Appendix A: Matrix Algebra Tutorial

Matrices and Their Basic Operations

A *matrix* is a two-dimensional ordered array of numbers. It can have any number of rows and columns, and it can contain any type of numbers (e.g., positive, negative, real, imaginary). The only requirement is that there are no missing values. The *order* of a matrix is its number of rows and columns. For example, the matrix

$$A = \begin{bmatrix} 2 & 3 & -1 \\ 0 & \frac{1}{2} & 4 \end{bmatrix}$$

is of order 2×3 because it has 2 rows and 3 columns. To enter this matrix into MATLAB, at the command line (denoted by >>) type

```
A = [2 3 -1; 0 .5 4]
```

Hitting the return key produces

```
        A =
2.0000  3.0000  -1.0000
0       0.5000   4.0000
```

By convention, matrices are typically identified by capital letters, and the order is sometimes expressed below this letter. For example, the 2×3 matrix A might be written as $\underset{2\times3}{A}$. A matrix in which the number of rows equals the number of columns is said to be *square* (i.e., the order is $n \times n$, for some value of n). A square matrix in which all entries not on the main diagonal equal 0 is called a *diagonal* matrix. For example,

$$D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -12 & 0 \\ 0 & 0 & \frac{1}{3} \end{bmatrix}$$

is a 3×3 diagonal matrix.

A matrix with one row or one column is called a *vector*. A column vector has a single column, and a row vector has a single row. Vectors are traditionally identified by lowercase underlined letters. So, for example, a 3×1 column vector might be written as

$$\underline{v} = \begin{bmatrix} 2 \\ -7 \\ 1 \end{bmatrix}.$$

Finally, within matrix algebra, single numbers are referred to as *scalars*.

A variety of mathematical operations can be performed on matrices and vectors. Although addition and multiplication are included in this list, not all pairs of matrices can be added or multiplied. In other words, certain conditions must be met before matrix addition or multiplication is defined. These conditions are different for addition and multiplication. Pairs of matrices that satisfy these conditions are said to be *conformable* for that operation.

Two matrices A and B are conformable for addition if and only if they have the same order. Matrices of different orders cannot be added. If A and B have the same order, then the sum C = A + B is defined as the matrix containing the term-by-term sums of the entries in A and B. Thus, if A and B are both 2×2, then

$$A + B = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} \\ a_{21} + b_{21} & a_{22} + b_{22} \end{bmatrix} = C.$$

In MATLAB, the command A + B will compute the sum, provided it exists. Note that matrix addition is commutative; that is, if A + B = C, then A and B are also conformable for the sum B + A, and B + A = A + B = C.

Suppose A is of order $n \times m$ and B is of order $p \times q$. Then A and B are conformable for the product AB if and only if $m = p$; that is, if and only if the number of columns of the pre-multiplier equals the number of rows of the post-multiplier. If this condition is met, then the product C = AB is of order $n \times q$. So C has the same number of rows as the pre-multiplier and the same number of columns as the post-multiplier. If the product C = AB exists, then the entry in row i and column j equals the sum of the term-by-term multiplication of the entries in row i of the pre-multiplier and column j of the post-multiplier. For example, if A is 3×2 and B is 2×2, then the product is of order 3×2 and is defined as

$$\underset{3\times 2}{A}\ \underset{2\times 2}{B} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \\ a_{31}b_{11} + a_{32}b_{21} & a_{31}b_{12} + a_{32}b_{22} \end{bmatrix} = \underset{3\times 2}{C}.$$

As numerical examples, note that

$$\begin{bmatrix} 2 & 3 \\ 4 & -1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 2 & 0 \end{bmatrix} = \begin{bmatrix} 8 & -2 \\ 2 & -4 \end{bmatrix} \tag{A.1}$$

and

$$[2 \quad 1] \begin{bmatrix} 3 & 0 \\ 4 & 1 \end{bmatrix} = [10 \quad 1].$$

To perform the equation A.1 multiplication in MATLAB, one first defines the two matrices

```
>> A = [2 3;4 -1];
>> B = [1 -1;2 0];
```

The semicolon at the end of each line suppresses printing. Next type

```
>> A*B
```

Hitting the return key produces the result

```
ans =
8     -2
2     -4
```

Unlike addition, matrix multiplication is not commutative. In fact, if A is $n \times m$ and B is $m \times p$, then the product AB is defined (and is of order $n \times p$), but note that BA is not defined unless $p = n$. Even in this case, however, AB will generally not equal BA. Because of this, great care must be taken about the order in which one writes the various terms in each product.

