Geographic Information Analysis, Second Edition. David O'Sullivan and David J. Unwin.

Appendix A

Notation, Matrices, and Matrix Mathematics

A.1. INTRODUCTION

In this appendix, we outline the notation that we use in this book and then some of the mathematics of *matrices* and closely related *vectors*. This material is worth mastering, because notation is important in ensuring consistency in many of the materials we present and, as will be discovered, matrices are vital to pursuing many topics in spatial analysis (and many other disciplines). In some cases, they provide a compact way of expressing questions and problems, but they also provide a useful generic way of representing the extremely important concept of adjacency in spatial systems.

We have two aims: (1) that you acquire familiarity with the notation and terminology of matrices and (2) that you become used to the way simple arithmetic operations are performed with them.

Before starting, we must introduce the basics of mathematical notation.

A.2. SOME PRELIMINARY NOTES ON NOTATION

In using mathematical notation in an introductory book, such as this, one has to steer a course between two extremes. Too rigorous adherence to a particular notation scheme can mystify the reader just as easily as a too casual approach can confuse. A further complication is that there are standard uses in the literature that need to be followed if possible. In developing this book, we have tried to be as consistent as possible and to follow some relatively straightforward basics. We hope that readers unfamiliar with the field will find this description of these basics useful.

A single instance of some variable or quantity is usually denoted by a *lowercase italicized* letter. Sometimes this is the initial letter of the quantity we're talking about—say, h for height or d for distance. More often, in

Table A.1 Commonly Used Symbols and Their Meaning in This Book

Symbol	*Meaning*
x	The Easting geographic coordinate or a general data value
y	The Northing geographic coordinate or a general data value
z, a, b	The numerical value of some measurement recorded at the geographic coordinates (x, y)
n, m	The number of observations in a data set
k	Either an arbitrary constant or the number of entities in a spatial neighborhood
d	Distance
w	The strength or "weight" of interaction between locations
$\mathbf{s}$	An arbitrary (x, y) location

introducing a statistical measure, we don't really care what the numbers represent (because they could be *anything*), so we employ one of the commonly used mathematical letters, say, x or y. Commonly used letters are x, y, z, n, m, and k. In the main text, these occur frequently, and generally have the meanings described in Table A.1. In addition to these six, you will note that d, w, and $\mathbf{s}$ also occur frequently in spatial analysis. The reason for use of an **upright bold** symbol for $\mathbf{s}$ is made clear later, where vectors and matrices are discussed.

A familiar aspect of mathematical notation is that letters from the Greek alphabet are used alongside the Roman alphabet letters that you are used to. You may already be familiar with mu (μ) for a population mean, sigma (σ) for population standard deviation, chi (χ) for a particular statistical distribution, and pi (π) for . . . well, just for "pi." In general, we try to avoid using any Greek symbols other than these, although lambda (λ) is commonly used for the intensity of a spatial process. In statistical logic, it is important to keep in mind the distinction between some parameter of an entire defined population and any estimate of that same parameter arrived at by analysis of a sample from that population. Usually, which is which will be evident from the context, but we also use Greek letters (as above) to indicate population parameters. Estimates of parameters are indicated by a "hat" symbol above the letter used for the parameter. Thus, the unknown intensity of a spatial process is indicated as λ and an estimate of it as $\hat{\lambda}$.

Symbols are introduced so that we can use mathematical notation to talk about related values or to indicate mathematical operations that we want to perform on sets of values. So, if h (or z) represents our height value, then h^2 (or z^2) indicates "height value squared." The symbols are a concise way of saying the same thing, and that's very important when we describe more complex operations on data sets.

Two symbols that you will see often are i and j. However, i and j normally appear in a particular way. To describe complex operations on sets of values, we need another notational device: the *subscript*. Subscripts are small italic letters or numbers below and to the right of normal mathematical symbols: the i in z_i is a subscript. A subscript is used to signify that there may be more than one item of the type denoted by the symbol, so z_i stands for a series or set of z values: z_1, z_2, z_3, and so on. This has various uses:

- A set of values is written between *braces*, so that $\{z_1, z_2, \ldots, z_{n-1}, z_n\}$ tells us that there are n elements in this set of z values. If required, the set as a whole may be denoted by a capital letter: Z. A typical value from the set Z is denoted z_i, and we can abbreviate the previous partial listing to simply $Z = \{z_i\}$, where it is understood that the set has n elements.
- In spatial analysis, it is common for the subscripts to refer to locations at which observations have been made and for the same subscripts to be used across a number of different data sets. Thus, h_7 and t_7 refer to the values of two different observations—say, height and temperature—at the same location ("location 7").
- Subscripts may also be used to distinguish different calculations of (say) the same statistic on different populations or samples. Thus, μ_A and μ_B denote the means of two different data sets, A and B.

