MODULE 13.2

Matrices for Population Studies—Linked for Life

Prerequisite: Module 13.1, "Computational Toolbox—Tools of the Trade: Tutorial 7," or "Alternative Tutorial 7" (up to section on "Eigenvalues and Eigenvectors"). Additional high-performance computing materials related to this module are available on the text's website.

Downloads

The text's website has the file *PopsAndMatOps*, which contains examples from this module, available for download for various computational tools and project data files *cell_trajectory_file.txt*, *cell_types_file.txt*, and *cell_vel_file.txt*.

Population Matrices and High-Performance Computing

Blue crabs (*Callinectes sapidus*) are very important to life along the Gulf Coast of the United States. Essential to the complex, estuarine food webs, these animals also represent the second-largest commercial fishery in the area, and thereby provide livelihoods for many—and they are delicious! Because the ecological and economic impact of population fluctuations of this species is immense, our understanding of the dynamics of crab populations is crucial. Human intrusion (oil spills, pollution, overfishing, habitat degradation, etc.) and natural disasters (hurricanes, etc.) in addition to natural oscillations (climate cycles, dispersal, etc.) all impact populations. Some environmentalists may decry the emphasis on the crabs' economic importance, but the reality is that proper management is necessary for healthy, sustained populations. We need to understand better the factors critical to the quality and quantity of native populations.

Actually, we do know quite a bit about blue crabs. They are found in the western Atlantic Ocean, from Nova Scotia to Argentina, in the Gulf of Mexico and in the

Caribbean. Also, blue crabs have been introduced into the North Sea, southwest of France, parts of the Mediterranean, and Japan. Found to depths of 90 m, they eat about anything—plants, benthic invertebrates, small fish, detritus, carrion—and do a great deal of cannibalism (FAO 2012; Zinski 2006).

Females mate only one time, right after what is called a "pubertal" molt. As she readies for this molt, she calls male crabs with chemical signals. Drawn to her allure (and attractants), as with many animals, the males may squabble over mating rights. The winner often cradles the female until her key molt. Once she is "softened up" for mating, sperm is transferred to storage sacs (seminal receptacles), from which the female will fertilize her eggs. When her shell hardens, the female migrates to estuaries, where she buries herself in mud to overwinter. With the arrival of spring, the female fertilizes and transfers her eggs to form a mass (sponge) attached to her body, which often contains about 2 million individuals (some up to 8 million; Zinski 2006).

Hatching into larvae (**zoeae**) in about 2 weeks, they are carried out into the open ocean, where some feed, grow, and molt several times over a month or more before they are transformed into **megalops**. Over 1 to 3 weeks, these swimming larvae are transported closer to shore. On shore, they molt into juvenile crabs and then head up the estuaries (primary habitat along the Gulf Coast), where they reside, grow, and undergo numerous molts. Maturity normally is achieved by the following summer. Adult males tend to stay in the upper estuaries (lower salinity), whereas adult females, after mating, remain in the lower reaches (higher salinity). Of course, most of the millions of fertilized eggs/larvae never reach adulthood, because they become food for other organisms—including their own kind (Zinski 2006).

Although we know a great deal, we still do not know enough to understand the population dynamics of this or any other species that is passively dispersed over large areas. The countless larval stages are of small sizes and at the mercy of predators, currents, and winds. There is considerable drift of immatures from their birth estuary among other estuaries connecting the adults of different sites. How is it possible to understand the population dynamics of this species when we so obviously do not understand dispersal?

Gulf coast populations are considered **metapopulations**, which means that they are spatially fragmented. The extent of **connectivity**, or exchange of individuals among these populations, is extremely important for population stability and recolonization following local extirpation events. Larval dispersal is very much influenced by mortality, duration of planktonic stages, and behavior in the water column and upon settling. To assess connectivity, scientists must quantify the controlling influences for transport, stocks, and maintenance.

Given the scale and complexity of this problem, scientists are turning to computer modeling and simulation to work out spatially explicit models for blue crab populations. These multifaceted, ecological models are now possible because finely tuned hydrodynamic models of coastal areas are available. Before they can develop any useful population model, scientists must determine the influence of larval dispersal, settlement, and survival rates on fluctuations in blue crab numbers and also a connectivity matrix for the estuaries, which is a rectangular array of numbers indicating contacts various estuary populations have with one another.

Using the Northern Gulf of Mexico Nowcast-Forecast System of the U.S. Navy, biologists at Tulane University are able to use a particle-tracking model (PTM) to simulate larval dispersal. The Navy system incorporates tides, freshwater runoff, winds, sea height, sea temperatures, and 3D current velocities. So, using this type of input and PTM, they can follow the trajectory of individual particles (larvae) with time. The resulting dispersal model can then be incorporated into the larger population model.

Over a 3-year period, scientists have collected more than a terabyte (space for 10^{12} characters) of data for their study. On a 2010 sequential computer, they estimate that the simulation time for one larva is 5 min and for 2000 larvae is 1 week. Thus, averaging the results of 300 simulations involving 2000 larvae each would take about 5.7 years! With such massive amounts of data and such intensive computations, researchers must use high-performance computing with multiple computer processors to store the data and large matrices and to perform the needed simulations in a reasonable amount of time (Taylor 2010).

In this module, we examine populations that change with time. To make long-term predictions about these populations, we store their data in structures called **vectors** and **matrices** and perform calculations on these structures.

Vectors

A **data structure** is a formal skeleton that can hold data and on which we can perform specific operations. One data structure that most computational tools and computer languages have is a **vector**, or a one-dimensional array. Vectors allow us to collect a great deal of similar data together under one name instead of thinking of perhaps hundreds of individual variable names. For example, Table 13.2.1 indicates simulated changes in populations of competing whitetip reef sharks (WTS) and blacktip sharks (BTS) in an area (from the model in Module 4.1). We can summarize these values in two vectors, $\mathbf{w} = (20.00, 6.57, 4.69, 3.08, 0.99, 0.02)$, or $\begin{bmatrix} 20.00 \\ 6.57 \\ 4.69 \\ 3.08 \\ 0.99 \\ 0.02 \end{bmatrix}$, and $\mathbf{b} = (15.00, 5.37, 4.84, 6.00, 10.83, 27.43)$, or $\begin{bmatrix} 15.00 \\ 5.27 \\ 4.84 \\ 6.00 \\ 10.83 \\ 27.43 \end{bmatrix}$, for WTS and BTS, respectively. We use boldface, such as **w**, or a line over the letter, such as $\overline{w}$, to indicate a vector. Each of the vectors **w** and **b** has six numbers, or **elements,** or **members**, so the **size** of each is 6. A subscript, or **index** (plural **indices**), indicates the particular item of the vector, and indices begin with 1 or 0. For a starting index of 1, $w_1 = 20.00$ is the number of whitetip sharks at month 0. By month 5, the simulated population dwindles to 0.02, which is w_6. Another advantage of vectors is the ability

Table 13.2.1
Simulated Populations

Time (months)	*WTS*	*BTS*
0	20.00	15.00
1	6.57	5.37
2	4.69	4.84
3	3.08	6.00
4	0.99	10.83
5	0.02	27.43

to use a variable like i as a subscript instead of a constant like 6. In a computational tool, we can employ this index to change values in a loop, allowing us to perform the same operations on all array elements. In mathematics, we can employ a variable index to express a general case.

Definitions A **vector v** is an ordered n-tuple, written as a row or a column,

$$\mathbf{v} = (v_1, v_2, \ldots, v_n) = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

where $v_1, v_2, \ldots, v_n$ are numbers, called **elements**, or **members**. The **size** of a vector is the number of elements, here n. A subscript is an **index** (plural **indices**), and in vector notation, indices begin with 1 or 0.

Quick Review Question 1

For $\mathbf{b} = (15.00, 5.37, 4.84, 6.00, 10.83, 27.43)$, where indices begin with 1, give the value of b_4.