Any matrix can be multiplied by a scalar in an operation known as scalar multiplication. The scalar simply multiplies each entry in the matrix. For example,

$$5 \begin{bmatrix} 1 & 2 \\ 3 & 0 \end{bmatrix} = \begin{bmatrix} 5 & 10 \\ 15 & 0 \end{bmatrix}.$$

This scalar multiplication is done in MATLAB via the following commands:

```
>> A = [1 2;3 0];
>> 5*A
```

Another useful operation is matrix transposition. The *transpose* of a matrix A, denoted by A′, is created by switching the rows and columns of A. Specifically, the i^{th} row of A becomes the i^{th} column of A′ (for all i). So if A is $n \times m$, then A′ is $m \times n$. For example,

$$\begin{bmatrix} 2 & 1 & -3 \\ 0 & 4 & 1 \end{bmatrix}' = \begin{bmatrix} 2 & 0 \\ 1 & 4 \\ -3 & 1 \end{bmatrix}.$$

MATLAB will produce the transpose of any matrix A simply by typing

>> A′

One useful rule regarding transposes is that the transpose of a product equals the product of the transposes in reverse order. So, for example,

$$\underset{n\times m}{\mathrm{A}}\ \underset{m\times p}{\mathrm{B}}\Big)' = \underset{p\times m}{\mathrm{B}'}\ \underset{m\times n}{\mathrm{A}'}.$$

Note that if the product AB is defined, then B′A′ must also be defined.

By convention, the transpose is also used to denote a row vector. Standard notation is to interpret a vector $\underline{v}$ as a column vector. If so, then to denote a row vector one would write $\underline{v}'$.

Matrix algebra is especially useful for simplifying and solving systems of simultaneous linear equations, as, for example, one finds in the GLM. To see how this is done, consider the equations

$$x - y + 2z = 2$$

$$3x + y - z = 4$$

$$5x + 2y - z = 9$$

Note that these equations can be rewritten in matrix form as

$$\begin{bmatrix} 1 & -1 & 2 \\ 3 & 1 & -1 \\ 5 & 2 & -1 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \\ 9 \end{bmatrix},$$

which we can rewrite in shorthand form as

$$\mathrm{A}\underline{x} = \underline{b} \qquad \text{(A.2)}$$

where

$$\mathrm{A} = \begin{bmatrix} 1 & -1 & 2 \\ 3 & 1 & -1 \\ 5 & 2 & -1 \end{bmatrix}, \quad \underline{x} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}, \quad \text{and} \quad \underline{b} = \begin{bmatrix} 2 \\ 4 \\ 9 \end{bmatrix}.$$

If equation A.2 was a univariate (i.e., scalar) algebraic equation, we would easily solve for $\underline{x}$ by dividing both sides by A. This does not work with matrix equations, though, because matrix division is not defined. However, note that dividing both sides of the scalar equation

$$ax = b$$

by a is the same as multiplying both sides by the inverse of a:

$$a^{-1}ax = a^{-1}b, \quad \text{which implies that } x = a^{-1}b.$$

Fortunately, the inverse of a matrix is defined, at least under certain special conditions.

The value 1/2 is the multiplicative inverse of 2 because their product is the identity element 1, and 1 is the multiplicative identity because

$$1 \times x = x \times 1 = x,$$

for any value of x. So to define a matrix inverse, we must first define an identity matrix. More specifically, we seek a matrix I such that

$$\mathrm{IA} = \mathrm{AI} = \mathrm{A}. \qquad \text{(A.3)}$$

Note that the only way that both products IA and AI are defined and equal to each other is if A and I are both square and of the same order (i.e., $n \times n$). For any value of n, it can be shown that the only matrix I that satisfies equation A.3 is the $n \times n$ matrix

$$\mathrm{I} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix},$$

which is a diagonal matrix (nonzero values only appear on the main diagonal) with every entry on the main diagonal equal to 1. To see that I satisfies equation A.3, note, for example, that

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & -1 & 2 \\ 3 & 1 & -1 \\ 5 & 2 & -1 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 2 \\ 3 & 1 & -1 \\ 5 & 2 & -1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 2 \\ 3 & 1 & -1 \\ 5 & 2 & -1 \end{bmatrix}.$$

