 1)$. *To obtain outcomes like* $(1, 0)$ *or* $(0, 1)$, *we need to start with an* X *such that* $(A - I)X$ *has these coordinates. In other words, we may take (c) the first and (d) the second column of* $(A - I)^{-1}$, *or any positive multiple of these; i.e., anything on the half-lines consisting of positive multiples of* $(-13, -23)$ *and* $(13, 7)$ *respectively. Try to recognize these directions in Fig.3.5.*

Exercises

Most of these exercises refer to one of the following nine matrices:

$$A_1 = \begin{bmatrix} 1 & 0.1 \\ 0.1 & 1 \end{bmatrix} \qquad A_2 = \begin{bmatrix} 1.14 & -0.12 \\ 0.08 & 0.86 \end{bmatrix} \qquad A_3 = \begin{bmatrix} 0.92 & -0.06 \\ 0.04 & 0.78 \end{bmatrix}$$

$$A_4 = \begin{bmatrix} 0.85 & 0.05 \\ 0.05 & 0.85 \end{bmatrix} \qquad A_5 = \begin{bmatrix} 0.2 & 1.8 \\ -0.2 & 1.4 \end{bmatrix} \qquad A_6 = \begin{bmatrix} 1.3 & -0.5 \\ 0.5 & 0.3 \end{bmatrix}$$

$$A_7 = \begin{bmatrix} 0.83 & -0.35 \\ 0.35 & 0.85 \end{bmatrix} \qquad A_8 = \begin{bmatrix} 1.18 & -0.7 \\ 0.35 & 0.48 \end{bmatrix} \qquad A_9 = \begin{bmatrix} 1.09 & -0.04 \\ -0.03 & 1.05 \end{bmatrix}$$

For exercises (1) – (4), use the approximate values $1.1^{41} \approx 50$, $0.9^{41} \approx 133 \times 10^{-4}$, and $0.8^{41} \approx 10^{-4}$ to compute A^{41}, where A is as below. In each case check whether both columns of A^{41} are approximate eigenvectors of A.

1. $A = A_2$. 2. $A = A_3$. 3. $A = A_6$. 4. $A = 0.9\,R(\pi/8)$.

In exercises (5) – (12), x_n denotes the number of rabbits (in hundred thousands) and y_n denotes the number of foxes (in thousands) on a certain arctic island, in the n-th year after counts were begun. The evolution of these numbers is governed by the equation $X_n = AX_{n-1}$, where X_n is the column (x_n, y_n) and A is as given below. For *plotting*, use graph-paper and a unit length of 2 cm or 1 inch.

5. Compute X_2, X_4, X_8, and X_{16}, if $A = A_2$ and $X_0 = (1, 2.1)$. item 6. (a) Plot X_n for $0 \le n \le 20$, with A and X_0 as above. (b) What has happened to the rabbits when $n = 15$? What to the foxes?
7. Redo (5) for $X_0 = (1.1, 2)$.

8. (a) As (6a), with $X_0 = (1.1, 2)$. (b) What happens for $0 \le n \le 11$, for $n > 11$, in the long run?
9. Compute X_2, X_4, X_8, and X_{16}, if $A = A_3$ and $X_0 = (2, .01)$.
10. (a) As (6a), with A and X_0 as in (9). (b) When does the fox population peak? What is the long term trend?
11. Compute X_2, X_4, X_8, and X_{16}, if $A = A_6$ and $X_0 = (2, .01)$.
12. (a) As (6a), with A and X_0 as in (11). (b) As (10b).

In exercises (13) – (20), identify the kind of dynamical system generated by the matrix A in question. Make a rough sketch of some typical orbits. Show eigenlines (if any); in the case of spirals, try to indicate the orientation of the underlying ellipse.

13. $A = A_1$. 14. $A = A_2$. 15. $A = A_3$. 16. $A = A_4$.
17. $A = A_5$. 18. $A = A_6$. 19. $A = A_7$. 20. $A = A_8$.
21. Sketch some typical orbits of A_9, showing the eigenlines.
22. Using graph paper and a unit length of 5 cm or 2 inches, plot $X_0, \ldots, X_4$ for

$$A = \begin{bmatrix} 0.7 & 0.3 \\ 2 & 0 \end{bmatrix} \qquad X_0 = \begin{bmatrix} 0 \\ 2 \end{bmatrix}.$$

Can you explain the zig-zags?

