CHAPTER 3

Discrete Linear models

3.1. A stratified population model

A single state variable, as we considered in the first two sections, is usually not enough to describe a system. In this section we will look at some simple examples of multi-dimensional dynamical systems.

Here is a population model that takes into consderation the fact that birth and death rates are connected to age. In this example we consider a plant population with the following characteristics: Plants live at most three years. In their first year they do not produce seeds. THos that survive to the second or third year produce seeds, which sprout in the next spring to become first-year plants. Moreover, the death rate varies from year to year, as does the number of seeds produced per plant. To describe this situation we use three states, x_1, x_2, x_3, to record the number of plants in their first, second or third year. We consider this as a discrete dynamical system, in the sense that we count the plants once per year. Of course, we have to assume that there is some way of identifying a plant as first, second or third year.

Suppose that a fraction, s_1, of the first year plants survive to the second year; as noted above, they do not produce seeds. Suppose that s_2 of the second year plants survive to the third year, and that, on average, each second year plant produces f_2 seeds that will germinate next year to form first year plants. Finally, none of the third year plants survive another year, but, on average, each third year plant produces f_3 seeds.

Suppose we have, as indicated above, a state vector $x = (x_1, x_2, x_3)$ which records the counts of the first, second and third year plants, and suppose the next year's counts are (X_1, X_2, X_3). We can calculate these figures

from the rules above. First year plants next year sprout from the seeds produced by second and third year plants this year, so the number of first year plants will be $X_1 = f_2x_2 + f_3x_3$. Second year plants next year are the first year plants from this year that survive; so there will be $X_2 = s_1x_1$ second year plants. Finally, third year plants next year are the second year plants from this year that survive; so there will be $X_3 = s_2x_2$ third year plants. The transition rule can then be expressed as

$F(x) = X$. In vector terms this is

$$F\left(\begin{bmatrix} x_1 \\ x_2 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} f_2x_2 + f_3x_3 \\ s_1x_1 \\ s_2x_2 \end{bmatrix}.$$

This transition rule is a linear function, so it can be represented as a matrix product:

$$F(x) = Lx, \text{ where } L = \begin{bmatrix} 0 & f_2 & f_3 \\ s_1 & 0 & 0 \\ 0 & s_2 & 0 \end{bmatrix}.$$

In general, the number of age groups may not be three, and we may not be talking about plants; but we can generalize to get the following model for age structured population dynamics:

EXAMPLE 3.1. The state of the system is an m dimensional vector representing populations in m different age groups. For each age group $j < m$ there is a *survival rate* s_j giving the proportion of individuals who survive to the next age group, and for each age group j there is a a *fecundity rate* f_j giving the rate at which new first year individuals are produced. We assume $0 \leq s_j \leq 1$ and $f_j \geq 0$; these are the parameters of the system.

The transition rule is $F(x) = Lx$ where

$$L = \begin{bmatrix} f_1 & f_2 & f_3 & \dots & f_{m-1} & f_m \\ s_1 & 0 & 0 & \dots & 0 & 0 \\ 0 & s_2 & 0 & \dots & 0 & 0 \\ 0 & 0 & s_3 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & s_{m-1} & 0 \end{bmatrix}.$$

This is known as the *Leslie Model*, and L is the *Leslie matrix*.

Before investigating the properties of this system we introduce some terminology and prove a simple lemma that will be useful throughout this section.

A matrix or vector A is *non-negative* if all its entries are non-negative. It is *positive* if all its entries are positive.

LEMMA 3.2. *Suppose A is an $m \times m$ matrix and v is an m-dimensional vector. Then*

(a) If A is non-negative, $A_{jk} \geq a \geq 0$, and v is non-negative then the j^{th} entry in Ax is $\geq ax_k$.
(b) If A and v are non-negative then Av is non-negative.
(c) If A is positive and v is non-negative and not zero then Av is positive.

PROOF. To prove part (a) note that $(Av)_j = \sum_i A_{ji}x_i = A_{jk}x_k + \sum_{i \neq k} A_{ji}v_i$. The last summation is non-negative and the term $A_{jk}v_k$ is $\geq av_k$. Now part (b) follows by selecting $a = 0$. Finally, part (c) follows from part (a) by choosing k so that $v_k \neq 0$, so $v_k > 0$, and choosing a to be the minimum of the entries in A. $\square$

Now, by definition, all Leslie matrices are non-negative, and all state vectors that represent populations are also non-negative. When we iterate the transition rule $F(x) = Lx$ we obtain $F^t(x) = L^t x$. From the lemma we see that the powers L^t are non-negative, as is the future state vectore $L^t x$.

As a concrete example we return to our plant model, and assign some values to the parameters to get the Leslie matrix

$$L = \begin{bmatrix} 0 & 7 & 6 \\ \frac{1}{4} & 0 & 0 \\ 0 & \frac{1}{2} & 0 \end{bmatrix}. \tag{3.1}$$

The first few powers of L are

$$L^1 = \begin{bmatrix} 0 & 7 & 6 \\ 1/4 & 0 & 0 \\ 0 & 1/2 & 0 \end{bmatrix},\ L^2 = \begin{bmatrix} 7/4 & 3 & 0 \\ 0 & 7/4 & 3/2 \\ 1/8 & 0 & 0 \end{bmatrix},\ L^3 = \begin{bmatrix} 3/4 & 49/4 & 21/2 \\ 7/16 & 3/4 & 0 \\ 0 & 7/8 & 3/4 \end{bmatrix},$$
$$L^4 = \begin{bmatrix} 49/16 & 21/2 & 9/2 \\ 3/16 & 49/16 & 21/8 \\ 7/32 & 3/8 & 0 \end{bmatrix},\ L^5 = \begin{bmatrix} 21/8 & 379/16 & 147/8 \\ 49/64 & 21/8 & 9/8 \\ 3/32 & 49/32 & 21/16 \end{bmatrix}.$$

As noted above, each power L^t is non-negative. Moreover, for this example, L^5 is positive. It follows from Lemma 3.2c applied to each column of L that L^6 is positive, and a simple induction following this pattern shows that L^t is positive for all $t \geq 5$.