A vector **equal** to **w** has size 6 and its numbers are identical to and in the same order as those of **w**. Thus, two vectors are equal if and only if they are of the same size and corresponding elements are equal.

Definition Vectors $(x_1, x_2, \ldots, x_n)$ and $(y_1, y_2, \ldots, y_n)$ are **equal** if and only if $x_i = y_i$ for $i = 1, 2, \ldots, n$.

Vector Addition

Suppose the only sharks in the given area are BTS and WTS. To obtain the total number of sharks each month (vector **s**), we add corresponding values of the vectors element by element, as follows:

$$\mathbf{s} = \mathbf{w} + \mathbf{b} = \begin{bmatrix} 20.00 \\ 6.57 \\ 4.69 \\ 3.08 \\ 0.99 \\ 0.02 \end{bmatrix} + \begin{bmatrix} 15.00 \\ 5.27 \\ 4.84 \\ 6.00 \\ 10.83 \\ 27.43 \end{bmatrix} = \begin{bmatrix} 20.00 + 15.00 \\ 6.57 + 5.27 \\ 4.69 + 4.84 \\ 3.08 + 6.00 \\ 0.99 + 10.83 \\ 0.02 + 27.43 \end{bmatrix} = \begin{bmatrix} 35.00 \\ 11.94 \\ 9.53 \\ 9.08 \\ 11.82 \\ 27.45 \end{bmatrix}$$

For instance, initially, the number of sharks is $s_1 = b_1 + w_1 = 20.00 + 15.00 = 35.00$, or the sum of the number of BTS and the number of WTS at the start of the simulation. The two vectors must be of the same size for their sum to make sense.

Definition Let $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ be vectors of n elements each. The **sum** of **x** and **y** is the vector

$$\mathbf{x} + \mathbf{y} = (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n).$$

Quick Review Question 2

Suppose two scientists, Drs. Chang and Morris, are leading research teams studying red-footed boobies on two islands in Galapagos. Each group counts the number of eggs, hatchlings, juveniles, and nesting pairs over a 1-week period. Suppose the values for Dr. Chang's team are 35, 16, 240, and 351 and for Dr. Morris' team are 18, 10, 103, and 153, respectively.

a. Express the values for Dr. Chang's team in a vector, **c**, and for Dr. Morris' team in a vector, **m**.
b. Compute **t** = **c** + **m**.
c. What does **t** represent?

Multiplication by a Scalar

Suppose that each member of the vectors **w** and **b** is in hundreds of sharks. In this case, $w_1 = 20.00$ indicates that the initial number of whitetip sharks is $100 \cdot 20.00 = 2000$ WTS. To carry the process through every month, we use **scalar multiplication**, as follows:

$$100(20.00, 6.57, 4.69, 3.08, 0.99, 0.02) = (2000, 657, 469, 308, 99, 2)$$

We multiply the **scalar**, or number, 100 by each element.

Definitions A **scalar** is a real number. Let $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ be a vector. The **scalar product** of a scalar a and the vector **x** is the vector

$$a\mathbf{x} = a(x_1, x_2, \ldots, x_n) = (ax_1, ax_2, \ldots, ax_n).$$

Quick Review Question 3

The scientists in Quick Review Question 2 estimate that the actual numbers of boobies in each category to be 1.1 as many as they observed. The vector of data for Dr. Chang's team is **c** = (35, 16, 240, 351).

a. Using 1.1 and the variable name **c**, give the expression for the vector of estimated values.
b. Give the vector with values rounded to the nearest integers.

Dot Product

As part of the effort to keep them from extinction, scientists around the world have studied the magnificent green sea turtle and used mathematics and computer science to make predictions about their populations. We deal with a different type of multiplication in estimating the number of eggs laid by Hawaiian green sea turtles in one year. We can consider their life cycle to be in five stages, with egg layers in two stages, novice breeders of age 25 years, and mature breeders from ages 26 through 50 years. On the average, a novice breeder lays 280 eggs in a year, and a mature breeder lays 70 eggs per year. We can combine these data in a vector **e** = (280, 70). Suppose also that there are 291 novice and 9483 mature breeders, which we store in the vector **b** = (291, 9483). To approximate the total green sea turtle egg production in a year, we multiply together corresponding terms and add the results, as follows:

$$\begin{aligned} \mathbf{e} \cdot \mathbf{b} &= (280, 70) \cdot (291, 9483) \\ &= 280 \cdot 291 + 70 \cdot 9483 \\ &= 81{,}480 + 663{,}810 \\ &= 745{,}290 \text{ eggs} \end{aligned}$$

This type of multiplication, the **dot product**, involves two vectors of the same size and results in a *number*, *not* another vector (Green Sea Turtle).

Definition Let $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ be vectors of n elements each. The **dot product** (or **scalar product**, or **inner product**) of **x** and **y** is

$$\mathbf{x} \cdot \mathbf{y} = x_1 \cdot y_1 + x_2 \cdot y_2 + \cdots + x_n \cdot y_n.$$

Often in writing the dot product, the first term is written as a row vector, such as (280, 70), while the second is written as a column vector, such as $\begin{bmatrix} 291 \\ 9483 \end{bmatrix}$. Multiplication of elements follows the arrows, elements from left to right corresponding to elements from top down:

$$\mathbf{e} \cdot \mathbf{b} = \overrightarrow{(280,\ 70)} \cdot \begin{bmatrix} 291 \\ 9483 \end{bmatrix} \downarrow = 280 \cdot 291 + 70 \cdot 9483 = 745{,}290 \text{ eggs}$$

This notation will be useful for computations we will be doing shortly.

Quick Review Question 4

The first stage in the life of the Hawaiian green sea turtle, consisting of eggs and hatchlings, occurs during the first year. Stage 2, juveniles, extends from year 1 to year 16. Suppose 23% of the hatchlings survive and move to stage 2, while 67.9% of those in Stage 2 remain in that stage each year. In one year, suppose Stage 1 has 808,988 individuals, and Stage 2 has 715,774 (Green Sea Turtle).

a. Give a vector, **p**, with real-number elements representing the percentages.
b. Give a vector, **s**, storing the individuals in Stages 1 and 2.
c. Using variables **p** and **s**, not the data, give the vector operation to determine the number of individuals that will be in Stage 2 the following year.
d. Calculate this value.

Matrices

In the section "Vectors," we considered the data structure of a one-dimensional array, or vector. One example involved vector **w**, which stored under that one name the simulated number of whitetip sharks from 0 through 5 months. Quite often, however, more features need to be stored and manipulated. In such cases 2D arrays may be helpful. For example, we can store the data from Table 13.2.1 for the number of whitetip reef sharks (WTS) and blacktip sharks (BTS) in the following 2D array:

$$S = \begin{bmatrix} 20.00 & 15.00 \\ 6.57 & 5.27 \\ 4.69 & 4.84 \\ 3.08 & 6.00 \\ 0.99 & 10.83 \\ 0.02 & 27.43 \end{bmatrix}$$

The name used in mathematics and in many computational tools for a 2D array is matrix. A **matrix** (plural, **matrices**) is a rectangular array of numbers, and we can think of a matrix as a table of numbers.

The symbol for an individual matrix element has two subscripts indicating its row and column. Assuming that the row and column indices for S shown previously begin with 1, the value 3.08, which is the number of WTS at month 3, is element s_{41}, the value in the fourth row and first column. Usually, we represent a matrix with an uppercase letter and an element with the corresponding lowercase letter and indices. Thus, we can abbreviate the array as $S = [s_{ij}]$. The **size** of a matrix is the number of rows by the number of columns. Thus, S is a 6×2, or a 6-by-2, matrix.

Definitions A **matrix** (plural, **matrices**), $S = [s_{ij}]$, is a rectangular array of numbers. Element s_{ij} is in row i and column j. A matrix with m rows and n columns has **size $m \times n$**, or **m by n**.