23. Explain how you can avoid the zig-zags of (22) by using A^2.
24. Two capacitors C_x and C_y are coupled in such a way that their charges x_n and y_n after n microseconds follow the equation $X_n = A_8 X_{n-1}$, where $X_n = (x_n, y_n)$. If $X_0 = (p, p)$ with $p > 0$, how many microseconds will it take before both charges are negative and approximately equal?
25. Let a dynamical system be defined by $\Delta X_n = P X_n$, where the entries of P are < 0.3 in absolute value. Show that the system is a "saddle" if and only if $\det P < 0$.

Appendix: Non-linear Models

The following material should not really be in this book — it will *not* be on the exam and can be skipped with a clear conscience. But truth will out: having so thoroughly decried the shortcomings of linear models, we are now obliged to show how matrices continue to play an important role even in the non-linear ones.

The most general difference equation has the the form $\Delta X_n = F(X_n)$, where F can be just about anything. However, if F is a half-decent function in a neighborhood of a particular point K, differential calculus will yield a constant matrix L_K such that

$$F(X) \approx F(K) + L_K \cdot (X - K)\,, \tag{3.10}$$

as long as $(X - K)$ is small. More precisely, the gap $d_F(X, K)$ between the two sides of Eq.(3.10) is small *relative* to the distance $d(X, K)$ between X and K, in the sense that the ratio $d_F(X, K) : d(X, K)$ tends to zero as X approaches K.

If it happens that $F(K) = 0$, the right hand side of Eq. (3.10) is just a matrix product, and the original difference equation can be approximated by the linear system

$$\Delta\tilde{X}_n = L_K\tilde{X}_n \tag{3.11}$$

for small $\tilde{X} = (X - K)$, i.e., for X near K. Now, $F(K) = 0$ means precisely that K is a stationary (or equilibrium) point, because $X_n = K$ entails $X_{n+1} - X_n = F(K) = 0$ — so the orbit never moves on. Equation (3.11) is called the *linearization* of $\Delta X_n = F(X_n)$ near K.

To give some substance to these fantasies, we shall follow them up for *quadratic* difference equations, which have the form

$$\Delta X_n = D(X_n)\big(C + QX_n\big)\,, \tag{3.12}$$

where $D(X) = D(x, y)$ is the obvious diagonal matrix, and Q is invertible. Except for the extra diagonal factor, Eq.(3.12) is much like Eq.(3.5) and will be treated similarly. Putting $K = -Q^{-1}C$ and $\tilde{X} = X - K$, we obtain $C + QX = Q(X - K) = Q\tilde{X}$ as before.

Substituting $X = K + \tilde{X}$ into Eq.(3.12), we get $\Delta\tilde{X}_n = \Delta X_n = D(K+\tilde{X}_n)Q\tilde{X}_n$. This yields the second of the following equations, the first just being an expanded version — noting that $D(X)C = D(C)X$ — of the original:

$$\begin{aligned}\Delta X_n &= D(C)X_n + D(X_n)QX_n\\ \Delta\tilde{X}_n &= D(K)Q\tilde{X}_n + D(\tilde{X}_n)Q\tilde{X}_n\,.\end{aligned} \tag{3.13}$$

When the components of X are sufficiently small, those of $D(X)QX$ can be counted on to be tiny even in comparison. Hence the first terms on the right hand side of Eq.(3.13) yield the linearizations of Eq.(3.12) at 0 and K, respectively. As dynamical systems, the latter are therefore driven by the matrices $I + D(C)$ and $I + D(K)Q$, respectively.

For the rest of this section, we shall try to get a feeling for what this means in particular settings — first in the one-dimensional case, and later in variants of Applications (1.) and (3.) from Section 3.