Again, by Lemma 3.2c, $L^t x$ will be positive if $t \geq 5$ and x is any non-negative, non-zero state vector. Hence, even if we start with only first year plants, after at most 5 years we will have plants in all age groups. As an example, suppose the initial state vector is $(100, 0, 0)$. Then

$$x = \begin{bmatrix} 100 \\ 0 \\ 0 \end{bmatrix}, \quad L^5 x = \begin{bmatrix} 262 \\ 77 \\ 9 \end{bmatrix} \text{ (rounded).}$$

Now notice that the smallest entry on the diagonal in L^5 is $21/16 > 1.3$. If we apply Lemma 3.2a for a non-negative vector v, with $j = k$, we see that each entry in $L^5 v$ is at lease 1.3 times the corresponding entry in v. Applying this with $v = L^5 x$, we conclude that each entry in $L^5 \cdot L^5 x = L^{10} x$ is at least $1.3 \cdot 9$ (since 9 is the smallest entry in $L^5 x$). If we iterate this argument we see that each entry in $L^{5n} x$ is at least

$(1.3)^{n-1} \cdot 9$. Hence each entry of $A^t x$ is unbounded as $t \to \infty$. (It is true that each entry approaches ∞ as $t \to \infty$. This requires slightly more work, although we will give a different proof later.)

We can see some further details of this behavior if we consider some larger values of t. Further calculation gives

$$y = L^{15}x = \begin{bmatrix} 19625 \\ 3304 \\ 1085 \end{bmatrix}, \quad z = L^{16}x = \begin{bmatrix} 29638 \\ 4906 \\ 1652 \end{bmatrix} \qquad \text{(rounded)}.$$

There are two interesting things to notice here. First,

$$\frac{z_1}{y_1} \approx 1.51, \quad \frac{z_2}{y_2} \approx 1.48, \quad \frac{z_3}{y_3} \approx 1.52$$

Hence z is essentially 1.5 times y. This indicates that the size of each component of the vector $L^t x$ is increasing by a factor of about 1.5 per year. This is much faster than our estimate above, which was 1.3 per 5 years, or $(1.3)^{1/5} \approx 1.05$ per year.

The other thing to notice is that y and z are almost parallel, since z is almost a scalar multiple of y. That is, it appears that y and z are pointing in approximately the same direction. In fact, we can make this explicit by rescaling so that their third components are equal to 1 (there are many other ways to "normalize" the vectors by rescaling). This leads to

$$\frac{1}{y_3} y \approx \begin{bmatrix} 18.09 \\ 3.05 \\ 1.00 \end{bmatrix}, \quad \frac{1}{z_3} z \approx \begin{bmatrix} 17.94 \\ 2.97 \\ 1.00 \end{bmatrix}.$$

So it appears that $L^t x$ is not just growing in magnitude at an exponential rate, but that its direction is converging to the direction of $\begin{bmatrix} 18 \\ 3 \\ 1 \end{bmatrix}$.

We need some more theory to confirm that this is indeed what is happening, and to generalize it.

3.2. Matrix powers, eigenvalues and eigenvectors.

Suppose we have an $m \times m$ matrix M and we need to calculate its powers M^n for $n = 1, 2, \ldots$. This is not generally easy to do, since even for integer matrices there is a lot of variation in the form of the powers. See Exercise 3.4 for some examples.

However, in some cases we can get a very good handle on the powers of a matrix. Suppose the matrix M is *diagonalizable*. This means that there is a basis $\{ v_1, v_2, \ldots, v_m \}$ of $\mathbb{R}^m$ consisting of *eigenvectors* of M; so corresponding to each v_k there is an *eigenvalue* λ_k, satisfying $Mv_k = \lambda_k v_k$. [Even if M is a real matrix the eigenvalues are, in general, complex numbers; they may not be real. We will discuss later how to deal with complex eigenvalues.]

There are two equivalent ways to see how to use the eigenvalues and eigenvectors of a diagonalizable matrix M to calculate powers of M.

For the first approach, form the matrix P whose columns are the eigenvectors of M; specifically, the k^{th} column of P is v_k. Since the vectors v_k form a basis, the matrix P is invertible. Next, form a diagonal matrix Λ using the eigenvalues as the diagonal entries. It is important to use the eigenvalues in the same order as the eigenvectors, so we should be explicit: the entry of Λ in the k^{th} row and k^{th} column is λ_k, and all other entries of Λ are 0. Then $P^{-1}MP = \Lambda$, or, by solving for M, $M = P\Lambda P^{-1}$. Now we can calculate M^n as

$$\begin{aligned}
M^n &= M \cdot M \cdot \ldots M && n \text{ copies of } M\\
&= (P\Lambda P^{-1}) \cdot (P\Lambda P^{-1}) \ldots (P\Lambda P^{-1}) && n \text{ copies of } \Lambda\\
&= P\Lambda(P^{-1}P)\Lambda(P^{-1}P) \ldots (P^{-1}P)\Lambda P^{-1} && \text{associative law}\\
&= P\Lambda \cdot \Lambda \cdot \ldots \Lambda P^{-1} && \text{since } P^{-1}P = I
\end{aligned}$$

$$M^n = P\Lambda^n P^{-1}. \tag{3.2}$$

It is easy to calculate Λ^n; it is still a diagonal matrix, with λ_k^n as the k^{th} diagonal entry.