In MATLAB, the $n \times n$ identity matrix is constructed via the command eye(n). For example, to construct a 3 × 3 identity matrix, type the command

>> I = eye(3)

Hitting the return key produces

I =

1 0 0
0 1 0
0 0 1

The identity matrix can be used to find the inverse of a matrix. If A is $n \times n$, then we seek another $n \times n$ matrix, which we denote A^{-1}, for which

$$A^{-1}A = AA^{-1} = I. \tag{A.4}$$

If A is the 2×2 matrix

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix},$$

then it turns out that

$$A^{-1} = \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}. \tag{A.5}$$

We can verify that this works, for example, by computing

$$A^{-1}A = \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

$$= \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{bmatrix} a_{11}a_{22} - a_{12}a_{21} & 0 \\ 0 & a_{11}a_{22} - a_{12}a_{21} \end{bmatrix}.$$

$$= I$$

MATLAB computes the inverse of a matrix A, if it exists, via the command inv(A). For example, to compute the inverse of the matrix

$$A = \begin{bmatrix} 7 & 9 \\ 3 & 4 \end{bmatrix}$$

Type

```
>> A = [7 9;3 4];
>> inv(A)
```

This produces the result

```
                ans =
 4.0000    -9.0000
-3.0000     7.0000
```

Note that the inverse defined by equation A.5 exists only if the denominator of the scalar multiple is nonzero; that is, only if

$a_{11}a_{22} - a_{12}a_{21} \neq 0$.

This value is so important that it is given its own name, the *determinant*, which is written as |A|. Specifically, the determinant of a 2×2 matrix A is defined as

$$|A| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}.$$

It turns out that every square matrix has a determinant, and every square matrix has an inverse if and only if its determinant is nonzero. For matrices 3×3 or larger, computing a determinant or inverse can be time consuming and tedious. With MATLAB, however, these computations are trivial. For any square matrix A, det(A) returns the determinant and inv(A) returns the inverse (if it exists). For example, the determinant of the 3×3 matrix A from equation A.2 is

$$|A| = \begin{vmatrix} 1 & -1 & 2 \\ 3 & 1 & -1 \\ 5 & 2 & -1 \end{vmatrix} = 5.$$

Therefore A has an inverse.

The following commands compute this determinant in MATLAB:

```
>> A = [1 -1 2;3 1 -1; 5 2 -1];
>> det(A)
```

which produces

```
ans =
5
```

The inverse of this matrix is computed from

```
>> inv(A)
```

which produces

```
                ans =
 .2          .6        -.2
-.44        -2.2        1.4
 .2         -1.4         .8
```

In summary, any matrix A has an inverse if and only if two conditions are met. First, A must be square, and second, the determinant of A must be nonzero. A matrix with an inverse is said to be *nonsingular*, whereas a square matrix without an inverse is *singular*. So

a matrix A is nonsingular if and only if $|A| \neq 0$. Another important property of the inverse is that if it does exist, then it is unique. In other words, for a matrix A satisfying these two conditions there exists only one matrix A^{-1} for which

$$A^{-1}A = AA^{-1} = I.$$

Rank

A set of vectors is said to be *linearly independent* if and only if it is impossible to write any one of them as a weighted linear combination of the others. A set of vectors that are not linearly independent are said to be *linearly dependent*. For example,

$$\begin{bmatrix}1\\2\\3\end{bmatrix}, \begin{bmatrix}1\\1\\1\end{bmatrix}, \text{ and } \begin{bmatrix}3\\4\\5\end{bmatrix}$$

are linearly dependent because

$$\begin{bmatrix}3\\4\\5\end{bmatrix} = \begin{bmatrix}1\\2\\3\end{bmatrix} + 2\begin{bmatrix}1\\1\\1\end{bmatrix}.$$

As another example,

$$\begin{bmatrix}1\\2\end{bmatrix} \text{ and } \begin{bmatrix}2\\3\end{bmatrix}$$

are linearly independent because neither one is a scalar multiple of the other.