One of the more striking weaknesses of linear models is that they sometimes predict unlimited growth. Already the one-dimensional (scalar) equation $\Delta x_n = \alpha x_n$ or $x_{n+1} = (1+\alpha)x_n$ suffers from this defect if $\alpha > 0$. In the context of population growth it is therefore often replaced by the *logistic equation*

$$\Delta x_n = x_n(\alpha - \varphi x_n)\,,$$

in which the term $-\varphi x_n$ (with $\varphi > 0$) puts the brakes on the growth-producing α. As x_n approaches the value $\kappa = \alpha/\varphi$, the difference $\Delta x_n = x_{n+1} - x_n$ tends toward 0, and at the point $x_n = \kappa$ itself there is no growth at all. Clearly, this is a scalar version of Eq.(3.12) — and the counterpart of Eq.(3.13) appears as

$$\begin{aligned}\Delta x_n &= \alpha x_n - \varphi x_n^2\\ \Delta\tilde{x}_n &= -\alpha\tilde{x}_n - \varphi\tilde{x}_n^2\,.\end{aligned}$$

When x is very small, x^2 is tiny even compared to x, and the second term of the first equation can be neglegted. Hence, near $x = 0$, the evolution of x_n is almost exponential: $x_{n+1} \approx (1+\alpha)x_n$. By the same token, near $x = \kappa$ where $\tilde{x}$ is very small, we get $\tilde{x}_{n+1} \approx (1-\alpha)\tilde{x}_n$, an almost exponential decay toward $\tilde{x} = 0$, i.e., $x = \kappa$. Figure 3.9 shows such a sequence for $\alpha = 0.2$ and $\varphi = 0.1$.

FIGURE. 3.9. A sequence of points spawned by the logistic equation.

Turning back to two dimensions, let us compare the straight linear model $\Delta X_n = PX_n$ with the quadratic $\Delta X_n = D(X_n)\big(C + QX_n\big) = D(C)X_n + D(X_n)QX_n$. The latter can be viewed as also being given by a matrix — but a non-constant one, namely $N(X) = D(C) + D(X)Q$. To interpret the meaning of their various entries, it is useful to see these two matrices side by side:

$$P = \begin{bmatrix} \alpha & b \\ \\ c & \delta \end{bmatrix} \qquad N(X) = \begin{bmatrix} \alpha - \varphi x & bx \\ \\ cy & \delta - \psi y \end{bmatrix},$$

where $\alpha = a - 1$ and $\delta = d - 1$, as well as b and c, come from the standard parameters of Eq.(3.9), but $-\varphi$ and $-\psi$ are new ingredients. They are the diagonal entries of the constant matrix Q, the off-diagonal ones being the old b and c.

In the top left corner of $N(X)$, we see the straight growth coefficient α of P replaced by the logistic $(\alpha - \varphi x)$. To the right of this, the interaction term b has changed to bx — meaning that the impact of the y's on the x's is proportional to the numerical strength of the latter. To use the predation metaphor: the amount of prey consumed depends on its availability, not only on the number of predators.

For a numerical example, consider

FIGURE. 3.10. A non-linear predation model.

$$P = \begin{bmatrix} 0.14 & -0.12 \\ \\ 0.08 & -0.14 \end{bmatrix} \qquad N(X) = \begin{bmatrix} 0.14 - 0.01x & -0.12x \\ \\ 0.08y & -0.14 \end{bmatrix}.$$

The linear difference equation $\Delta X_n = PX_n$ produces the dynamical system mentioned in the preamble to this chapter and depicted in Fig.3.2. The non-linear $\Delta X_n = N(X_n)X_n$, which incorporates the logistic correction $\varphi = .01$ (with $\psi = 0$) as well as the internal multipliers x and y, yields a completely different picture. Figure 3.10 shows some of its orbits including the equilibrium point $K = (1.75, 1.02)$. As a model of predator y and prey x, it suggests that these populations would tend to remain constant at K or to oscillate around it in cyclic fashion. Of course, this is an idealization which would never occur exactly like that in reality. The only reason we got close to it here is our cunning choice of φ. For $\varphi = .02$, the orbits would swiftly contract toward the equilibrium instead of cycling.

`Section 4 still needs to be repared and completed.`

This article was processed by the author using the TEX macro package from Springer-Verlag.