In the second approach we use directly the fact that the eigenvectors form a basis. Hence any vector x in $\mathbb{R}^m$ can be written as a linear combination of the eigenvectors; that is, $x = c_1v_1 + c_2v_2 + \cdots + c_mv_m$. It is not hard to calculate the coefficients c_k: If we form a column vector c with entries given by the coefficients then Pc is the linear combination of the columns v_k of P with the coefficients c_k. Hence we have $Pc = x$ and, given x, we can solve this for c, either by row reduction or by $c = P^{-1}x$. Now we can calculate

$$M^nx = M^n(c_1v_1 + c_2v_2 + \cdots + c_mv_m) = c_1M^nv_1 + c_2M^nv_2 + \cdots + c_mM^nv_m.$$

But $Mv_k = \lambda_kv_k$, so $M^2v_k = M(Mv_k) = M(\lambda_kv_k) = \lambda_k(Mv_k) = \lambda_k(\lambda_kv_k) = \lambda_k^2v_k$. Continuing in this way we obtain $M^nv_k = \lambda_k^nv_k$. Plugging this into the equation above produces

$$M^nx = c_1\lambda_1^nv_1 + c_2\lambda_2^nv_2 + \cdots + c_m\lambda_m^nv_m. \tag{3.3}$$

Equations (3.2) and (3.3) greatly simplify analysing iterative matrix multiplication. There are a couple of problems. First, we may need to deal with non-real complex numbers, as noted above. More seriously, there are square matrices that are not diagonalizable. The main criterion for an $m \times m$ matrix to be diagonalizable is that it have m different eigenvalues, for then the corresponding eigenvalues are guaranteed to be linearly independent. If there are repeated eigenvalues then there will not necessarily be a basis of eigenvectors, and in this case the matrix is not diagonalizable. This occurs rarely, especially with matrices that are determined empirically or experimentally. However, when it does occur the analogs of (3.2) and (3.3) are more complicated and harder to analyse. We will take the approach in this book that results will be stated, whenever possible, without assuming that the matrices involved are diagonalizable; but any derivations will assume diagonalizable matrices.

Here is a typical result along these lines. We say an eigenvalue λ_1 is a *dominant eigenvalue* if $|\lambda_1| > |\lambda_k|$ for all $k > 1$. This implies that λ_1 is a *simple eigenvalue*; that is, it has multiplicity 1 as a root of the characteristic equation. Moreover, it implies that a dominant eigenvalue of a real matrix must be real. Some texts use the term "dominant eigenvalue" in a weaker sense, which does not imply simple. For that reason we will usually sue the phrase "simple dominant eigenvalue" in formal statements.

THEOREM 3.3. *Suppose λ_1 is a simple dominant eigenvalue. Then any vector x can be written uniquely in the form $x = c_1 v_1 + w$ where*

$$\lim_{n\to\infty} \frac{1}{\lambda_1^n} M^n w = 0.$$

Moreover,

$$\lim_{n\to\infty} \frac{1}{\lambda_1^n} M^n x = c_1 v_1,$$

so we have the following possibilities:

(a) If $|\lambda_1| < 1$ then $M^n x \to 0$ as $n \to \infty$.
(b) If $|\lambda_1| = 1$ and $c_1 \neq 0$ then $M^n x$ is bounded but does not converge to 0.
(c) If $|\lambda_1| > 1$ and $c_1 \neq 0$ then $M^n x$ is not bounded, and its norm converges to ∞.
(d) If $|\lambda_1| \geq 1$ and $c_1 = 0$ then nothing can be said about the limiting behavior of $M^n x$ without further information about M and x.

PROOF. This actually follows easily from (3.3), in case M is diagonal. We define $c_1 v_1$ as in (3.3), and we define $w = c_2 v_2 + c_3 v_3 + \cdots + c_m v_m$. Then

$$\begin{aligned}\frac{1}{\lambda_1^n} M^n w &= \frac{1}{\lambda_1^n}\left(c_2 \lambda_2^n v_2 + c_3 \lambda_3^n v_3 + \cdots + c_m \lambda_m^n v_m\right) \\ &= \left(\frac{\lambda_2}{\lambda_1}\right)^n c_2 v_2 + \left(\frac{\lambda_3}{\lambda_1}\right)^n c_3 v_3 + \cdots + \left(\frac{\lambda_m}{\lambda_1}\right)^n c_m v_m,\end{aligned}$$

and this clearly converges to 0 since each of the quotients in parentheses is less than 1 in absolute value. The limit of $\frac{1}{\lambda_1^n} M^n x$ is calculated in the same way, except that

$c_1 v_1$ is present with a multiplier of $\left(\dfrac{\lambda_1}{\lambda_1}\right)^n = 1$. The remainder of the theorem follows by multiplying this limit by λ_1^n. □

Note that, in case $|\lambda_1| > 1$ and $c_1 \neq 0$, the fact that $\dfrac{1}{\lambda_1^n} M^n x$ converges to $c_1 v_1$ can be interpreted as saying that the limiting direction of the vectors $M^n x$ is parallel to the eigenvector v_1.

This theorem explains our numerical results in the previous section about powers L^t of the Leslie matrix (3.1). If we calculate the characteristic polynomial $\det L - \lambda I$ we obtain $-\lambda^3 + \frac{7}{4}\lambda + \frac{3}{4}$. This has the three roots $\lambda_1 = \dfrac{3}{2}$, $\lambda_2 = -\dfrac{1}{2}$, $\lambda_3 = -1$. These are the eigenvalues of L, and we can calculate the corresponding eigenvectors as

$$v_1 = \begin{bmatrix} 18 \\ 3 \\ 1 \end{bmatrix}, \; v_2 = \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}, \; v_3 = \begin{bmatrix} 8 \\ -2 \\ 1 \end{bmatrix}.$$

Therefore the matrix L is diagonalizable, and $\lambda_1 = \frac{3}{2}$ is the dominant eigenvalue. We can express our initial state x in terms of the eigenvectors, as

$$x = \begin{bmatrix} 100 \\ 0 \\ 0 \end{bmatrix} = \frac{5}{2} v_1 - \frac{25}{2} v_2 + 10 v_3,$$

so $c_1 = 5/2$. Hence we see that $L^t x \to \infty$ as $t \to \infty$, at an approximate exponential rate of $\lambda_1 = 1.5$; and the limiting direction of $L^t x$ is in the same direction as v_1. Both of these correspond closely to what we guessed from the numerical evidence.