The symbols i and j usually appear as subscripts in one of these ways. A particularly common usage is to denote *summation* operations, indicated by the Σ symbol (another Greek letter, this time capital sigma). This is where subscripts come into their own, because we can specify a range of values that are summed to produce a result. Thus, the sum

$$a_1 + a_2 + a_3 + a_4 + a_5 + a_6 \tag{A.1}$$

is denoted

$$\sum_{i=1}^{i=6} a_i \tag{A.2}$$

indicating that summation of a set of a values should be carried out on all the elements from a_1 to a_6. For a set of n "a" values, this becomes

$$\sum_{i=1}^{i=n} a_i \tag{A.3}$$

which is usually abbreviated to either

$$\sum_{i=1}^{n} a_i \tag{A.4}$$

or

$$\sum_i a_i \tag{A.5}$$

where the number of values in the set of a's is understood to be n. If, instead of the simple sum, we wanted the sum of the squares of the a values, then we would have

$$\sum_{i=1}^{n} a_i^2 \tag{A.6}$$

instead. Or perhaps we have two data sets, A and B, and we want the sum of the products of the a and b values at each location. This would be denoted

$$\sum_{i=1}^{n} a_i b_i \tag{A.7}$$

In spatial analysis, more complex operations might be carried out *between* two sets of values, and then we may need two summation operators. For example,

$$c = k \sum_{i=1}^{n} \sum_{j=1}^{n} \left(z_i - z_j\right)^2 \tag{A.8}$$

indicates that c is to be calculated in two stages. First, we take each z value in turn (the outer i subscript) and sum the square of its value minus every z value in turn (the j subscript). You can figure this out by imagining first setting i to 1 and calculating the inner sum, which would be $\sum_j \left(z_1 - z_j\right)^2$. We then set i to 2, and do the summation $\sum_j \left(z_2 - z_j\right)^2$, and so on all the way to $\sum_j \left(z_n - z_j\right)^2$. The final "double summation" is the sum of all of these individual sums, and c is equal to this sum multiplied by k. This will seem complex at first, but you will get used to it.

A.3. MATRIX BASICS AND NOTATION

A *matrix* is a rectangular array of numbers arranged in rows and columns; for example,

$$\begin{bmatrix} 2 & 4 & 7 & -2 \\ 0 & 1 & -3 & 3 \\ 5 & -1 & 7 & 1 \end{bmatrix} \tag{A.9}$$

As shown above, a matrix is usually written enclosed in square brackets. This matrix has three *rows* and four *columns*. The size of a matrix is described in terms of the number of rows by the number of columns, so the example above is a "3 by 4" matrix. A *square matrix* has equal numbers of rows and columns. For example,

$$\begin{bmatrix} 3 & 1 & 2 \\ 1 & -3 & 4 \\ 6 & -1 & 0 \end{bmatrix} \tag{A.10}$$

is a 3 by 3 square matrix. When we wish to talk about matrices in general terms, it is usual to represent them using uppercase **ROMAN BOLD** characters:

$$\mathbf{A} = \begin{bmatrix} 2 & 4 & 7 & -2 \\ 0 & 1 & -3 & 3 \\ 5 & -1 & 7 & 1 \end{bmatrix} \tag{A.11}$$

Individual elements in a matrix are generally referred to using *lowercase italic* characters, with their row and column numbers written as subscripts. The element in the top left corner of the above matrix is $a_{11} = 2$, and element a_{24} is the entry in row 2, column 4, and is equal to 3. In general, the subscripts i and j are used to represent rows and columns, and a general matrix has n rows and p columns, so we have

$$\mathbf{B} = \begin{bmatrix} b_{11} & \cdots & b_{1j} & \cdots & b_{1p} \\ \vdots & \ddots & & & \vdots \\ b_{i1} & & b_{ij} & & b_{ip} \\ \vdots & & & \ddots & \vdots \\ b_{n1} & \cdots & b_{nj} & \cdots & b_{np} \end{bmatrix} \tag{A.12}$$

Vectors and Matrices

A *vector* is a quantity that has size and direction. It is convenient to represent a vector graphically by an arrow of length equal to its size, pointing in the vector's direction. Typical vectors are shown in Figure A.1. In geography, vectors might be used to represent winds or current flows. In a more abstract application, they might represent migration flows. In terms of a typology of spatial data (see Chapter 1), we can add vectors to our list of types of quantity so that we have nominal, ordinal, interval, ratio, and vector types. In

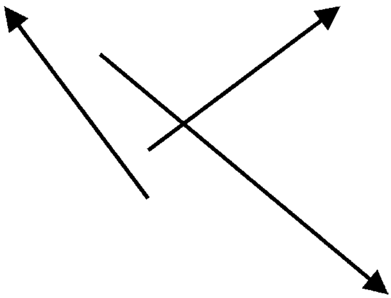

Figure A.1 Typical vectors.

particular, we can imagine a *vector field* representing, for example, the wind patterns across a region, as shown in Figure A.2.

How do we represent a vector mathematically, and what do vectors have to do with matrices? In two-dimensional space (as in the diagrams), we can use two numbers, representing the vector *components* in two perpendicular directions. This should be familiar from geographic grid coordinate systems