Quick Review Question 5

For matrix S, assume the row and column indices begin with 1. Give each of the following.

a. The value of s_{12}
b. The notation for the element with value 6.00

Scalar Multiplication and Matrix Sums

As with vectors, two matrices are equal if they have the same size and corresponding elements are identical. To compute the sum of two matrices that have the same size, we add corresponding elements. For the product of a scalar times a matrix, we multiply each element by the scalar.

Definitions Let $A = [a_{ij}]$ and $B = [b_{ij}]$ be two $m \times n$ matrices. A and B are **equal** if and only if $a_{ij} = b_{ij}$ for all i and j. The **product of scalar c and matrix A** is $m \times n$ matrix

$$cA = [ca_{ij}]$$

That is, each element of A is multiplied by c. The **matrix sum** of A and B is an $m \times n$ matrix

$$A + B = [a_{ij} + b_{ij}]$$

That is, corresponding elements are added.

Quick Review Question 6

For the following, calculate the indicated matrices:

$$A = \begin{bmatrix} 1 & 3 & 9 \\ 0 & 5 & 6 \end{bmatrix},\ B = \begin{bmatrix} -1 & 2 & 0 \\ 1 & 4 & 3 \end{bmatrix},\ \text{and } C = \begin{bmatrix} 7 & 4 \\ -2 & 8 \end{bmatrix}$$

a. $3B$
b. $A + B$
c. $C + 2A + B$

Matrix Multiplication

The ability to multiply matrices allows us to model many problems, including those involving changes in populations. The foundation for the operation of matrix multiplication is the concept of dot product of vectors. Suppose we wish to estimate the total shark mass at each month of the simulation involving whitetip reef sharks (WTS) and blacktip sharks (BTS). An estimate of the average mass of a whitetip is 20 kg, while that of a blacktip is 18 kg. Suppose we started the simulation with 20 WTS and 15 BTS. Thus, the initial total shark mass is the following dot product:

$$(20, 15) \cdot (20, 18) = 20 \cdot 20 + 15 \cdot 18 = 670 \text{ kg}$$

As pointed out before, we can write the second vector as a column vector. Then, we are actually multiplying a 1×2 matrix by a 2×1 matrix to find a 1×1 matrix, as follows:

$$\begin{bmatrix} 20 & 15 \end{bmatrix} \begin{bmatrix} 20 \\ 18 \end{bmatrix} = [670]$$

To compute the shark mass totals at each month, we multiply the shark matrix S by the mass vector $\mathbf{g} = (20, 18)$, written as a column. We take the dot product of each row of S with $\mathbf{g}$ to compute a 6×1 matrix of monthly masses (kg) rounded to the nearest integer, as follows:

$$\begin{array}{l} \\ \text{month 0} \\ \text{month 1} \\ \text{month 2} \\ \text{month 3} \\ \text{month 4} \\ \text{month 5} \end{array}
\begin{array}{c} \begin{array}{cc} \text{WTS} & \text{BTS} \end{array} \\ \begin{bmatrix} 20.00 & 15.00 \\ 6.57 & 5.27 \\ 4.69 & 4.84 \\ 3.08 & 6.00 \\ 0.99 & 10.83 \\ 0.02 & 27.43 \end{bmatrix} \end{array}
\begin{array}{c} \text{mass} \\ \begin{bmatrix} 20 \\ 18 \end{bmatrix} \\ \\ \\ \\ \\ \end{array}
\begin{array}{l} \\ \text{WTS} \\ \text{BTS} \\ \\ \\ \\ \\ \end{array}
=
\begin{bmatrix} 20.00 \cdot 20 + 15.00 \cdot 18 \\ 6.57 \cdot 20 + 5.27 \cdot 18 \\ 4.69 \cdot 20 + 4.84 \cdot 18 \\ 3.08 \cdot 20 + 6.00 \cdot 18 \\ 0.99 \cdot 20 + 10.83 \cdot 18 \\ 0.02 \cdot 20 + 27.43 \cdot 18 \end{bmatrix}
=
\begin{array}{c} \text{mass} \\ \begin{bmatrix} 670 \\ 228 \\ 181 \\ 170 \\ 215 \\ 494 \end{bmatrix} \end{array}
\begin{array}{l} \\ \text{month 0} \\ \text{month 1} \\ \text{month 2} \\ \text{month 3} \\ \text{month 4} \\ \text{month 5} \end{array}$$

As we move from left to right on a row of the first matrix, we go down on the second, multiplying corresponding elements and then adding the results. Consequently, the number of columns in the first matrix must equal the number of rows in the column vector. The resulting vector has the same number of rows as the first matrix and the same number of columns as the second. In this example, S has size 6×2 while $\mathbf{g}$ has size 2×1, and the result is a 6×1 matrix.

Quick Review Question 7

a. For the vector $\mathbf{v} = (5, 0, -1)$, written as a column vector, and the matrix $A = \begin{bmatrix} 1 & 3 & 9 \\ 0 & 5 & 6 \end{bmatrix}$, calculate $A\mathbf{v}$.

b. For a 5×8 matrix B, give the size of a vector $\mathbf{w}$ for which we can calculate $B\mathbf{w}$.

c. Give the resulting size of $B\mathbf{w}$.

Suppose scientists observed that 25% of the WTS have wounds, while none of the BTS do. Such wounds could contribute to an animal's decreased hunting ability. In calculating the total number of wounded sharks, we need to consider only the WTS. Again, the computation can be accomplished with a dot product. At the start of the simulation, we have the following computation:

$$\begin{bmatrix} 20.00 & 15.00 \end{bmatrix} \begin{bmatrix} 0.25 \\ 0.00 \end{bmatrix} = [5.00]$$

Zero in the second row eliminated the effect of the number of BTS.

Additionally, suppose scientists noted that 30% of the WTS and 20% of the BTS have lesions. The total number of sharks with lesions at a particular month is the dot product of a vector of the numbers of sharks and a column vector with these percentages. For example, as the following computation shows, 9 sharks have lesions in month 0:

$$\begin{bmatrix} 20.00 & 15.00 \end{bmatrix} \begin{bmatrix} 0.30 \\ 0.20 \end{bmatrix} = [9.00]$$

Certainly, we can take the shark-numbers matrix, S, and multiply by any 2×1 attribute vector, such as $\begin{bmatrix} 20 \\ 18 \end{bmatrix}$, $\begin{bmatrix} 0.25 \\ 0.00 \end{bmatrix}$, or $\begin{bmatrix} 0.30 \\ 0.20 \end{bmatrix}$. However, we can perform all these calculations together. We form a 2×3 attribute matrix, $A = \begin{bmatrix} 20 & 0.25 & 0.30 \\ 18 & 0.00 & 0.20 \end{bmatrix}$, and we dot each row of the first matrix (S) by each column of a second matrix (attribute matrix, A) to get a resulting totals matrix (T) containing the totals for shark mass, wounded sharks, and sharks with lesions by month. There are six months, thus six rows, one for each month, in the resulting totals matrix. Because there are three attributes to total, the totals matrix has three columns, or three totals, for each month. Six rows in the first matrix along with three columns in the second yield a 6×3 totals matrix, as follows:

$$SA = \begin{array}{l} \\ \text{month 0} \\ \text{month 1} \\ \text{month 2} \\ \text{month 3} \\ \text{month 4} \\ \text{month 5} \end{array}
\begin{array}{c} \begin{array}{cc} \text{WTS} & \text{BTS} \end{array} \\ \begin{bmatrix} 20.00 & 15.00 \\ 6.57 & 5.27 \\ 4.69 & 4.84 \\ 3.08 & 6.00 \\ 0.99 & 10.83 \\ 0.02 & 27.43 \end{bmatrix} \end{array}
\begin{array}{c} \begin{array}{ccc} \text{mass} & \text{\% wounded} & \text{\% lesions} \end{array} \\ \begin{bmatrix} 20 & 0.25 & 0.30 \\ 18 & 0.00 & 0.20 \end{bmatrix} \\ \\ \\ \\ \\ \end{array}
\begin{array}{l} \\ \text{WTS} \\ \text{BTS} \\ \\ \\ \\ \\ \end{array}$$

$$
= T = \begin{array}{l} \\ \text{month 0} \\ \text{month 1} \\ \text{month 2} \\ \text{month 3} \\ \text{month 4} \\ \text{month 5} \end{array}
\begin{array}{c} \begin{array}{ccc} \text{mass (kg)} & \text{\# wounded} & \text{\#lesioned} \end{array} \\
\begin{bmatrix} 670 & 5.00 & 9.00 \\ 228 & 1.64 & 3.03 \\ 181 & 1.17 & 2.38 \\ 170 & 0.77 & 2.12 \\ 215 & 0.25 & 2.46 \\ 494 & 0.005 & 5.49 \end{bmatrix} \end{array}
$$