3.3. Non negative matrices

In the last section we used eigenvalues and eigenvectors to analyse a specific Leslie model, but we need some guidance as to when such an approach will be appropriate. It turns out that a classic theorem on positive matrices is just what we need.

THEOREM 3.4 (Perron-Frobeniuus). *Suppose that A is a non-negative $m \times m$ matrix and that some power of A is positive. Then A has a simple dominant eigenvalue, λ_1. Moreover,*

(a) λ_1 is a positive real number.
(b) There is a positive eigenvector v_1 corresponding to λ_1.
(c) For all non-zero non-negative vectors x, $\lim_{n\to\infty} \dfrac{1}{\lambda_1^n} A^n x = c_1 v_1$ with $c_1 > 0$.
(d) No other eigenvector of A is positive, or even non-negative.

Sometimes we will refer to λ_1 as the Perron-Frobenius eigenvalue of A, and to a positive eigenvector as the Perron-Frobenius eigenvector.

We will not prove this result now; it requires several concepts that belong to the realms of analysis and topology. The theorem is usually stated under the assumption that A is positive, but the version above follws easily from the usual statement.

Use of the Perron-Frobenius Theorem simplifies our previous discussion of plant populations using the Leslie matrix (3.1). If we calculate L^5 we see that L satisfies the hypotheses of the Perron-Frobenius Theorem, so we know that it has a dominant positive real eigenvalue. If we calculate the eigenvalues we see that the dominant eigenvalue is $\lambda_1 = 1.5$, with eigenvector $v_1 = \begin{bmatrix} 18 \\ 3 \\ 1 \end{bmatrix}$. According to the Perron-Frobenius Theorem, for any non-negative non-zero initial state x, all components of $L^t x$ approach ∞ at an approximate exponential rate of $3/2$; and the direction of $L^t x$ converges to the direction of the eigenvector v_1.

In fact, this situation applies to any Leslie model: As long as some power of L is positive then all we need to do to predict the limiting behavior of $L^t x$ is to find the Perron-Frobenius eigenvalue and corresponding eigenvector. If $\lambda_1 > 1$ then the analysis is as above; if $\lambda_1 < 1$ any initial population eventually dies out; and if $\lambda_1 = 1$ any non-zero initial population vector will converge to a stable population vector which points in the direction given by v_1.

For a case in which L is guaranteed to have a positive power see Exercise 3.6.

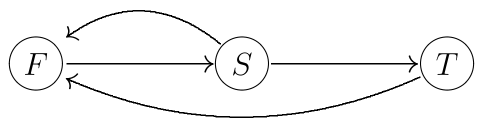
3.4. Networks; more examples

Many dynamical systems models can be described using a type of *network*. There are a number of different formulations of this; we'll use the following.

A *directed graph*, or *digraph*, is a finite collection of objects called *vertices* or *nodes*, together with a collection of ordered pairs of nodes, called *edges*. We say an edge $e = (a, b)$ is the edge *from a to b*; we say the node a is the *tail* or *source* of e and the node b is the *head* or *target* of e. We will usually designate an edge from a to b as $e \to b$ rather than as (a, b). Note that it is possible for an edge to connect a node to itself.

Warning: There are a number of different definitions of digraph available, depending on the context. According to the most prevalent definition, what we are talking about is technically a *pseudograph with no multiple edges*. Use caution when consulting other sources.

We will visualize a digraph as a diagram in which the nodes are represented by numbers or other symbols, and the edges are represented as arrows connecting two nodes (which may not be different). Here is an example:

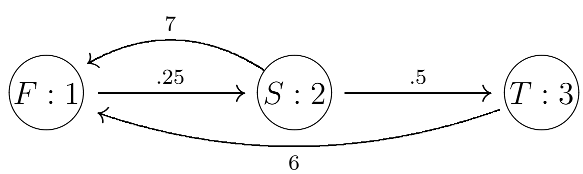


In this digraph there are three nodes, F, S, T, and there are four edges. We will only need digraphs with the property that there can be at most one edge connecting any pair of nodes, so we can identify the edges by their starting and ending nodes. So in this example the nodes are $F \to S$, $S \to T$, $S \to F$ and $T \to F$.

If we number the nodes in a digraph then we can completely specify the digraph by a matrix A in which the ij entry is 1 if there is an edge from i to j, and otherwise is zero. This matrix is called the *adjacency matrix* of the digraph. If we number the nodes in the example above in the order F, S, T then the adjancy matrix is

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}.$$

For our uses we need to decorate a digraph with extra information. Specifically, we will attach quantities, called *weights*, to the edges of a digraph; the result is called a *weighted digraph*. The weights can be represented graphically; for example, here is the example above, with weights attached to the edges. We have also added index numbers to the nodes, to make the ordering explicit.



We can modify the adjancy matrix to record the weight information, to construct the *weight matrix* W: we just enter the weight for the edge from i to j in location ij. If there is no edge from i to j we enter 0. If no weights are 0 then we can reconstruct the adjacency matrix from the weight matrix, since in this case $A_{ij} = 1$ if and only if $W_{ij} \neq 0$. Here is the weight matrix for our example:

$$W = \begin{bmatrix} .25 & 0 & 0 \\ 7 & 0 & .5 \\ 6 & 0 & 0 \end{bmatrix}.$$

You have probably already noticed the connection with the stratified population model discussed in section 3.1. For example, we interpret the weighted edge $F \xrightarrow{.25} S$ as representing the fraction of the first year plants (F) that survive be part of the second year group (S); and the weighted edge $S \xrightarrow{7} F$ represents the number of viable seeds produced by each second year plant.