Every matrix can be considered either as a collection of row vectors or column vectors. A well-known result in matrix algebra is that the number of linearly independent columns in any matrix must equal the number of linearly independent rows. For example, consider the matrix

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

As column 2 is not a scalar multiple of column 1, there are two linearly independent columns in this matrix. Therefore, there must also be two linearly independent rows. The first two rows are linearly independent because the second row is not a scalar multiple of the first. Row 3, however, equals

$$\begin{bmatrix}-1\\3\end{bmatrix} = 2\begin{bmatrix}1\\2\end{bmatrix} - \begin{bmatrix}3\\1\end{bmatrix},$$

so row 3 is not linearly independent of rows 1 and 2. Similarly, note that row 4 equals row 2 minus row 1. Therefore, there are also two linearly independent rows in this matrix.

The *rank* of a matrix equals the number of linearly independent rows or columns. So for example, the rank of A in the equation A.6 matrix is 2. Rank is a useful construct that has a number of important applications. For example, as we will see, the number of solutions of any set of simultaneous linear equations can be determined by comparing the ranks of two appropriate matrices. Computing the rank of any matrix in MATLAB is simple. For example, the rank of the equation A.6 matrix is computed via

```
>> A = [1 2;3 1;-1 3;2 -1];
>> rank(A)
```

which produces

```
ans =
2
```

Computing the rank of a matrix by hand can be difficult. The following properties of rank, however, can simplify this process.

Property 1 *The rank of a matrix equals 0 if and only if every entry in the matrix is 0.*

Property 2 *If A is of order $n \times m$, then rank(A) $\leq$ min(n,m).* This result follows because the number of linearly independent rows equals the number of linearly independent columns. So for example, if a matrix has fewer rows than columns (i.e., so $n < m$), then at most there are n linearly independent rows, and therefore also at most n linearly independent columns. A corollary to this result says that the maximum rank of an $n \times n$ square matrix is n. In this case, note that all rows and all columns are linearly independent. An $n \times n$ square matrix with rank n is said to be *full rank*.

Property 3 *If rank(A) = r, then there must exist an $r \times r$ submatrix of full rank.* A submatrix is created by striking out any number of rows or columns. For example, the rank of the matrix

$$A = \begin{bmatrix}1 & 1 & 0 & 0\\4 & 7 & 3 & 0\\0 & 1 & 1 & 0\\0 & 0 & 0 & 1\end{bmatrix}$$

is no greater than 3 because column 2 is the sum of columns 1 and 3. A 3×3 submatrix of full rank can be created by striking out row 2 and column 2. Note that this process leaves the 3×3 identity matrix, which is full rank. Therefore, this matrix has rank 3.

Property 4 *Suppose A is $n \times n$. Then rank(A) = n if and only if $|A| \neq 0$.* This is a very important property. Note that it provides another way to determine whether a matrix is nonsingular (i.e., has an inverse).

In summary, if A is a square $n \times n$ matrix, then the following statements are all equivalent.

1. rank(A) = n.
2. $|A| \neq 0$.
3. A is nonsingular (i.e., A^{-1} exists).

Similarly, the following statements are also equivalent.

1. rank(A) < n.
2. $|A| = 0$.
3. A is singular.

Solving Linear Equations

Any set of simultaneous linear equations must have 0, 1, or an infinite number of solutions. Examples of these three possibilities are shown in figure A.1. For example, the equations

$x + y = 0$

$x + y = 2$

have zero solutions because if $x + y$ equals 0, it cannot also equal 2. Graphically, these equations describe parallel lines with slope −1 and y-intercepts 0 and 2 (see the top panel of figure A.1). A solution to these equations would be a point (x, y) that falls on both lines and therefore simultaneously satisfies both equations. Of course, parallel lines share no points in common, so these equations have no solution.

Simultaneous equations with no solutions are said to be *inconsistent*. If at least one solution exists, then the equations are *consistent*. With linear equations, there are only two possibilities if the equations are consistent. They either have one solution or they have an infinite number of solutions.

The middle panel of figure A.1 shows an example of equations with one solution:

$x - y = 0$

$x + y = 2.$

Solving these equations produces $x = 1$ and $y = 1$, and figure A.1 shows that this is the point where the two lines intersect. In contrast, the equations