Usually we write the matrix product without row and column headings, as follows:

$$
SA = \begin{bmatrix} 20.00 & 15.00 \\ 6.57 & 5.27 \\ 4.69 & 4.84 \\ 3.08 & 6.00 \\ 0.99 & 10.83 \\ 0.02 & 27.43 \end{bmatrix} \begin{bmatrix} 20 & 0.25 & 0.30 \\ 18 & 0.00 & 0.20 \end{bmatrix} = \begin{bmatrix} 679 & 5.00 & 9.00 \\ 228 & 1.64 & 3.03 \\ 181 & 1.17 & 2.38 \\ 170 & 0.77 & 2.12 \\ 215 & 0.25 & 2.46 \\ 494 & 0.005 & 5.49 \end{bmatrix} = T
$$

In the totals matrix, T, the third-row, first-column element ($t_{31} = 181$) indicates that at month 2 of the simulation, the total mass of WTS and BTS in the area is 181 kg. The rounded second-row, second-column element ($t_{22} = 1.64$, rounded to 2) indicates that in month 1, two of the sharks have wounds. The rounded sixth-row, third-column element ($t_{63} = 5.49$, rounded to 5) shows that five of the sharks have lesions in month 5. For the dot products to be possible, the number of columns in the first matrix (here S) and the number of rows in the second (here A) have to be identical; here both are 2.

Definition Let $A = [a_{ij}]_{m \times q}$ be an $m \times q$ matrix and $B = [b_{ij}]_{q \times n}$ a $q \times n$ matrix. The **matrix product** of A and B is an $m \times n$ matrix AB, or $A \cdot B = C = [c_{ij}]_{m \times n}$, where c_{ij} is the dot product of the ith row of A and the jth column of B.

Quick Review Question 8

Consider the following matrices:

$$
A = \begin{bmatrix} 8 & 5 & 3 & -4 \\ -5 & 1 & 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 2 & -6 \\ 7 & 1 \\ 4 & 3 \\ -9 & -2 \end{bmatrix}, \quad C = \begin{bmatrix} 6 & 5 \\ 1 & -3 \\ 2 & -8 \end{bmatrix}, \quad I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \text{ and}
$$

$$
I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}
$$

Evaluate each of the following:

a. AB
b. BA
c. CI_2
d. I_3C

Square Matrices

Each of I_2 and I_3 in Parts c and d of the last Quick Review Question is a **square matrix**, having the same number of rows as columns. Moreover, each has 1s along its **diagonal**, the line of elements from the top left to the bottom right. A number of examples in biology involve square matrices. For example, the hypothetical data in Table 13.2.2 presents the distribution of ABO blood types for mothers and newborns (multiple births omitted) in a county over a year. The corresponding matrix is as follows:

$$
M = \begin{bmatrix} \mathbf{1068} & 53 & 68 & 516 \\ 37 & \mathbf{273} & 88 & 601 \\ 60 & 58 & \mathbf{21} & 0 \\ 491 & 189 & 0 & \mathbf{2059} \end{bmatrix}
$$

According to the datum in the second row, first column, only 37 mothers with type B blood gave birth to a child with type A blood in that county during the year of the study. The diagonal values, which are in boldface, indicate the numbers of mother-newborn pairs that share the same blood type. For instance, 1068 type A mothers gave birth to type A children in the county that year.

Definitions An $n \times n$ matrix is called a **square matrix**. In an $n \times n$ square matrix M, the **diagonal** is the set of elements $\{m_{11}, m_{22}, \ldots, m_{nn}\}$.

As another example involving a square matrix, Table 13.2.3 presents similarity measures (specifically, Euclidean distances) of the 18S rRNA sequences of pairs of animals, where smaller numbers indicate closer relationships. Thus, with a Euclid-

Table 13.2.2
Hypothetical Data for the Distribution of ABO Blood Types of Mothers and Newborns (Multiple Births Omitted) in a County Over a Year. (Similar to Table 1 in Bottini et al. 2001)

Mother\Newborn	*A*	*B*	*AB*	*O*
A	1068	53	68	516
B	37	273	88	601
AB	60	58	21	0
O	491	189	0	2059

Table 13.2.3
Similarity Measures (Specifically, Euclidean Distances) of the 18S rRNA Sequences of Pairs of Animals (Lockhart et al. 1994, Table 3)

	Frog	*Bird*	*Human*	*Rabbit*
Frog	0	0.316	0.350	0.336
Bird	0.316	0	0.130	0.102
Human	0.350	0.130	0	0.028
Rabbit	0.336	0.102	0.028	0

ean distance of 0.028, the rRNA sequences of a human and a rabbit are more closely related than that of a human and a frog (distance 0.350). Because the distance from animal A's rRNA sequence to animal B's sequence is the same as the distance from B to A, the table and the corresponding matrix, which follows, are **symmetric** around the diagonal:

$$M = \begin{bmatrix} 0 & 0.316 & 0.350 & 0.336 \\ 0.316 & 0 & 0.130 & 0.102 \\ 0.350 & 0.130 & 0 & \mathbf{0.028} \\ 0.336 & 0.102 & \mathbf{0.028} & 0 \end{bmatrix}$$

Thus, as the boldface emphasizes, the distance in row 3, column 4, namely, 0.028, is the same as the number in row 4, column 3. In general, elements $m_{ij} = m_{ji}$. For symmetry, the values on the diagonal do not have to be zero as they are in this example.

Definition An $n \times n$ square matrix M is **symmetric** if $m_{ij} = m_{ji}$ for all i and j.

Matrices and Systems of Equations

The section "Dot Product" indicates that a Hawaiian green sea turtle novice breeder lays an average of 280 eggs per year, while a mature breeder only lays 70. We considered a specific number of turtles in each category, 291 and 9483, respectively, and calculated the total yearly egg production as the following dot product:

$$\mathbf{e} \cdot \mathbf{b} = (280, 70) \cdot (291, 9483)$$

Instead of specifying the number of turtles in each category, let n be the number of novice breeders and m the number of mature breeders with $\mathbf{b} = (n, m)$. In general, the average annual egg production, a, is computed as follows:

$$\mathbf{e} \cdot \mathbf{b} = (280, 70) \cdot (n, m)$$
$$= 280n + 70m = a$$

Thus, the dot product translates into one side of a linear equation.

The following are examples of linear equations:

$$6x = 1$$
$$5x + 7y = 3$$
$$-2x + \pi y + \sqrt{3}\,z = 9$$
$$1/2x_1 + 33.2x_2 + 15x_3 + 13x_4 = 33/4$$

The equations derive their name from the fact that when they have only one, two, or three variables, as in the first three examples, their graphs are straight lines. The general **linear equation** is

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = c$$

where a_i and c are numbers for $i = 1, 2, \ldots, n$.