In general, we can interpret a weighted digraph as specifying the following dynamical system: There is a state vector x, with each entry x_i corresponding to the node in the digraph labelled i. The weight W_{ij} attached to an edge $i \to j$ is interpreted as the fraction of the state variable x_i that is contributed to the next value of the state variable x_j in one time unit. Hence the new value of x_j is the sum of all such contributions, or

$$\text{new } x_j = x_1 \cdot W_{1j} + x_2 \cdot W_{2j} + \cdots + x_m \cdot W_{mj} = \sum_i x_i W_{ij}.$$

In matrix terms, this says that

$$\text{new } \begin{bmatrix} x_1 & x_2 \ldots x_m \end{bmatrix} = \begin{bmatrix} x_1 & x_2 \ldots x_m \end{bmatrix} \cdot W.$$

However, we are consistently interpreeting state vectors x as *column vectors*, not as *row vectors*. To convert the row vector $\begin{bmatrix} x_1 & x_2 \ldots x_m \end{bmatrix}$ into the column vector x we take the transpose, and remember that transposing reverses the order of matrix multiplication, so the new value of x is $\left(\begin{bmatrix} x_1 & x_2 \ldots x_m \end{bmatrix} \cdot W\right)^T = W^T \begin{bmatrix} x_1 & x_2 \ldots x_m \end{bmatrix}^T = W^T x$. So we have a general procedure for creating a linear dynamical system:

The discrete dynamical system associated to a weighted digraph with vertices $\{1, 2, \ldots, m\}$ and weight matrix W is

$$F(x) = W^T x \tag{3.4}$$

where x is an m-dimensional state vector.

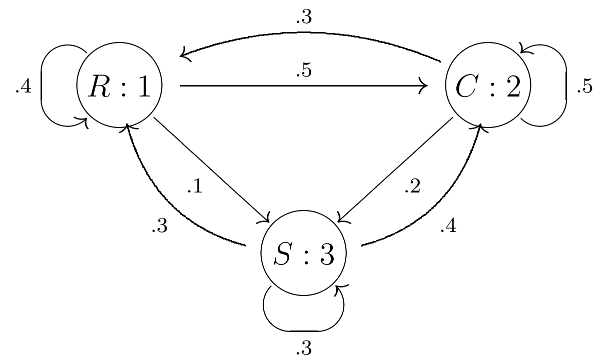
This is exactly the procedure that leads to the Leslie matrices discussed in sections 3.1, 3.2 and 3.3, and it is easily adapted to similar problems in population dynamics and other studies. See Exercise 3.7 for an ecology problem based on a network.

Here is a very different kind of application of digraph methods. We shall consider the weather in Binghamton. We suppose that every day can be classified into exactly one of the following categories:

R: Rainy (or snowy).
C: Cloudy.
S: Sunny.

Historically we know that each day's weather is closely related to the next: For example, if today is rainy then 40% of the time it will be rainy tomorrow; 50% of the time it will be cloudy tomorrow; and so on. There are then nine possible transitions from the weather on one day to the weather on the next day, each with an associated probability; we can summarize this via a digraph:

(3.5)



(The numbers are just guesses; a study of several years of weather data would lead to more accurate probabilities.)

The corresponding weight matrix is

$$W = \begin{bmatrix} .4 & .5 & .1 \\ .3 & .5 & .2 \\ .3 & .4 & .3 \end{bmatrix}. \tag{3.6}$$

This matrix has two important properties, which constitute the definition of a *stochastic matrix* (more precisely, a *right stocahastic matrix*): All entries in W are non-negative, and the entries in each row add up to 1. You can check this in the specific example above, but it is much more generally true: The entries in W must be non-negative since they are probabilities, which cannot be negative; and the entries in any row must add to 1 because these entries give the probabilities for the weather on the next day, and the next day's weather must fit into exactly one of our categories.

The corresponding dynamical system has the form $F(x) = Mx$ where $M = W^T$. We interpret the state vector x as a vector of probabilities, so the j^{th} component of $F^t(x)$ is the probability that the weather will be in category j (R, C or S) after t days. The initial condition x is the actual weather at time $t = 0$; since today is sunny I'll take the initial value of x to be $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$. The first thing to notice is that $F^t(x)$ is always a *stochastic vector*: this means that all entries of x are non-negative and that their sum is 1. This is exactly what we should expect, but it is instructive to prove this, as a property of any stochastic matrix.

Suppose that W is an $m \times m$ stochastic matrix. Let u be the m-dimensional column vector in which all entries are equal to 1. Then the j^{th} entry of Wu is

$\sum_k W_{jk}u_k$. since the u_k entries are all 1, this just becomes $\sum_k W_{jk}$. In other words, the entries of Wu are the sums of the rows of W. Hence the condition that the row sums of W are all equal to 1 can be expressed neatly as $Wu = u$. In other words, the row sums of W are equal to 1 if and only if u is an eigenvector of W with eigenvalue 1.

From the discussion about eigenvectors of powers of matrices in section 3.2, it follows that u is an eigenvector of M^n with eigenvalue 1, for any positive integer n. Since powers of non-negative matrices are non-negative, we conclude that powers of a stochastic matrix are still stochastic.