$x + y = 2$

$2x + 2y = 4$

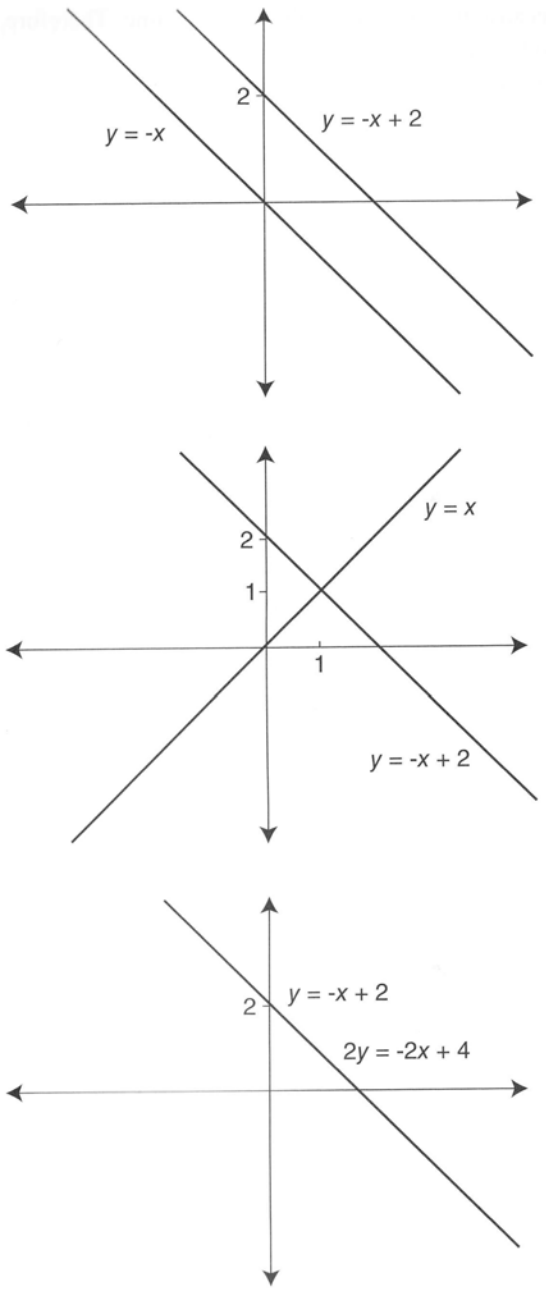

Figure A.1
Three possible outcomes when trying to solve a set of simultaneous linear equations. Either there are 0 solutions (top), one solution (middle), or an infinite number of solutions (bottom).

have an infinite number of solutions because they both describe the same line. Therefore, any point on this line is a solution of both equations.

When faced with a set of simultaneous linear equations therefore, our first question is to ask whether they are consistent. If not, then nothing more can be done. If they are consistent, then the next question is to ask whether they have one solution or an infinite number. The question of consistency can be answered by comparing the ranks of two different matrices. Specifically, the equations

$$A\underline{x} = \underline{b}$$

are consistent if and only if

$$\text{rank}(A) = \text{rank}(A \vdots \underline{b}), \tag{A.7}$$

where $A \vdots \underline{b}$ is the matrix A augmented with the vector $\underline{b}$. For example, if A is $n \times n$, then $\underline{x}$ and $\underline{b}$ must both be $n \times 1$. The matrix $A \vdots \underline{b}$, which is of order $n \times n + 1$, contains A in its first n columns and $\underline{b}$ in column $n + 1$.

If A is $n \times n$, then $\text{rank}(A) \leq n$, and adding another column to A cannot decrease its rank. Therefore, note that equation A.7 holds if $\underline{b}$ is linearly dependent on the columns of A and it is violated if $\underline{b}$ is linearly independent of the columns of A. Furthermore, note that by rank property no. 2, $\text{rank}(A \vdots \underline{b}) \leq n$ (because this augmented matrix has only n rows). So if A is full rank [i.e., $\text{rank}(A) = n$], then because adding $\underline{b}$ cannot reduce the rank of $A \vdots \underline{b}$, equation A.7 must hold. In other words, the equations $A\underline{x} = \underline{b}$ are always consistent if A is full rank. This means that if $\underline{x}$ and $\underline{b}$ are $n \times 1$, then the only conditions under which the equations $A\underline{x} = \underline{b}$ are not consistent is if $\text{rank}(A) < n$.

To illustrate these results, consider our earlier examples. First, rewriting

$$x + y = 0 \quad \text{and} \quad x + y = 2$$

in matrix form produces

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \end{bmatrix}.$$

Note that

$$\text{rank}(A) = \text{rank}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = 1 \quad \text{and} \quad \text{rank}(A \vdots \underline{b}) = \text{rank}\begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 2 \end{bmatrix} = 2,$$

so these equations are not consistent. As another example, consider the equations

$$x + y = 2 \quad \text{and} \quad 2x + 2y = 4.$$

In matrix form these become

$$\begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}.$$

Because

$$\text{rank}\begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix} = \text{rank}\begin{bmatrix} 1 & 1 & 2 \\ 2 & 2 & 4 \end{bmatrix} = 1,$$

these equations are consistent.