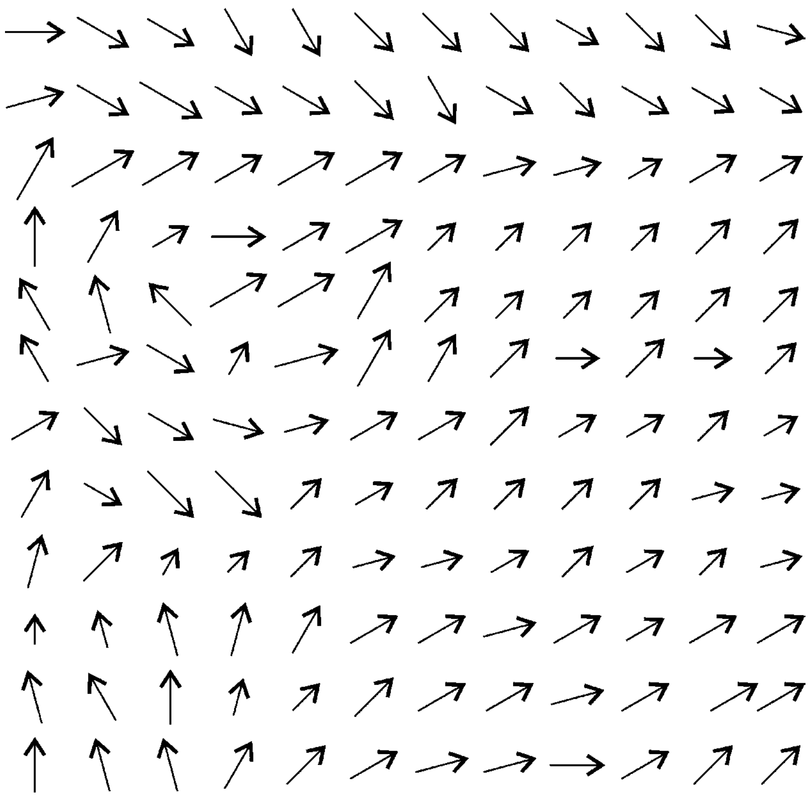

Figure A.2 A vector field.

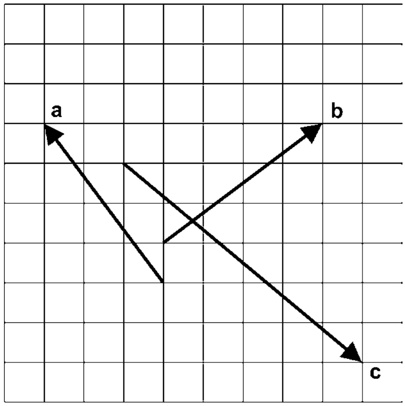

Figure A.3 Vectors in a coordinate space.

and is shown in Figure A.3. The three vectors shown have components **a** = (–3, 4), **b** = (4, 3), and **c** = (6, –5) in the east–west and north–south directions, respectively, relative to the coordinate system shown on the grid.

An alternative way to represent vectors is as *column matrices*, that is, as 2 by 1 matrices:

$$\mathbf{a} = \begin{bmatrix} -3 \\ 4 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 4 \\ 3 \end{bmatrix}, \text{ and } \mathbf{c} = \begin{bmatrix} 6 \\ -5 \end{bmatrix} \tag{A.13}$$

Thus, a vector is a particular type of matrix with only one column. As here, vectors are usually denoted by a lowercase **roman bold** symbol. In the same way, point locations relative to an origin can be represented as vectors. This is why we sometimes use the notation in the main text where a point is represented as

$$\mathbf{s} = \begin{bmatrix} x \\ y \end{bmatrix} \tag{A.14}$$

Note also that we can represent a location in three dimensions in exactly the same way. Instead of a 2 by 1 column matrix, we use a 3 by 1 column matrix. More abstractly, in n-dimensional space, a vector will have n rows, so that it is an n by 1 matrix.

A.4. SIMPLE MATRIX MATHEMATICS

Now let us review the mathematical rules by which matrices are manipulated.

Addition and Subtraction

Matrix addition and subtraction are straightforward. Corresponding elements in the matrices in the operation are simply added (or subtracted) to produce the result. Thus, if

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \tag{A.15}$$

and

$$\mathbf{B} = \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix} \tag{A.16}$$

then

$$\begin{aligned} \mathbf{A} + \mathbf{B} &= \begin{bmatrix} 1+5 & 2+6 \\ 3+7 & 4+8 \end{bmatrix} \\ &= \begin{bmatrix} 6 & 8 \\ 10 & 12 \end{bmatrix} \end{aligned} \tag{A.17}$$

Subtraction is defined similarly. It follows from this that $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$. It also follows that $\mathbf{A}$ and $\mathbf{B}$ must each have the same number of rows and columns for addition (or subtraction) to be possible.

For vectors, subtraction has a specific useful interpretation. If $\mathbf{s}_1$ and $\mathbf{s}_2$ are two locations, then the vector from $\mathbf{s}_1$ to $\mathbf{s}_2$ is given by $\mathbf{s}_2 - \mathbf{s}_1$. This is illustrated in Figure A.4, where the vector $\mathbf{x}$ from $\mathbf{s}_1$ to $\mathbf{s}_2$ is given by

$$\begin{aligned} \mathbf{x} &= \mathbf{s}_2 - \mathbf{s}_1 \\ &= \begin{bmatrix} 5 \\ 7 \end{bmatrix} - \begin{bmatrix} 8 \\ 3 \end{bmatrix} \\ &= \begin{bmatrix} -3 \\ 4 \end{bmatrix} \end{aligned} \tag{A.18}$$

Multiplication

Multiplication of matrices and vectors is more involved. The easiest way to think of the multiplication operation is that we "multiply rows into columns." Mathematically, we can define multiplication as follows: If

$$\mathbf{C} = \mathbf{AB} \tag{A.19}$$

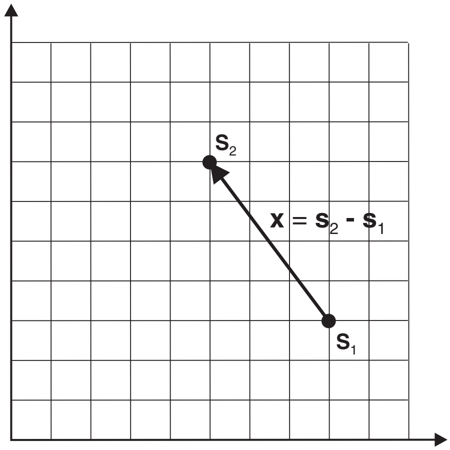

Figure A.4 Vector subtraction gives the vector between two point locations.