Moreover, a non-negative vector x is a stochastic vector if and only if $u^T x = 1$. It then follows that, if W and x are stochastic then $M^n x$ is stochastic for all positive n, since $M^n x$ is non-negative and

$$u^T \cdot M^n \cdot x = u^T \cdot \left(W^T\right)^n \cdot x = u^T \cdot (W^n)^T \cdot x = (W^n u)^T \cdot x = u^T x = 1.$$

A dynamical system of the form $F^t(x) = M^t x$ where x is a stochastic matrix and M is the transpose of a stochastic matrix is called a *Markov chain.* There is a very nice result for the limiting behavior of Markov chains:

THEOREM 3.5. *Suppose that M is the transpose of a stochastic matrix, and that some power of M is positive. Then 1 is the dominant eigenvalue of M, and it has a unique stochastic eigenvector p. If x is any stochastic vector then $M^t x$ converges to p as $t \to \infty$. In particular, M^t converges to the matrix in which all columns are equal to p.*

PROOF. Let λ be the Perron-Frbenius eigenvalue of M, and let v be a corresponding positive eigenvector. Since v is positive, $u^T v$ is a positive scalar, so we can define $p = \dfrac{v}{u^T v}$. Then p is positive, and $u^T p = 1$, so p is a stochastic vector. Since p is a scalar multiple of the eigenvector v we have $Mp = \lambda p$.

Now $M = W^T$ where W is a stochastic matrix, so $Wu = u$. Transposing this yields $u^T M = u^T W^T = u^T$. Hence $u^T Mp = u^T p = 1$. On the other hand, $Mp = \lambda p$ so $u^T Mp = \lambda u^T p = \lambda$. We have shown that $\lambda = 1$.

The rest of Theorem 3.5 now follows from the Perron-Frobenius Theorem. □

As an application, let us return to the Binghamton weather example, 3.5, corresponding to the stochastic matrix 3.6. We know that 1 is a simple eigenvalue of $M = W^T$, so we can find the limit state vector p of Theorem 3.5 as follows:

(a) Find a basis for the one-dimensional null space of $M - I$; in other words, find a non-zero solution v of $(M - I)v = 0$. Via row reduction we get $v = \begin{bmatrix} 1.8 \\ 2.6 \\ 1 \end{bmatrix}$.

(b) Divide v by $u^T v = 5.4$, the sum of the entries in v, to get $p = \begin{bmatrix} 0.33 \\ 0.48 \\ 0.19 \end{bmatrix}$.

We can interpret this as saying that eventually the probabilities of rainy, cloudy and sunny days will stabilize at 33%, 48% and 19%.

You should be aware that there is an alternative to constructing the dynamical system in terms of the transpose of the weight matrix W. In many applications you will find that the weight matrix is not transposed, but that the dynamical system is expressed in terms of the transition rule $F(y) = yW$. In this case, y is the state vector expressed as a row vector; it is the same as the traspose of our x. In this case we iterate F to get $F^t(y) = yW^t$. All the techniques of this chapter can be modified to handle this formulation. For example, the eigenvalue and eigenvector analysis can be adapted, but we need to talk about *left eigenvalues* and *left eigenvectors*.

This alternative approach is usually used for Markov chains, but our original approach is generally used for Leslie matrices. There is little uniformity in other types of models based on matrix powers.

3.5. Complex eigenvalues

Even though the problems that we are looking at are stated with only real numbers we often find it necessary to use the complex number system to solve them. One way that complex numbers arise is in finding eigenvalues: you need to solve a polynomial equation, and generally, many of the roots are non-real.

In this section we explain one way to deal with non-real eigenvalues without leaving the real domain. The basic facts about algebra in the complex number system are summarized in Appendix A.1.

Start with a $m \times m$ real matrix A. There is no real difference in finding eigenvalues and eigenvectors in the complex domain. We start by solving the characteristic polynomial of A to find the eigenvalues. The characteristic polynomial has degree m and the Fundamental Theorem of Algebra (Theorem A.1) guarantees that there will be m complex solutions, counted with multiplicity.

We will simplify things by assuming that the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_m$ are all distinct.

The next step is to find an eigenvector v_k corresponding to each eigenvalue λ_k. This is done by row reduction of $A - \lambda_k I$, so if λ_k is non-real then v_k will be non-real. When we string the eigenvectors together to form the change of basis matrix P we will have $A = P\Lambda P^{-1}$ where Λ is the diagonal matrix with the eigenvalues along the diagonal. Each non-real eigenvalue will correspond to a non-real column in P.

Here's a simple example: If $A = \begin{bmatrix} 3 & 2 \\ -1 & 1 \end{bmatrix}$ then the characteristic polynomial is $\lambda^2 - 4\lambda + 5$. Using the quadratic formula we find that the eigenvalues are $\lambda_1 = 2 + i$ and $\lambda_2 = 2 - i$. The corresponding eigenvectors are $v_1 = \begin{bmatrix} -1 - i \\ 1 \end{bmatrix}$ and $v_2 = \begin{bmatrix} -1 + i \\ 1 \end{bmatrix}$. We have $P = \begin{bmatrix} -1 - i & -1 + i \\ 1 & 1 \end{bmatrix}$ and $P^{-1} = \begin{bmatrix} \frac{i}{2} & \frac{1+i}{2} \\ -\frac{i}{2} & \frac{1-i}{2} \end{bmatrix}$, so

$$A = \begin{bmatrix} 3 & 2 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} -1 - i & -1 + i \\ 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 2 + i & 0 \\ 0 & 2 - i \end{bmatrix} \cdot \begin{bmatrix} \frac{i}{2} & \frac{1+i}{2} \\ -\frac{i}{2} & \frac{1-i}{2} \end{bmatrix}$$

This is usable in, for example, calculations of A^n. However, it is often easier to keep in the real domain.