While we can employ a dot product in representing one linear equation, we can use matrix multiplication for a system of linear equations. Returning to the example involving whitetip sharks (WTS) and blacktip sharks (BTS) from the section "Matrix Multiplication," suppose the number of each kind of shark by month is as in Table 13.2.1. That section represented the data in the following matrix, with the number of WTS in the first column and the number of BTS in the second:

$$S = \begin{bmatrix} 20.00 & 15.00 \\ 6.57 & 5.27 \\ 4.69 & 4.84 \\ 3.08 & 6.00 \\ 0.99 & 10.83 \\ 0.02 & 27.43 \end{bmatrix}$$

Let x be the percentage of whitetip sharks with lesions, y be the percentage of blacktip sharks with lesions, and $\mathbf{h} = (x, y)$. Suppose the total number of sharks with lesions from month 0 through 5 is 9.00, 3.04, 2.38, 2.12, 2.46, and 5.49, respectively, with vector representation $\mathbf{v} = (9.00, 3.04, 2.38, 2.12, 2.46, 5.49)$. Thus, we have the following linear equation for the total number of sharks with lesions in month 0:

$$20.00x + 15.00y = 9.00$$

which we can write as the following dot product:

$$(20.00, 15.00) \cdot (x, y) = 9.00$$

or

$$[20.00, 15.00] \cdot \begin{bmatrix} x \\ y \end{bmatrix} = [9.00]$$

Quick Review Question 9

Use the preceding shark data for month 1.

a. Write the linear equation.

b. Write the corresponding equation using a dot product.

Instead of writing each equation separately, we can employ matrix multiplication to system of six equations, as follows:

$$\begin{bmatrix} 20.00 & 15.00 \\ 6.57 & 5.27 \\ 4.69 & 4.84 \\ 3.08 & 6.00 \\ 0.99 & 10.83 \\ 0.02 & 27.43 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 9.00 \\ 3.04 \\ 2.38 \\ 2.12 \\ 2.46 \\ 5.49 \end{bmatrix}$$

or

$$S\mathbf{h} = \mathbf{v}$$

As we see in other modules, besides providing a useful abbreviation for systems of equations, matrices can simplify the process of finding solutions.

Quick Review Question 10

Express the following system of equations using a matrix-vector notation:

$$\begin{cases} 2x - y = 7 \\ 6x \quad\;\; = 5 \end{cases}$$

Exercises

Given the scalars a = 7 and b = 3 and the vectors $\mathbf{u} = (3, -4, 8, 0)$, $\mathbf{v} = (-9, 4, 21, 2)$, $\mathbf{y} = (8, 8, 1, -2)$, *and* $\mathbf{x} = (7, 17, 6)$, *where possible, compute the values of Exercises 1–20. Check your work with a computational tool.*

1. $a\mathbf{u}$
2. $b\mathbf{v}$
3. $a\mathbf{u} + b\mathbf{v}$
4. $\mathbf{u} + \mathbf{v}$
5. $\mathbf{v} + \mathbf{u}$
6. $(\mathbf{u} + \mathbf{v}) + \mathbf{y}$
7. $\mathbf{u} + (\mathbf{v} + \mathbf{y})$
8. $\mathbf{u} + \mathbf{x}$
9. $(a + b)\mathbf{y}$
10. $a\mathbf{y} + b\mathbf{y}$
11. $0\mathbf{x}$
12. $\mathbf{u} \cdot \mathbf{v}$
13. $\mathbf{y} \cdot (2\mathbf{v})$
14. $2(\mathbf{y} \cdot \mathbf{v})$
15. $(2\mathbf{y}) \cdot \mathbf{v}$
16. $\mathbf{x} \cdot \mathbf{y}$
17. $\mathbf{v} - \mathbf{y}$
18. $a(\mathbf{u} + \mathbf{y})$
19. $a\mathbf{u} + a\mathbf{y}$
20. $(0, 0, 0) \cdot \mathbf{x}$

Compute, if possible, the dot products in Exercises 21–23. Check your work with a computational tool.

21. $(5,7) \cdot \begin{bmatrix} -1 \\ 4 \end{bmatrix}$
22. $(6,2,3) \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix}$
23. $(7,-7,1) \cdot \begin{bmatrix} 3 \\ 3 \\ 1 \end{bmatrix}$

24. Suppose the following items are for sale one week by a scientific supply house at the indicated prices: a particular bacterial culture, \$17; case of pipettes, \$310; case of Petri dishes, \$190; case of beakers, \$40.
   a. Write these prices in a vector $\mathbf{v}$.
   b. Suppose there is a 25%-off sale. What scalar is multiplied by $\mathbf{v}$ to give the sale prices?

   c. Perform this scalar multiplication.
   d. Suppose 83 bacterial cultures, 18 cases of pipettes, 145 cases of Petri dishes, and 108 cases of beakers are sold during the sale. Write a dot product of vectors to calculate the amount of money from the sale, and evaluate this dot product.
   e. Suppose the next week the store sells 20 bacterial cultures, 3 cases of pipettes, 76 cases of Petri dishes, and 37 cases of beakers. Write the vector sum to indicate the number of each item sold during the 2-week period, and evalulate this addition.

Determine the values of the unknowns to make the vectors equal in Exercises 25–27.

25. $(3, 5, 7) = (a, b, 7)$
26. $(-6, 2, 1) = (-6, 2, 1, a)$
27. $2(6, 1, a) = b(3, c, 4)$
28. Consider the matrix $A = [a_{ij}] = \begin{bmatrix} 6 & 3 & -2 \\ 0 & -8 & 4 \end{bmatrix}$
   a. What is A's size?
   b. Find a_{21}, a_{12}, a_{31}, and a_{13}.

Using $A = \begin{bmatrix} 6 & 3 & -2 \\ 0 & -8 & 4 \end{bmatrix}$, $B = \begin{bmatrix} -1 & 2 & 3 \\ -7 & 2 & 1 \end{bmatrix}$, *and* $C = \begin{bmatrix} 0 & -4 & 1 \\ 3 & 1 & -8 \end{bmatrix}$, *calculate the matrices in Exercises 29–40. Check your work with a computational tool.*

29. $3A$
30. $3B$
31. $3A + 3B$
32. $A + B$
33. $3(A + B)$ (Compare to Exercise 31.)
34. $B + A$ (Compare to Exercise 32.)
35. $-A$
36. $B + C$
37. $(A + B) + C$ (Use Exercise 32.)
38. $A + (B + C)$ (Use Exercise 36; compare to Exercise 37.)
39. $2(3A)$ (Use Exercise 29.)
40. $6A$ (Compare to Exercise 39.)

If possible, compute the matrices in Exercises 41–43 using $A = [3 \quad 5]$ *and* $B = \begin{bmatrix} 1 & 1 \\ 3 & 4 \\ -1 & 4 \end{bmatrix}$. *Check your work with a computational tool.*

41. $2A$
42. $A + A$
43. $A + B$
44. How many elements are in matrix A if it is of each given size?
   a. 20×5
   b. $m \times n$
   c. 5×5
   d. $n \times n$
45. Consider the square matrix

$$A = \begin{bmatrix} 1 & 5 & 3 \\ 5 & 4 & 6 \\ 3 & 6 & 8 \end{bmatrix} = [a_{ij}]_{3\times 3}.$$

   a. Find a_{21} and a_{12}.
   b. Why is A symmetric?
   c. Give the diagonal elements of A.
   d. Suppose B is a symmetric matrix. Fill in the blanks.

$$B = \begin{bmatrix} -7 & - & - & - \\ 2 & 3 & - & - \\ -1 & 4 & -4 & - \\ 6 & 5 & 0 & -3 \end{bmatrix}$$

46. A lab is using spectrophotometer to indicate the number of bacteria in a broth. From a reading, they determine absorbance, a value between 0.0 and 2.0. As the number of bacteria increases, so does the absorbance. Each team takes measurements at 10-min intervals. Suppose following measurements are made for *E. coli* at 15 °C from 70 min to 130 min: 0.041, 0.055, 0.064, 0.062, 0.089, 0.097, 0.103. The following measurements are for *E. coli* at 21 °C: 0.055, 0.070, 0.077, 0.095, 0.105, 0.115, 0.124. Place the values in a matrix, and indicate the meanings of the rows and columns (Johnson and Case 2009).