then the element in row i, column j of $\mathbf{C}$ is given by

$$c_{ij} = \sum_k a_{ik} b_{kj} \tag{A.20}$$

Thus, element in the ith row and jth column of the product of $\mathbf{A}$ and $\mathbf{B}$ is the sum of the products of corresponding elements from the ith row of $\mathbf{A}$ and the jth column of $\mathbf{B}$. Working through an example will make this clearer. If

$$\mathbf{A} = \begin{bmatrix} 1 & -2 & 3 \\ -4 & 5 & -6 \end{bmatrix} \tag{A.21}$$

and

$$\mathbf{B} = \begin{bmatrix} 6 & -5 \\ 4 & -3 \\ 2 & -1 \end{bmatrix} \tag{A.22}$$

then, for the element in *row 1, column 1* of the product $\mathbf{C}$, we have the sum of products of corresponding elements in *row 1* of $\mathbf{A}$ and *column 1* of $\mathbf{B}$, that is,

$$\begin{aligned} c_{11} &= a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31} \\ &= (1 \times 6) + (-2 \times 4) + (3 \times 2) \\ &= 6 - 8 + 6 \\ &= 4 \end{aligned} \tag{A.23}$$

Similarly, we have

$$\begin{aligned}
c_{12} &= (1 \times -5) + (-2 \times -3) + (3 \times -1) \\
&= -5 + 6 + (-3) \\
&= -2 \\
c_{21} &= (-4 \times 6) + (5 \times 4) + (-6 \times 2) \\
&= -24 + 20 + (-12) \\
&= -16 \\
c_{22} &= (-4 \times -5) + (5 \times -3) + (-6 \times -1) \\
&= 20 + (-15) + 6 \\
&= 11
\end{aligned} \tag{A.24}$$

This gives us the final product matrix

$$\mathbf{C} = \begin{bmatrix} 4 & -2 \\ -16 & 11 \end{bmatrix} \tag{A.25}$$

Figure A.5 shows how multiplication works schematically. Corresponding elements from a row of the first matrix and a column of the second are multiplied together and summed to produce a single element of the product

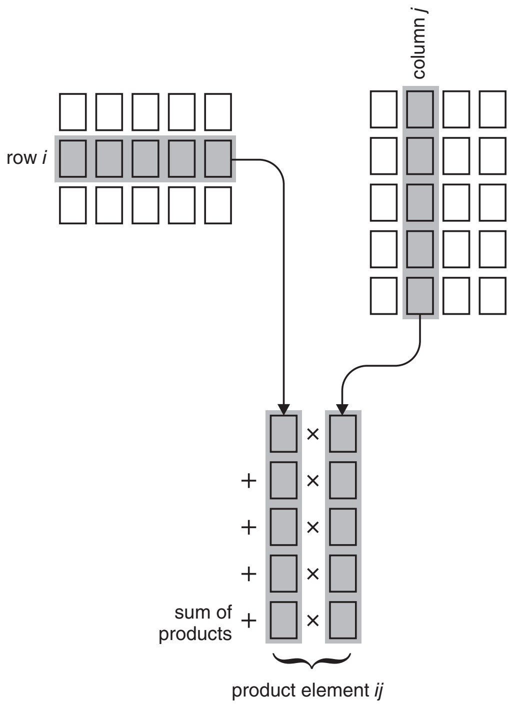

Figure A.5 Matrix multiplication.

matrix. This element's position in the product matrix corresponds to the row number from the first matrix and the column number from the second. Because of the way matrix multiplication works, it is necessary that the first matrix has the same number of columns as the second has rows. If this is not the case, then the matrices cannot be multiplied. If you write the matrices you want to multiply as ${}_n\mathbf{A}_p$ (n rows, p columns) and ${}_x\mathbf{B}_y$ (x rows, y columns), then you can determine whether they multiply by checking that the subscripts between the two matrices are equal:

$$ {}_n\mathbf{A}_p \ {}_x\mathbf{B}_y \tag{A.26}$$

If $p = x$, then this multiplication is possible and the product $\mathbf{AB}$ exists. Furthermore, the product matrix has dimensions given by the "outer" subscripts, n and y, so that the product will be an n by y matrix. On the other hand, for

$$ {}_x\mathbf{B}_y \ {}_n\mathbf{A}_p \tag{A.27}$$

if $y \neq n$, then $\mathbf{BA}$ *does not exist* and multiplication is not possible. Note that this means that, in general, for matrices

$$\mathbf{AB} \neq \mathbf{BA} \tag{A.28}$$

and multiplication is not *commutative*: it is *order dependent*. This is important when matrices are used to transform between coordinate spaces (see Section A.6).

In the example above,

$$\mathbf{C} = \mathbf{AB} = \begin{bmatrix} 4 & -2 \\ -16 & 11 \end{bmatrix} \tag{A.29}$$

but

$$\mathbf{D} = \mathbf{BA} = \begin{bmatrix} 26 & -37 & 48 \\ 16 & -23 & 30 \\ 6 & -9 & 12 \end{bmatrix} \tag{A.30}$$

Here the product $\mathbf{D}$ is not even the same size as $\mathbf{C}$, and this is not unusual. However, it is useful to know that $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$. The rule is that, provided the written order of multiplications is preserved, multiplications may be carried out in any sequence.