To do this we reconsider the problem of finding complex eigenvectors. Suppose that λ is a non-real eigenvalue, with a corresponding eigenvector v. From Proposition A.2 we know that $\overline{\lambda}$ is also an eigenvalue. In fact, if we conjugate the eigenvector equation $Av = \lambda v$ we obtain $\overline{Av} = A\overline{v} = \overline{\lambda v} = \overline{\lambda}\overline{v}$. This shows directly that $\overline{\lambda}$ is an eigenvalue, and it also demonstrates that $\overline{v}$ is an eigenvector corresponding to $\overline{\lambda}$.

Now, in this case, suppose that $\lambda = a + bi$ and define $x = \frac{1}{2}(v + \overline{v})$ and $y = \frac{1}{2i}(v - \overline{v})$. These are real vectors; in fact, each component of x is the real part of the corresponding component of v and each component of y is the imaginary component of the corresponding component of v. In other words, $v = x + iy$ and $\overline{v} = x - iy$.

Of course x and y are not eigenvectors of A, but we have the following calculation:

$$\begin{aligned} Ax &= \tfrac{1}{2}(Av + A\overline{v}) = \tfrac{1}{2}\left(\lambda v + \overline{\lambda}\overline{v}\right) \\ &= \tfrac{1}{2}((a + bi)(x + iy) + (a - ib)(x - iy)) \\ &= \tfrac{1}{2}(ax + iay + ibx - by + ax - iay - ibx - by) \\ &= \tfrac{1}{2}(2ax - 2by) = ax - by. \end{aligned}$$

A similar calculation shows that $Ay = bx + ay$. This leads to the following, which is an analog of the diagonalization theorem but which only uses real matrices.

THEOREM 3.6 (Real canonical form, special case). *Suppose A is a real $m \times m$ matrix with m distinct eigenvalues, listed as $r_1, \ldots, r_p, c_1, \overline{c}_1, \ldots, c_q, \overline{c}_q$ where the r's are real and the c's are non-real, and $p + 2q = m$. Then there is a real invertible matrix Q so that $A = QMQ^{-1}$, where all the entries in M are zero except:*

(a) $M_{jj} = r_j$.

(b) If $c_j = a_j + ib_j$ and $n = p + 2j - 1$ then the submatrix $\begin{bmatrix} M_{nn} & M_{n,n+1} \\ M_{n+1,n} & M_{n_1,n+1} \end{bmatrix}$ is $\begin{bmatrix} a_j & -b_j \\ b_j & a_j \end{bmatrix}$.

Proof. Proposition A.2 shows how to list the eigenvalues in the given order. Since the eigenvalues are distinct, A is diagonalizable, so we can assemble the eigenvectors in a matrix P so that $P^{-1}AP$ is diagonal with the eigenvalues on the diagonal in the given order.

Q is formed from P by replacing each pair of columns given by a conjugate pair $v_j, \overline{v}_j$ of non-real eigenvectors with the real vectors $x_j = \frac{1}{2}(v_j + \overline{v}_j)$ and $y_j = \frac{1}{2i}(v_j - \overline{v}_j)$. The calculations above the statement of the theorem show that $M = Q^{-1}AQ$ has the desired form. □

Continuing our example from above, we have $x = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ and $y = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, as the real and imaginary parts of the eigenvector $v = \begin{bmatrix} -1-i \\ 1 \end{bmatrix}$. Hence $Q = \begin{bmatrix} -1 & 1 \\ 1 & 0 \end{bmatrix}$, with inverse $Q^{-1} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}$. The eigenvalue corresponding to v is $a + bi = 2 + i$, so the real canonical form of A is $M = \begin{bmatrix} a & -b \\ b & a \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ 1 & 2 \end{bmatrix}$. Therefore,

$$A = \begin{bmatrix} 3 & 2 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} -1 & 1 \\ 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} 2 & -1 \\ 1 & 2 \end{bmatrix} \cdot \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}.$$

We can further analyze 2×2 blocks of the form $M = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$ by putting the complex number $a + bi$ in polar form, so $a + bi = r(\cos\theta + i\sin\theta)$. Substituting this into M and factoring out the scalar r, we have $M = \begin{bmatrix} a & -b \\ b & a \end{bmatrix} = r\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$, with $r = |a + bi| = \sqrt{a^2 + b^2}$. This matrix represents a transformation of the plane consisting of a rotation through θ, followed by a scale change given by r. This is the same as the geometric interpretation of multiplication by $a + bi$ as a transformation of the complex plane (see Appendix A.1). It is then easy to interpret M^n by reinterpreting De Moivre's formula (A.1):

$$M^n = r^n \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}^n = r^n \begin{bmatrix} \cos(n\theta) & -\sin(n\theta) \\ \sin(n\theta) & \cos(n\theta) \end{bmatrix}. \tag{3.7}$$

That is, geometrically, M^n represents a rotation through $n\theta$, followed by a scale change of r^n.

In our example above, $r = |2 + i| = \sqrt{5}$ and $\theta = \arctan(1/2)$, so M^n represents a rotation through $n\theta$ followed by a scale change of $5^{n/2}$. Since $A^n = QM^nQ^{-1}$, we see that the entries in A^n can be expected to grow essentially as a constant times $5^{n/2}$, since the rotational part of M^n does not affect sizes of vectors.

Exercises

3.1. In the year 1202 Fibonacci published the mathematics book *Liber abaci*, which was one of the first books to promote the use of Arabic numerals in Europe. Here is one of the exercises from the book:

A certain man put a pair of rabbits in a place surrounded on all sides by a wall. How many pairs of rabbits can be produced from that pair in a year if it is supposed that every month each pair begets a new pair which from the second month on becomes productive?

This is perhaps the first age-structured population system. It is not a Leslie model (for example, the rabbits never die), but it leads to the same kind of dynamical system. Represent the state of the population by a vector $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ where x_1 is the number of juvenile rabbit pairs (less than a month old) and x_2 is the number of mature rabbit pairs (1 month or older). Of course, time is measured in months.