47. Suppose a certain animal has a maximum life span of 3 years. This example predicts populations in each age category: year 1 (0–1 year), year 2 (1–2 years), and year 3 (2–3 years). We consider only females. A year 1 female animal has no offspring; a year 2 female has 3 daughters on the average; and a year 3 female has a mean of 2 daughters. A year 1 animal has a 0.3 probability of living to year 2. A year 2 animal has a 0.4 probability of living to year 3. Suppose at one instance, the number of year 1, 2, and 3 females are 2030, 652, and 287, respectively.

a. Write a row vector of three elements giving the mean number of female offspring in each age category.

b. Write a row vector triple giving the probabilities that in the next year a year 1 animal lives to years 2, 3, and 4.

c. Write a row vector triple giving the probabilities that in the next year a year 2 animal lives to years 2, 3, and 4.

d. Place the row vectors from Parts a–c in a matrix, *L*.

e. Write a column vector, **c**, of the female counts in each year.

f. Using Parts d and e, estimate the female numbers in each age category a year later.

g. Using Parts d and f, estimate the female numbers in each age category 2 years after the initial counts.

h. Using Part d, calculate L^2, or *LL*.

i. Using Parts h and e, calculate $L^2\mathbf{c}$.

j. How do your answers from Parts g and i compare?

48. Consider the matrix

$$T = \begin{bmatrix} 0 & 50 & 20 \\ 100 & 150 & 120 \\ 90 & 170 & 200 \end{bmatrix}.$$

For any 3×3 matrix *M* with elements from the set of nonnegative integers, apply the function *f* to each element, defined as

$$f(m_{ij}) = \begin{cases} 0, & \text{if } m_{ij} < \text{corresponding threshold value, } t_{ij} \\ 1, & \text{if } m_{ij} \geq \text{corresponding threshold vlaue, } t_{ij}. \end{cases}$$

T is called a **threshold matrix**, and each t_{ij} is a threshold value. An element of the matrix *M* is mapped to 1 if and only if it is at least as big as the corresponding threshold value. Fill in the blanks for the image of the elements of the following matrix *M*:

$$M = \begin{bmatrix} 110 & 112 & 100 \\ 100 & 70 & 75 \\ 90 & 80 & 90 \end{bmatrix} \rightarrow \begin{bmatrix} - & 1 & - \\ - & 0 & - \\ - & - & - \end{bmatrix}.$$

49. A dither matrix can be used to enhance a digital image, such as a medical image from a CT (computerized tomography) scan of the body. A computer can analyze the degree of grayness of each dot, or **pixel**, of one such black-and-white image and assign it a value for intensity, say from 0 (white) to 255 (black). One method of enhancing the picture is **dithering**. Each digitized pixel is compared with an individual threshold value to determine if a dot will or will not be placed at that point on the reconstructed picture. There are no gray dots in the reconstructed picture; a black dot is either present or not present at each position, depending on the presence of a 1 or 0 in the corresponding position of the final matrix. To accomplish this procedure a threshold matrix, called a **dither matrix**, is needed. Much experimentation has been done in dithering to find the best threshold matrix to help produce a clear, apparently continuous image using black dots on a white background. The construction of one dither matrix is presented in this problem. Let

$$D_2 = \begin{bmatrix} 0 & 2 \\ 3 & 1 \end{bmatrix} \text{ and } V_2 = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

To develop the dither matrix, find the following matrices:

a. $4D_2$ **b.** $4D_2 + 2V_2$ **c.** $4D_2 + 3V_2$ **d.** $4D_2 + V_2$

e. Construct the 4×4 matrices D_4 and V_4 with the 2×2 matrices from the previous parts placed in the indicated positions.

$$D_4 = \begin{bmatrix} 4D_2 & 4D_2 + 2V_2 \\ 4D_2 + 3V_2 & 4D_2 + V_2 \end{bmatrix}, \quad V_4 = \begin{bmatrix} V_2 & V_2 \\ V_2 & V_2 \end{bmatrix}$$

f. Calculate the dither matrix $16D_4 + 8V_4$, which is the threshold matrix that will be used in reconstructing a picture below.

g. Consider the 4×4 matrix *M* containing pixel intensities transmitted from space.

$$M = \begin{bmatrix} 100 & 145 & 100 & 178 \\ 111 & 60 & 250 & 102 \\ 40 & 200 & 20 & 73 \\ 254 & 198 & 223 & 204 \end{bmatrix}.$$

With function *f* defined as in Exercise 48, find the image of *M* after applying *f* to every point.

h. Draw the picture in a 4×4 array. *Note:* If the picture were larger, we could use the same dither matrix by applying that threshold matrix in a checkerboard fashion over the entire picture.

Let $A=\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, $B=\begin{bmatrix} 6 & 2 & 0 \\ 0 & -1 & 4 \end{bmatrix}$, $C=\begin{bmatrix} -2 & 1 & -5 \\ 7 & 1 & 0 \end{bmatrix}$. *Where possible, perform the indicated operation or answer the question in each of the Exercises 50–63. Check your work with a computational tool.*

50. AB **51.** AC **52.** $AB + AC$ (Use Exercises 50 and 51.)
53. $B + C$ **54.** $A(B + C)$ (Use Exercise 53. Compare to Exercise 52.)
55. BA **56.** BC **57.** $3(AC)$ (Use Exercise 51.)
58. $3A$ **59.** $(3A)C$ (Use Exercise 58. Compare to Exercise 57.)
60. $A^2 = A \cdot A$ **61.** B^2 **62.** $\mathbf{0}_{2\times 2} \cdot A$, where $\mathbf{0}_{2\times 2}$ is a 2×2 matrix of all zeros.
63. $B \cdot \mathbf{0}_{3\times 3}$, where $\mathbf{0}_{3\times 3}$ is a 3×3 matrix of all zeros.

For Exercises 64–67, perform the indicated matrix multiplication. Check your work with a computational tool.

64. $\begin{bmatrix} 0.1 & 0.2 & 0.9 \\ 1.3 & 0.5 & 0.7 \end{bmatrix}\begin{bmatrix} 2.2 & 3.9 \\ 0.6 & 0.4 \\ 1.1 & 2.8 \end{bmatrix}$ **65.** $\begin{bmatrix} 1 & -3 \\ 8 & 5 \\ 0 & 7 \\ 2 & -2 \end{bmatrix}\begin{bmatrix} -6 & 5 & 5 \\ -5 & 1 & 0 \end{bmatrix}$

66. $\begin{bmatrix} 9 & 0 \\ 3 & 2 \end{bmatrix}\begin{bmatrix} -6 & 10 & 3 & 8 \\ -3 & 11 & 20 & 7 \end{bmatrix}$ **67.** $\begin{bmatrix} 0.4 & 3.2 & 4.9 & 1.1 \\ 8.4 & 2.6 & 3.6 & 8.8 \end{bmatrix}\begin{bmatrix} 2 & 1 & 0 \\ 3 & 0 & -9 \\ -7 & 5 & 5 \\ 3 & 9 & 8 \end{bmatrix}$

Using the matrix $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, *where possible, perform the indicated operation or answer the question in each of the Exercises 68–74.*

68. Find the matrix H such that $HA = \begin{bmatrix} 5\cdot 1 & 5\cdot 2 \\ 7\cdot 3 & 7\cdot 4 \end{bmatrix}$.

69. Find the matrix J such that $AJ = \begin{bmatrix} 1\cdot 4 & 2\cdot 9 \\ 3\cdot 4 & 4\cdot 9 \end{bmatrix}$.

70. $[6 \quad 1]A$ **71.** $A\begin{bmatrix} 6 \\ 1 \end{bmatrix}$ **72.** $A\begin{bmatrix} x \\ y \end{bmatrix}$ **73.** $\begin{bmatrix} x \\ y \end{bmatrix}A$ **74.** $A[6 \quad 1]$

75. Write the following system of equations as $AX = B$ using a matrix and vectors:

$$4x_1 + 5x_2 = -3$$
$$7x_1 + 9x_2 = 4$$

Projects

To complete the following projects, use a computational tool.