Matrix Transposition

The *transpose* of a matrix is obtained by swapping rows for columns. This operation is indicated by a superscript $^{\mathrm{T}}$, so that the transpose of $\mathbf{A}$ is written $\mathbf{A}^{\mathrm{T}}$. Hence,

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}^{\mathrm{T}} = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} \tag{A.31}$$

Note that this definition, combined with the row-column requirement for multiplication, means that $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ and $\mathbf{A}\mathbf{A}^{\mathrm{T}}$ always exist. The product $\mathbf{a}^{\mathrm{T}}\mathbf{a}$ is of particular interest when $\mathbf{a}$ is a vector, because it is equal to the sum of the squares of the components of the matrix. This means that the *length* of a vector $\mathbf{a}$ is given by $\sqrt{(\mathbf{a}^{\mathrm{T}}\mathbf{a})}$, from Pythagoras's Theorem. See Section A.6 for more on this topic.

A.5. SOLVING SIMULTANEOUS EQUATIONS USING MATRICES

We now come to one of the major applications of matrices. Suppose we have a pair of equations in two unknowns, x and y, for example:

$$\begin{aligned} 3x + 4y &= 11 \\ 2x - 4y &= -6 \end{aligned} \tag{A.32}$$

The usual way to solve this is to add a multiple of one of the equations to the other, so that one of the unknown variables is eliminated, leaving an equation in one unknown, which we can solve. The second unknown is then found by substituting the first known value back into one of the original equations. In this example, if we add the second equation to the first, we get

$$(3 + 2)x + (4 - 4)y = 11 + (-6) \tag{A.33}$$

which gives us

$$5x = 5 \tag{A.34}$$

so that $x = 1$. Substituting this into (say) the first equation, we get

$$3(1) + 4y = 11 \tag{A.35}$$

so that

$$4y = 11 - 3 \tag{A.36}$$

which we easily solve to get $y = 2$. This is simple enough. But what if we have 3 unknowns, or 4, or 100, or 10,000? This is where matrix algebra comes into its own. To understand how, we must introduce two more matrix concepts: the identity matrix and the inverse matrix.

The Identity Matrix and the Inverse Matrix

The identity matrix, written **I**, is defined such that

$$\mathbf{IA} = \mathbf{AI} = \mathbf{A} \tag{A.37}$$

Think of the identity matrix as the matrix equivalent of the number 1, since $1 \times z = z \times 1 = z$, where z is any number. It turns out that the identity matrix is always a square matrix with the required number of rows and columns for the multiplication to go through. Elements in **I** are all equal to 1 on the *main diagonal* from top left to bottom right. All other elements are equal to 0. The 2 by 2 identity matrix is

$$\mathbf{I} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{A.38}$$

The 5 by 5 identity matrix is

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{A.39}$$

and so on.

We now define the *inverse* $\mathbf{A}^{-1}$ of matrix **A**, such that

$$\mathbf{AA}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I} \tag{A.40}$$

Finding the inverse of a matrix is tricky and is not always possible. For 2 by 2 matrices it is simple:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \tag{A.41}$$

For example if

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \tag{A.42}$$

then we have

$$\begin{aligned} \mathbf{A}^{-1} &= \frac{1}{(1\times 4)-(2\times 3)}\begin{bmatrix} 4 & -2 \\ -3 & 1 \end{bmatrix} \\ &= -\frac{1}{2}\begin{bmatrix} 4 & -2 \\ -3 & 1 \end{bmatrix} \\ &= \begin{bmatrix} -2 & 1 \\ \frac{3}{2} & -\frac{1}{2} \end{bmatrix} \end{aligned} \tag{A.43}$$

We can check that this really is the inverse of **A** by calculating $\mathbf{AA}^{-1}$:

$$\begin{aligned} \mathbf{AA}^{-1} &= \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \times \begin{bmatrix} -2 & 1 \\ \frac{3}{2} & -\frac{1}{2} \end{bmatrix} \\ &= \begin{bmatrix} (1\times -2)+\left(2\times\frac{3}{2}\right) & (1\times 1)+\left(2\times -\frac{1}{2}\right) \\ (3\times -2)+\left(4\times\frac{3}{2}\right) & (3\times 1)+\left(4\times -\frac{1}{2}\right) \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \end{aligned} \tag{A.44}$$

We leave it to you to check that the product $\mathbf{A}^{-1}\mathbf{A}$ also equates to **I**.

Unfortunately, finding the inverse for bigger matrices rapidly becomes much more difficult as the matrix gets bigger. Fortunately, it isn't necessary for you to know how to perform matrix inversion. The important things to remember are its definition and its relation to the identity matrix. Almost invariably, computer routines using well-known and reliable algorithms will be employed to invert any large matrices you come across.

Some other points are also worth noting:

- The quantity $ad - bc$ is known as the matrix *determinant* and is usually denoted $|\mathbf{A}|$. If $|\mathbf{A}| = 0$, then the matrix **A** has no inverse. The determinant of a larger square matrix can be found recursively from the determinants of smaller matrices known as the *cofactors* of the matrix. You will find details in books on linear algebra (Strang, 1988, is recommended).