If the equations $A\underline{x} = \underline{b}$ are consistent, then they have either one solution or an infinite number. The difference depends on A. If A is square and full rank, then A^{-1} exists and there is only one solution, which is easily found by solving for $\underline{x}$:

$$A\underline{x} = \underline{b}$$

implies that

$$A^{-1}A\underline{x} = A^{-1}\underline{b}$$

and so

$$\underline{x} = A^{-1}\underline{b}. \tag{A.8}$$

For example, consider the three simultaneous equations described by equation A.2. Using A.8 produces the following unique solution to these equations:

$$\underline{x} = \begin{bmatrix} 1 & -1 & 2 \\ 3 & 1 & -1 \\ 5 & 2 & -1 \end{bmatrix}^{-1}\begin{bmatrix} 2 \\ 4 \\ 9 \end{bmatrix} = \begin{bmatrix} .2 & .6 & -.2 \\ -.4 & -2.2 & 1.4 \\ .2 & -1.4 & .8 \end{bmatrix}\begin{bmatrix} 2 \\ 4 \\ 9 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}.$$

We can verify that $x = 1$, $y = 3$, and $z = 2$ is the solution by substituting these values back into the original equations. MATLAB solves these equations via the commands

```
>> A = [1 -1 2;3 1 -1;5 2 -1];
>> b = [2; 4; 9];
>> x = inv(A)*b
```

The result is

```
x =
1.0000
3.0000
2.0000
```

If the equations are consistent but A is not square and full rank, then there are an infinite number of solutions. In this case A has no inverse and may not even be a square matrix. The

infinite number of solutions that exist can be found by using a generalization of the inverse called the generalized inverse, which is often denoted by A^-. Every matrix has a generalized inverse, even matrices that are not square. For details on using the generalized inverse to solve linear equations, see for example, Searle (1966).

Eigenvalues and Eigenvectors

One other useful topic in matrix algebra, which forms the basis of PCA (see chapter 10) for example, is eigenvalues and eigenvectors.

Definitions

Consider an $n \times n$ matrix A. Suppose we are able to find an $n \times 1$ vector $\underline{v}$ and a scalar d such that

$$A\underline{v} = d\underline{v}, \tag{A.9}$$

or in other words, when we post-multiply A by the vector $\underline{v}$, the result is still $\underline{v}$, except scaled by the constant d. In such a case, the vector $\underline{v}$ is called an *eigenvector* of A and the constant d is called an *eigenvalue*. At least at the level of mathematics considered in this book, equation A.9 by itself does not offer any profound insights into the matrix A. So the definition of eigenvalues and eigenvectors is not particularly illuminating. Even so, eigenvectors and eigenvalues have many properties that are extremely useful, and it is for these properties, rather than for the definition, that eigenvectors and eigenvalues are so frequently used in statistics.

Next we consider methods for finding the eigenvectors and eigenvalues of a matrix. If equation A.9 holds, then

$$A\underline{v} - d\underline{v} = \underline{0},$$

where $\underline{0}$ is a vector of all zeroes. And therefore,

$$(A - dI)\underline{v} = \underline{0}. \tag{A.10}$$

Note that we needed to add the matrix I to make the difference inside the parentheses conformable.

An obvious solution to equation A.10 is that $\underline{v} = \underline{0}$, but this is not interesting because it is a solution no matter what the matrix A, and for this reason, it certainly cannot tell us anything useful about A. Thus, the only solutions of equation A.10 that could be of interest are when $\underline{v} \neq \underline{0}$. Now if $(A - dI)$ is nonsingular (i.e., full rank), then $\underline{v} = \underline{0}$ is the only solution. For this reason, we are interested in finding values of d that make $(A - dI)$ singular. As we saw earlier, a square matrix is singular if and only if its determinant is zero. Therefore, our task is to find values of d for which

$$|A - dI| = 0. \tag{A.11}$$

This is called the *characteristic equation* of the matrix A.