1. The Network Dynamics and Science Simulation Laboratory (NDSSL) at Virginia Technical University generated from real data a synthetic dataset for the activities of the population of Portland, Ore. Various NDSSL datasets are available at http://ndssl.vbi.vt.edu/opendata/download.php (NDSSL 2009a, b, c, d). Scientists use such data "for simulating the spread of epidemics at the level of individuals in a large urban region, taking into account realistic contact patterns and disease transmission characteristics" (VBI 2008). Files in the *Details* column describe the datasets; files in the *Samples* column display small example files; and files in the *Download* column are compressed large datasets. Omitting the header line in the file, cut and paste a Data Set Release 1 or 2 sample activities file into a text file. For this dataset develop code to accomplish the following tasks, which can be used for epidemiological studies:
   - **a.** Form the vector, *personIdLst*, of person IDs.
   - **b.** Form the vector, *locationIDLst*, of location IDs.
   - **c.** Generate connection matrix, *connMat*, with people indices representing row labels and location indices representing column labels. The *ij* element of *connMat* is 1 if the *i*th person visits the *j*th location; otherwise, the element is 0.
   - **d.** Write a function to return the number of locations a person visits.
   - **e.** Write a function to return the number of people that visit a location.
   - **f.** Generate a people-to-people connection matrix, *connPeopleMat*, with people indices representing row labels and column labels. The *ij* element of *connPeopleMat* is 1 if the *i*th and the *j*th people visit the same location in a day but not necessarily at the same time; otherwise, the element is 0.
   - **g.** Using *connPeopleMat* from Part f, write a function to return the degree of a person ID, that is, to return the number of people that go to locations visited by the individual.
   - **h.** Calculate the square of the matrix *connPeopleMat* from Part f. Develop a function that returns true if two people, A and B, have direct or indirect contact, that is if A and B were at the same location in a day or if there is a person C such that A and C were at the same location and C and B were at the same location in a day. As before, ignore times people visit locations. Explain why the square of *connPeopleMat* and the sum of *connPeopleMat* and its square are useful for this task.
2. Using the NDSSL site listed in Project 1, download and uncompress a Data Set Release 1 or 2 activities file. Generate 1000 random unique person IDs. From the dataset, create a data file with the activities lines for only these individuals. Repeat Project 1 with this new dataset.
3. Using the NDSSL site listed in Project 1, download and uncompress a Data Set Release 1 or 2 activities file. Repeat Project 1 using high-performance computing.

*Projects 4–7 use data from simulations with Cancer Chaste. "Chaste (***C***ancer,* ***H****eart* ***a****nd* ***S****oft* ***T****issue* ***E****nvironment) is a general purpose simulation package aimed at multi-scale, computationally demanding problems arising in biology and physiol-*

ogy. Current functionality includes tissue and cell level electrophysiology, discrete tissue modelling, and soft tissue modelling. The package is being developed by a team mainly based in the Computational Biology Group at Oxford University Computing Laboratory, and development draws on expertise from software engineering, high-performance computing, mathematical modeling and scientific computing. While Chaste is a generic extensible library, software development to date has focused on two distinct areas: continuum modelling of cardiac electrophysiology (Cardiac Chaste); and discrete modeling of cell populations (Systems Biology Chaste), with specific application to tissue homeostasis and carcinogenesis (Cancer Chaste)" (Chaste 2012). The initial focus of Cancer Chaste was on colorectal cancer, which it is believed originates in tiny ***crypts of Lieberkühn*** *that descend from the colon's epithelium into the underlying connective tissue (Cancer Chaste 2012).*

At Oxford, using Cancer Chaste, Ornella Cominetti and Angela Shiflet, in consultation with George Shiflet, developed simulations to see the impact of ***differential cell adhesion****, or variations in the level of adhesion between cells of various types, in the crypt. The categories of cells are* ***stem*** *(generation 0);* ***transit*** *categories* ***TA1*** *(generation 1),* ***TA2*** *(generation 2),* ***TA3*** *(generation 3), and* ***TA4*** *(generation 4); and* ***differentiated*** *(generation 5). Stem cells are anchored at the bottom of the crypt. Except for differentiated cells, cells of all other categories can divide. Using Cancer Chaste, the researchers' work attempts to reproduce the work of Wong et al. (2010) using a cellular Potts model. Files* cell_trajectory_file.txt, cell_types_file.txt, and cell_vel_file.txt *generated by some of the Oxford simulations are available for download from the text's website (Cominetti et al. 2010).*

4. (See the italicized description immediately before Project 4.) The file *cell_trajectory_file.txt*, which is available for download from the text's website, has simulated data about the location of a cell in the crypt each simulation hour until the cell leaves the crypt. Each line of the file contains the simulation time, generation, and x- and y-coordinates of the cell's location. Plot the trajectory of the cell using a different color for each generation. Have a legend indicating the generation. See Figure 4b of Wong et al. (2010) for a similar figure. Discuss the results.
5. (See the italicized description immediately before Project 4.) The file *cell_types_file.txt*, which is available for download from the text's website, has simulated data for 20 runs (experiments) of the simulation about the total number of each cell type every half hour for times 70 to 170 simulated hours. Each line of the file contains the simulation time and the number of cells in each category (stem, TA1, TA2, TA3, TA4, differentiated). Generate a stacked bar chart of the average number of cells in each category by time. See Figure 7 of Wong et al. (2010) for a similar figure. Are there any anomalies in the figure? Discuss the results.
6. (See the italicized description immediately before Project 4.) The file *cell_vel_file.txt*, which is available for download from the text's website, has simulated data about the velocities of cells in the crypt one simulation hour before the end of the simulation and at the end of the simulation for 20 runs (experiments) of the simulation. Each line of the file contains the simulation time, cell number, generation number, and x- and y-coordinates of the cell's location. Averaging over the 20 datasets, generate a plot of the mean migration velocities (change in y-coordinate over 1 h) of cells at different heights (y-coordinates) in the crypt. See Figure 5b, graph with triangles, of Wong et al. (2010) for a similar figure. Discuss the results.
7. (See the italicized description immediately before Project 4.) Scientists have found that cultured epithelial cells move collectively in sheets. For a simulation of cells in the crypt, we can use **spatial correlation of velocity**, $C(r)$, as a metric of the amount of coordinated movement of the cells. With r being the distance between two cell centroids, or centers of mass for the cells, for all pairs of cells at distance r from each other, we add the cosines of the angles between their velocity vectors and divide by the number of such pairs. For cells i and j with average velocities over 1 h (change in position from one hour to the next), $\mathbf{v}_i$ and $\mathbf{v}_j$, the cosine of the angle between $\mathbf{v}_i$ and $\mathbf{v}_j$ is $\dfrac{(\mathbf{v}_i \cdot \mathbf{v}_j)}{|\mathbf{v}_i||\mathbf{v}_j|}$, where $|\mathbf{v}_i|$ is the length of vector $\mathbf{v}_i$. Because the cos(0) = 1, its maximum, the fraction is largest when the angle is zero and the two velocity vectors point in the same direction, or the two cells are headed in the same direction. Thus, a large value for the spatial correlation of velocity, $C(r) = \dfrac{1}{N_r} \sum_{i,j}^{r=|r_i - r_j|} \dfrac{(\mathbf{v}_i \cdot \mathbf{v}_j)}{|\mathbf{v}_i||\mathbf{v}_j|}$, where N_r is the number of cell pairs with distance r, indicates a high correlation of velocities of pairs of cells at distance r from each other (Haga et al. 2005).