- It is also useful to know that

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1} \tag{A.45}$$

You can verify this from

$$\begin{aligned}\mathbf{B}^{-1}\mathbf{A}^{-1}\mathbf{AB} &= \mathbf{B}^{-1}\left(\mathbf{A}^{-1}\mathbf{A}\right)\mathbf{B} \\ &= \mathbf{B}^{-1}(\mathbf{I})\mathbf{B} \\ &= \mathbf{B}^{-1}\mathbf{B} \\ &= \mathbf{I}\end{aligned} \tag{A.46}$$

- Also useful is

$$\left(\mathbf{A}^{\mathrm{T}}\right)^{-1} = \left(\mathbf{A}^{-1}\right)^{\mathrm{T}} \tag{A.47}$$

Now, Back to the Simultaneous Equations

Now we know about inverting matrices, we can get back to the simultaneous equations:

$$\begin{aligned}3x + 4y &= 11 \\ 2x - 4y &= -6\end{aligned} \tag{A.48}$$

The key is to realize that these can be rewritten as the matrix equation:

$$\begin{bmatrix} 3 & 4 \\ 2 & -4 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 11 \\ -6 \end{bmatrix} \tag{A.49}$$

Now, to solve the original equations, if we can find the inverse of the first matrix on the left-hand side, we can premultiply both sides of the matrix equation by the inverse matrix to obtain a solution for x and y directly. The inverse of

$$\begin{bmatrix} 3 & 4 \\ 2 & -4 \end{bmatrix} \tag{A.50}$$

is

$$-\frac{1}{20}\begin{bmatrix} -4 & -4 \\ -2 & 3 \end{bmatrix} \tag{A.51}$$

Doing the premultiplication on both sides, we get

$$-\frac{1}{20}\begin{bmatrix} -4 & -4 \\ -2 & 3 \end{bmatrix}\begin{bmatrix} 3 & 4 \\ 2 & -4 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = -\frac{1}{20}\begin{bmatrix} -4 & -4 \\ -2 & 3 \end{bmatrix}\begin{bmatrix} 11 \\ -6 \end{bmatrix} \tag{A.52}$$

which gives us

$$\begin{aligned}
\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} &= -\frac{1}{20} \begin{bmatrix} (-4 \times 11) + (-4 \times -6) \\ (-2 \times 11) + (3 \times -6) \end{bmatrix} \\
\begin{bmatrix} x \\ y \end{bmatrix} &= -\frac{1}{20} \begin{bmatrix} -44 + 24 \\ -22 - 18 \end{bmatrix} \\
&= -\frac{1}{20} \begin{bmatrix} -20 \\ -40 \end{bmatrix} \\
\begin{bmatrix} x \\ y \end{bmatrix} &= \begin{bmatrix} 1 \\ 2 \end{bmatrix}
\end{aligned} \tag{A.53}$$

which is the same solution for x and y that we obtained before. This all probably seems a bit laborious for just two equations! The point is that this approach can be scaled up very easily to much larger sets of equations, and provided we can find the inverse of the matrix on the left-hand side, the equations can be solved. We can generalize this result. Any system of equations can be written

$$\mathbf{Ax} = \mathbf{b} \tag{A.54}$$

and the solution is given by premultiplying both sides by $\mathbf{A}^{-1}$ to get

$$\mathbf{A}^{-1}\mathbf{Ax} = \mathbf{A}^{-1}\mathbf{b} \tag{A.55}$$

Since $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$, we then have

$$\mathbf{Ix} = \mathbf{x} = \mathbf{A}^{-1}\mathbf{b} \tag{A.56}$$

This is an amazingly compressed statement of the problem of solving any number of equations. Remember that the matrix equation $\mathbf{Ax} = \mathbf{b}$ can represent a system of hundreds or even thousands of equations, not just two or three. Note also that if we calculate the determinant of $\mathbf{A}$ and find that it is zero, then we know that the equations cannot be solved, since $\mathbf{A}$ has no inverse. Furthermore, having solved this system once by finding $\mathbf{A}^{-1}$, we can quickly solve it for any values on the right-hand side of the equation.

Because of this general result, matrices have become central to modern mathematics, statistics, computer science, and engineering. In a smaller way, they are important in spatial analysis, as will become clear in the main text.

A.6. MATRICES, VECTORS, AND GEOMETRY

Another reason for the importance of matrices is their usefulness in representing coordinate geometry. We have already seen that a vector (in two or more dimensions) may be considered a *column vector* where each element represents the vector's length parallel to each of the axes of the coordinate space. We expand here on a point that we have already touched on relating to the calculation of the quantity $\mathbf{a}^{\mathrm{T}}\mathbf{a}$ for a vector. As we have already mentioned, this quantity is equal to the sum of the squares of the components of $\mathbf{a}$, so that the length of $\mathbf{a}$ is given by

$$|\mathbf{a}| = \sqrt{\mathbf{a}^{\mathrm{T}}\mathbf{a}} \tag{A.57}$$

This result applies regardless of the number of dimensions of $\mathbf{a}$.

We can use this result to determine the angle between any two vectors $\mathbf{a}$ and $\mathbf{b}$. In Figure A.6, the vector $\mathbf{a}$ forms an angle A with the positive x axis, and $\mathbf{b}$ forms angle B. The angle between the two vectors $(B - A)$ we label θ.