The transition function has the form $F(x) = Ax$ for some 2×2 matrix. What is A? What is the initial state corresponding to Fibonacci's problem? What numbers did Fibonacci invent when he solved the problem? Can you express $A^t x$ in terms of these numbers?

3.2. Replace the Leslie matrix in the discussion of 3.1 with $L = \begin{bmatrix} 0 & 3 & 7/3 \\ 1/4 & 0 & 0 \\ 0 & 1/2 & 0 \end{bmatrix}$, and use the same initial state vector $x = \begin{bmatrix} 100 \\ 0 \\ 0 \end{bmatrix}$. Use Maple or another computer package to answer the following.
(a) Is some power of L positive?
(b) Calculate $L^{20}x$.
(c) Calculate the eigenvalues of L.
(d) How long will it be until the first component of $L^t x$ is greater than 10000.

3.3. Consider the following generalization of the Leslie matrix in the discussion of 3.1: $L = \begin{bmatrix} 0 & a & b \\ 1/4 & 0 & 0 \\ 0 & 1/2 & 0 \end{bmatrix}$, with $a \geq 0$ and $b \geq 0$. Under what conditions on a and b will 1 be an eigenvalue of L? [Suggestion: Find the characteristic polynomial of L. Now plug $\lambda = 1$ into the characteristic polynomial and set the result equal to 0. This will give you an equation involving a and b. Now find the restrictions necessary so that $a \geq 0$ and $b \geq 0$.]

3.4. Find a general formula for A^n, or at least a non-recursive procedure for determining A^n for any n, for each of the following:,

(a) $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$

(b) $A = \begin{bmatrix} 2 & -2 & 8 \\ -1 & 2 & -6 \\ -1 & 1 & -4 \end{bmatrix}$

(c) $A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$

(d) $A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$

(e) $A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$

(f) $A = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$

3.5. For each of the matrices in Exercise 3.4 find the eigenvalues and decide whether there is a dominant eigenvalue. In cases where there is a dominant eigenvalue, verify that the conclusion of Theorem 3.3 is satisfied. Use $x =$ the unit vector with 1 in the first entry and 0's elsewhere.

3.6. In the general form of the Leslie matrix in Example 3.1 suppose that all the survival rates s_j and all the fecundity rates f_j for $j > 0$ are positive. Then some power of L is a positive matrix.

3.7. This is a classical example of a linear dynamical system. We want to predict polution levels in the Great Lakes, assuming all sources of pollution are turned off; the goal is to see how long it will take before the water in the lakes is reasonably free of pollution. (This might be a good time to look at a map of the Great Lakes.)

The basic model is very simplistic: There is an amount of pollutant, measured in tons, dispersed evenly throughout each of the lakes (a different amount for each lake). Dividing this amuount by the volume of the lake, measured in cubic miles, gives the amount of pollution per unit volume, or the concentration, in the lake. If water flows from lake A to lake B then, in one year, some of the pollutant will be removed from lake A and added to lake B. Since the pollutant is assumed to be evenly distributed, the amount of pollutant transferred from lake A to B is just the volume of water that flows from lake A to lake B in one year, times the concentration of pollutant in lake A. If, during the year, some water flows out of a lake, but not to another lake in the system, then the corresponding amount of pollutant is just removed from the lake. If there is inflow into a lake from outside the system then we presume that it does not carry any pollution into the lake.

Here are the current volumes and flow rates, in cubic miles:

Lake	Volume	Inflow	Outflow	Flow
Superior	2900	15		15 → Huron
Michigan	1180	38		38 → Huron
Huron	850	15		68 → Erie
Erie	116	17		85 → Ontario
Ontario	393	14	99	

Suppose the original pollutant totals 5439 tons, for an overall concentration of 1 ton per cubic mile. Select an initial distribution of the 5439 tons among the 5 lakes. You can just assume that all have the same concentration, or you can assume, for example, that the concentration in Superior is only a 0.1 ton per cubic mile, with correspondingly larger concentrations in the other lakes. Just make sure that, in your initial distribution, the total amount of pollutant is 5493, that all initial concentrations are at least 0.1, and that no lake has a concentration greater than 5.

(a) The state vector x is 5 dimensional, and contains the *concentration* of pollutant in each lake. Set up the transition rule for the state vector; it will have the form $F(x) = Ax$ where A is a non-negative matrix. Be careful: For flow from one lake to another you must account for changes in both lakes.

(b) Use Maple to find the first time t at which all the lakes have a concentration less than .01.

(c) Use Maple to determine the largest concentration that will occur in any lake.

I do not expect that you willl need any advanced features of Maple to answer these questions. You just need to calculate values of $F^t(x)$ and inspect the answers. Turn in a copy of your Maple session. If you want, send it to me by email, as a text file, not as a live Maple worksheet.

3.8. Let $B = \begin{bmatrix} -4 & 4 & 6 \\ 6 & -2 & -6 \\ -8 & 7 & 10 \end{bmatrix}$; its characteristic polynomial is $-\lambda^3 + 4\lambda^2 - 14\lambda + 20$. The calculations below do not involve any messy numbers, so they can be done by hand. Optionally, use Maple.

(a) Find the eigenvalues. [Hint: 2 is an eigenvalue, so you can factor $2 - \lambda$ out of the characteristic polynomial. Now use the quadratic formula. You should get two non-real answers, $a \pm bi$, where a and b are integers.]

(b) Find the eigenvectors, v_1, v_2, v_3, corresponding, in order, to the eigenvalues 2, $a + bi$, $a - bi$.

(c) Calculate Q as the real matrix with columns v_1, $\frac{1}{2}(v_2 + v_3)$, $\frac{1}{2i}(v_2 - v_3)$.

(d) Find $Q^{-1}BQ$. You can just write down the answer and check it, or you can calculate it.