   The file *cell_vel_file.txt*, which is available for download from the text's website, has simulated data about the velocities of cells in the crypt one simulation hour before the end of the simulation and at the end of the simulation for 20 runs (experiments) of the simulation. Each line of the file contains the simulation time, cell number, generation number, and x- and y-coordinates of the cell's location. Using only data for differentiated cells, produce a plot of the mean $C(r)$ values along with **standard error bars**, or symmetric error bars that are two standard deviation units in length, and use intervals of length 1/6 for r. We consider only differentiated cells for two reasons. Because differentiated cells do not divide, cell division does not affect their velocities as much as it does cells of other types. Moreover, differentiated cells compose the largest category of cells. See Figure 6b, graph with circles, of Wong et al. (2010) for a similar figure.

Answers to Quick Review Questions

1. 6.00
2. a. $\mathbf{c}$ = (35, 16, 240, 351), $\mathbf{m}$ = (18, 10, 103, 153)
   b. $\mathbf{t}$ = (53, 26, 343, 504)
   c. data totals by category
3. a. 1.1$\mathbf{c}$
   b. (39, 18, 264, 386)
4. a. (0.23, 0.679)
   b. (808988, 715774)

c. $\mathbf{p} \cdot \mathbf{s}$
d. $672078 = 0.23 \cdot 808988 + 0.679 \cdot 715774$

5. a. 15.00
b. s_{42}

6. a. $\begin{bmatrix} -3 & 6 & 0 \\ 3 & 12 & 9 \end{bmatrix}$

b. $\begin{bmatrix} 0 & 5 & 9 \\ 1 & 9 & 9 \end{bmatrix}$

c. cannot be done because C has size 2×2, not 2×3, the size of $2A + B$

7. a. $A\mathbf{v} = \begin{bmatrix} 1 & 3 & 9 \\ 0 & 5 & 6 \end{bmatrix} \begin{bmatrix} 5 \\ 0 \\ -1 \end{bmatrix} = \begin{bmatrix} -4 \\ -6 \end{bmatrix}$ because

$(1)(5) + (3)(0) + (9)(-1) = -4$
$(0)(5) + (5)(0) + (6)(-1) = -6$

b. 8×1
c. 5×1

8. a. $A \cdot B = \begin{bmatrix} 8 & 5 & 3 & -4 \\ -5 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & -6 \\ 7 & 1 \\ 4 & 3 \\ -9 & -2 \end{bmatrix} = \begin{bmatrix} 99 & -26 \\ -12 & 29 \end{bmatrix}$ because

$(8)(2) + (5)(7) + (3)(4) + (-4)(-9) = 99$
$(8)(-6) + (5)(1) + (3)(3) + (-4)(-2) = -26$
$(-5)(2) + (1)(7) + (0)(4) + (1)(-9) = -12$
$(-5)(-6) + (1)(1) + (0)(3) + (1)(-2) = 29$

b. $B \cdot A = \begin{bmatrix} 2 & -6 \\ 7 & 1 \\ 4 & 3 \\ -9 & -2 \end{bmatrix} \begin{bmatrix} 8 & 5 & 3 & -4 \\ -5 & 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 46 & 4 & 6 & -14 \\ 51 & 36 & 21 & -27 \\ 17 & 23 & 12 & -13 \\ -62 & -47 & -27 & 34 \end{bmatrix}$

c. C
d. C

9. a. $6.57x + 5.27y = 3.04$
b. $(6.57, 5.27) \cdot (x, y) = 3.04$

10. $\begin{bmatrix} 2 & -1 \\ 6 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 7 \\ 5 \end{bmatrix}$

References

Bottini, Nunzio, Gian Franco Meloni, Andrea Finocchi, Giuseppina Ruggiu, Ada Amante, Tullio Meloni, and Egidio Bottini. 2001. "Maternal-Fetal Interaction in the ABO System: A Comparative Analysis of Healthy Mothers and Couples with

Recurrent Spontaneous Abortion Suggests a Protective Effect of B Incompatibility." *Human Biology*, 73(2): 167–174.

Cancer Chaste. 2012. "Cell-Based Chaste: A Multiscale Computational Framework for Modeling Cell Populations." http://www.cs.ox.ac.uk/chaste/cell_based_index.html (accessed December 30, 2012)

Chaste (Cancer, Heart and Soft Tissue Environment). 2012. http://www.cs.ox.ac.uk/chaste/about.html (accessed December 30, 2012)

Cominetti, Ornella, Angela Shiflet, and George Shiflet. 2010. Files *cell_trajectory_file*, *cell_types_file*, and *cell_vel_file* available on this website.

FAO (Food and Agriculture Organization of the United Nations). 2012. "*Callinectes sapidus* Species Fact Sheets," Fisheries and Aquaculture Department. http://www.fao.org/fishery/species/2632/en (accessed December 30, 2012)

Green Sea Turtle. Demographics of the Hawaiian Green Sea Turtle: Modeling Population Dynamics Using a Linear Deterministic Matrix Model: The Leslie Matrix. http://isolatium.uhh.hawaii.edu/linear/ch6/green.htm (accessed December 30, 2012)

Haga, H., Irahara, C., Kobayashi, R., Nakagaki, T. and Kawabata, K. 2005. "Collective Movement of Epithelial Cells on a Collagen Gel Substrate," *Biophys. J.*, 88: 2250– 2256.

Johnson, Ted R., and Christine L. Case. 2009. *Laboratory Experiments in Microbiology*. San Francisco, CA: Benjamin Cummings.

Lockhart, Peter J., Michael A. Steel, Michael D. Hendy, and David Penny. 1994. "Recovering Evolutionary Trees under a More Realistic Model of Sequence Evolution." *Mol. Biol. Evol.*, 11(4): 605–6 12.

LANL (Los Alamos National Laboratory). " EpiSimS: Epidemic Simulation System" http://www.lanl.gov/programs/nisac/episims.shtml (accessed January 26, 2013)

NDSSL (Network Dynamics and Simulation Science Laboratory, Virginia Polytechnic Institute and State University). 2009a. "NDSSL Proto-Entities" http://ndssl.vbi.vt.edu/opendata/ (accessed December 30, 2012)

———. 2009b. Synthetic Data Products for Societal Infrastructures and Proto-Populations: Data Set 1.0. ndssl.vbi.vt.edu/Publications/ndssl-tr-06-006.pdf (accessed December 30, 2012)

———. 2009c. Synthetic Data Products for Societal Infrastructures and Proto-Populations: Data Set 2.0. ndssl.vbi.vt.edu/Publications/ndssl-tr-07-003.pdf (accessed December 30, 2012)

———. 2009. Synthetic Data Products for Societal Infrastructures and Proto-Populations: Data Set 3.0. ndssl.vbi.vt.edu/Publications/ndssl-tr-07-010.pdf (accessed December 30, 2012)

Taylor, Caz, and Erin Grey. 2010. "Population Dynamics of Gulf Blue Crabs" Tulane University. http://leag.tulane.edu/PDFs/Grey-LEAG-4.28.10.pdf (accessed December 30, 2012)

VBI (Virginia Bioinformatics Institute at Virginia Tech). 2008. "EpiSims." http://ndssl.vbi.vt.edu/episims.php (accessed December 30, 2012)

Wong, Shek Yoon, K.-H. Chiam, Chwee Teck Lim, and Paul Matsudaira. 2010. "Computational Model of Cell Positioning: Directed and Collective Migration in the Intestinal Crypt Epithelium." *J. R. Soc. Interface*, 7: S351–S363. First published online March 31, 2010; *doi: 10.1098/rsif.2010.0018.focus)

Zinski, Steven C. 2006. "Blue Crab Life Cycle." http://www.bluecrab.info/lifecycle.html (accessed December 30, 2012)