Using the well-known trigonometric equality

$$\cos(B - A) = \cos A\,\cos B + \sin A\,\sin B \tag{A.58}$$

we have

$$\begin{aligned}\cos\theta &= \cos A\,\cos B + \sin A\,\sin B\\ &= \left(\frac{x_a}{|\mathbf{a}|}\times\frac{x_b}{|\mathbf{b}|}\right) + \left(\frac{y_a}{|\mathbf{a}|}\times\frac{y_b}{|\mathbf{b}|}\right)\\ &= \frac{x_a x_b + y_a y_b}{|\mathbf{a}||\mathbf{b}|}\\ &= \frac{\mathbf{a}^{\mathrm{T}}\mathbf{b}}{\sqrt{\mathbf{a}^{\mathrm{T}}\mathbf{a}}\sqrt{\mathbf{b}^{\mathrm{T}}\mathbf{b}}}\end{aligned} \tag{A.59}$$

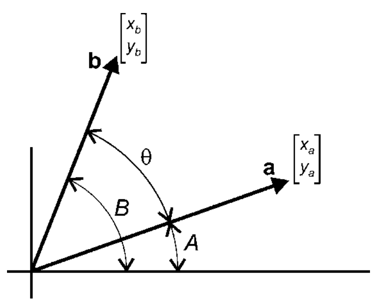

Figure A.6 Derivation of the expression for the angle between two vectors (see text).

The quantity $\mathbf{a}^{\mathrm{T}}\mathbf{b}$ is known as the *dot product* or *scalar product* of the two vectors and is simply the sum of products of corresponding vector components. One of the most important corollaries of this result is that two vectors whose dot product is equal to zero are perpendicular or *orthogonal*. This follows directly from the fact that cos 90° is equal to zero. Although we have derived this result in two dimensions, it again scales to any number of dimensions, even if we have trouble understanding what "perpendicular" means in nine dimensions! The result is also considered to apply to matrices, so that if $\mathbf{A}^{\mathrm{T}}\mathbf{B} = 0$, then we say that matrices **A** and **B** are orthogonal.

The Geometric Perspective on Matrix Multiplication

In this context, it is useful to introduce an alternative way of understanding the matrix multiplication operation. Consider the the 2×2 matrix, **A**, and the spatial location vector, **s**

$$\mathbf{A} = \begin{bmatrix} 0.6 & 0.8 \\ -0.8 & 0.6 \end{bmatrix}, \mathbf{s} = \begin{bmatrix} 3 \\ 4 \end{bmatrix} \tag{A.60}$$

The product, **As**, of these matrices is

$$\mathbf{As} = \begin{bmatrix} 5 \\ 0 \end{bmatrix} \tag{A.61}$$

We can look at a diagram of this operation in two-dimensional coordinate space, as shown on the left-hand side of Figure A.7. The vector **As** is a rotated version of the original vector **s**. If we perform the same multiplication on a series of vectors, collected together in the two-row matrix **S** so that each column of **S** is a vector,

$$\begin{aligned} \mathbf{AS} &= \begin{bmatrix} 0.6 & 0.8 \\ -0.8 & 0.6 \end{bmatrix} \begin{bmatrix} 1 & 3 & 0 & -1 & -2.5 \\ 1 & -2 & 5 & 4 & -4 \end{bmatrix} \\ &= \begin{bmatrix} 1.4 & 0.2 & 4 & 2.6 & -4.7 \\ -0.2 & -3.6 & 3 & 3.2 & -0.4 \end{bmatrix} \end{aligned} \tag{A.62}$$

then we can see that multiplication by the matrix **A** may be considered equivalent to a clockwise *rotation* of the vectors (through 53.13° for the record). These operations are shown on the right-hand side of Figure A.7 for confirmation.

In fact, *any* matrix multiplication may be thought of as a transformation of some coordinate space. This property of matrices has ensured their

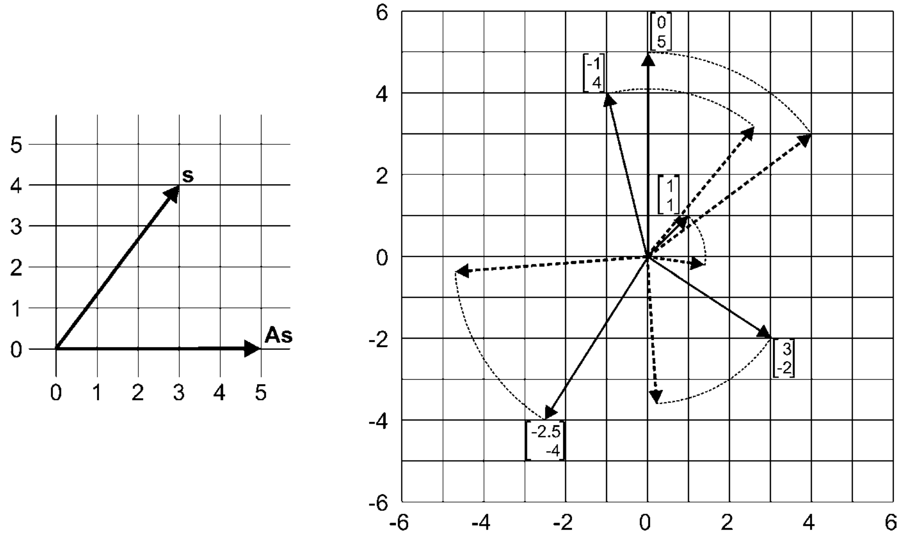

Figure A.7 Matrix multiplication as a transformation of coordinate space. In the left-hand grid, the multiplication **As** is shown. In the right-hand grid, each column of **S** is shown as a vector that is rotated after multiplication by **A** (see text).

widespread use in computer graphics, where they are an efficient way of doing the calculations required for drawing perspective views. Transformation matrices have the special property that they *project* the three dimensions of the objects displayed into the two dimensions of the screen. By changing the projection matrices used, we change the viewer's position relative to the displayed objects. This perspective on matrices is also important for transforming between geographic projections (see Chapter 11).

This perspective also provides an interpretation of the inverse of a matrix. Since multiplication of a vector **s** by a matrix, followed by multiplication by its inverse, returns **s** to its original value, the inverse of a matrix performs the opposite coordinate transformation to that of the original matrix. The inverse of the matrix above therefore performs a 53.13° *counterclockwise* rotation. You may care to try this on some examples.

A.7. EIGENVECTORS AND EIGENVALUES

Two properties important in statistical analysis are the *eigenvectors* and *eigenvalues* of a matrix. These only make intuitive sense in light of the geometric interpretation of matrices we have just introduced—although you will probably still find it a stretch. The eigenvectors $\{\mathbf{e}_1 \ldots \mathbf{e}_n\}$ and eigenvalues $\{\lambda_1 \ldots \lambda_n\}$ of an $n \times n$ matrix $\mathbf{A}$ each satisfy the following equation:

$$\mathbf{A}\mathbf{e}_i = \lambda\mathbf{e}_i \tag{A.63}$$

Seen in terms of the multiplication-as-transformation view, this means that the eigenvectors of a matrix are directions in coordinate space that are unchanged under transformation by that matrix. Note that the equation means that the eigenvalues and eigenvectors are associated with one another in pairs $\{(\lambda_1, \mathbf{e}_1), (\lambda_1, \mathbf{e}_1), \ldots (\lambda_n, \mathbf{e}_n)\}$. The scale of the eigenvectors is arbitrary, since they appear on both sides of the above equation, but normally they are scaled so that they have unit length. We won't worry too much about how the eigenvectors and eigenvalues of a matrix are determined (see Strang, 1988, for details). As an example, the eigenvalues and eigenvectors of the matrix in our simultaneous equations

$$\begin{bmatrix} 3 & 4 \\ 2 & -4 \end{bmatrix} \tag{A.64}$$

are

$$\left(\lambda_1 = 4,\ \mathbf{e}_1 = \begin{bmatrix} 0.9701 \\ 0.2425 \end{bmatrix}\right) \text{ and } \left(\lambda_2 = -5,\ \mathbf{e}_2 = \begin{bmatrix} -0.4472 \\ 0.8944 \end{bmatrix}\right) \tag{A.65}$$

It is straightforward to check this result by substitution into the defining equation above.

Figure A.8 may help to explain the meaning of the eigenvectors and eigenvalues. The unit circle shown is transformed to the ellipse shown under multiplication by the matrix we have been discussing. However, the eigenvectors have their direction unchanged by this transformation. Instead, they are each scaled by a factor equal to the corresponding eigenvalue.

An important result (again, see Strang, 1988) is that *the eigenvectors of a symmetric matrix are mutually orthogonal*. That is, if $\mathbf{A}$ is symmetric about its main diagonal, then any pair of its eigenvectors $\mathbf{e}_i$ and $\mathbf{e}_j$ have a dot product $\mathbf{e}_i^T\mathbf{e}_j = 0$. For example, the symmetric matrix

$$\begin{bmatrix} 1 & 3 \\ 3 & 2 \end{bmatrix} \tag{A.66}$$

has eigenvalues and eigenvectors

$$\left(4.541,\ \begin{bmatrix} 0.6464 \\ 0.7630 \end{bmatrix}\right) \text{ and } \left(-1.541,\ \begin{bmatrix} -0.7630 \\ 0.6464 \end{bmatrix}\right) \tag{A.67}$$

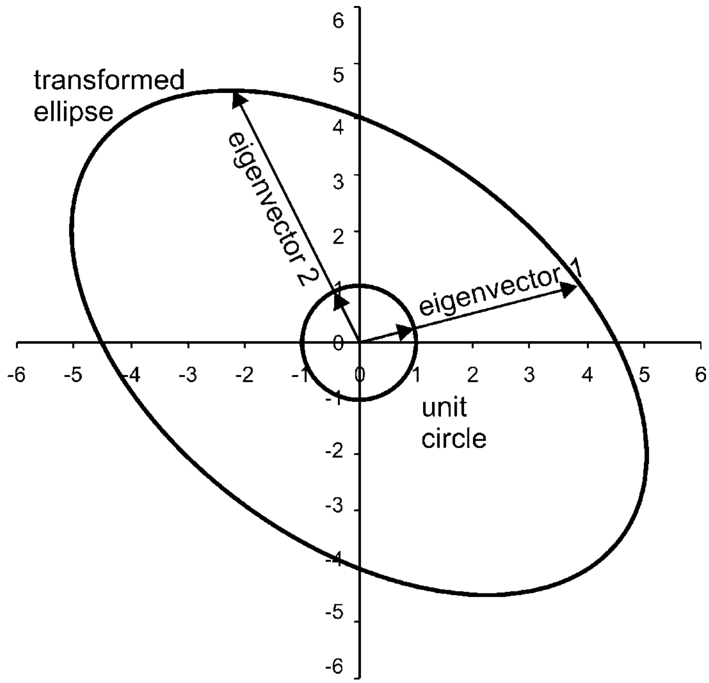

Figure A.8 The geometric interpretation of eigenvectors and eigenvalues (see text).

and it is easy to confirm that these vectors are orthogonal. The widely used method, *principal components analysis*, makes use of this result.

REFERENCES

Strang, G. (1988) *Linear Algebra and Its Applications*, 3rd ed., (Fort Worth, TX: Harcourt Brace Jovanovich).