Note that unlike equation A.10, both sides of equation A.11 are scalars (rather than vectors). In fact, if A is $n \times n$, then its characteristic equation is an n^{th} order polynomial. For example, in the case of the matrix

$$A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix},$$

the characteristic equation is

$$0 = \left| \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} - \begin{bmatrix} d & 0 \\ 0 & d \end{bmatrix} \right|,$$

$$= \begin{vmatrix} 3-d & 1 \\ 1 & 3-d \end{vmatrix}$$

$$= (3-d)^2 - 1$$

$$= d^2 - 6d + 8$$

$$= (d-4)(d-2).$$

The two roots of this quadratic equation and, consequently, the two eigenvalues of A are $d = 4$ and $d = 2$. MATLAB will produce these eigenvalues in response to the command eig(A). For example, the commands

```
>> A = [3 1;1 3];
>> eig(A)
```

produce

```
ans =
2
4
```

Once the eigenvalues are computed, the eigenvectors can be determined by solving equations A.10. In the current example

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} = \left(\begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} - \begin{bmatrix} d & 0 \\ 0 & d \end{bmatrix} \right) \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

$$= \begin{bmatrix} 3-d & 1 \\ 1 & 3-d \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}.$$

We begin by substituting in the first eigenvalue $d = 4$:

$$\begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

which leads to two equations and two unknowns

$$-v_1 + v_2 = 0 \quad \text{and} \quad v_1 - v_2 = 0. \tag{A.12}$$

Note that these two equations have an infinite number of solutions. Of course, this was inevitable because we selected the eigenvalues $d = 4$ and $d = 2$ precisely because they are the only possible values of d that lead to an infinite number of solutions of equations A.10.

Any set of v_1 and v_2 that satisfy equations A.12 define a legitimate eigenvector of the matrix A. Note that all such solutions fall on the line

$$v_2 = v_1.$$

Any vector can be considered as a directed line segment beginning at the origin (0,0) and ending at the vector coordinates. Thus, although there are an infinite number of solutions to equations A.12, they all point in the same direction. They differ only in length. The convention is to choose a solution so that the resulting eigenvector has a length of 1; that is, to choose v_1 and v_2 so that $\underline{v}'\underline{v} = 1$.

An easy way to do this is to choose any solution, compute $\underline{v}'\underline{v}$, and then divide both v_1 and v_2 by the square root of this value. In the example of equations A.12, we could choose $v_1 = 1$ and $v_2 = 1$.

Then

$$\underline{v}'\underline{v} = [1 \quad 1] \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 2,$$

and so the eigenvector associated with the eigenvalue $d = 4$ is

$$\underline{v}_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}.$$

We use the subscript 1 to signify that this is the eigenvector associated with the first, or largest, eigenvalue. We can verify that the length of this eigenvector is 1 via

$$\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} = \frac{1}{2} + \frac{1}{2} = 1.$$

By a similar process, we can determine that the eigenvector associated with the second eigenvalue $d = 2$ is

$$\underline{v}_2 = \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}.$$

Properties

As previously mentioned, the definitions of eigenvalues and eigenvectors do not lead to any immediate insights into the matrix A. However, eigenvalues and eigenvectors have many properties that are extremely useful in a wide variety of applications, and it is for these properties that eigenvalues and eigenvectors are considered core topics in matrix or linear algebra. As we saw in chapter 10, several of these properties lie at the heart of PCA.

This section describes a few of the most important properties of eigenvalues and eigenvectors. Any text on linear algebra will include a variety of others.

Property 1 *An $n \times n$ matrix has n eigenvalues, some of which may equal 0 and some of which may be repeated.* This follows because the characteristic equation of an $n \times n$ matrix is an n^{th} order polynomial, which has n roots. In our earlier numerical example, the 2×2 matrix

$$A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$$

has two eigenvalues; namely, 4 and 2.

Property 2 *The number of nonzero eigenvalues of A equals the rank of A.* This property is extremely useful because it can be quite tedious to compute the rank of a matrix by determining the number of linearly independent rows or columns. Note that an important corollary to this property is that A is singular if and only if A has at least one eigenvalue equal to 0. In our earlier numerical example, A had two eigenvalues $d = 4$ and $d = 2$. These are both nonzero, so the rank of A is 2 and therefore A is nonsingular.

Property 3 *The determinant of A equals the product of its eigenvalues.* So in our numerical example, the determinant of A must equal $4 \times 2 = 8$. Because A is 2×2, this is easily verified:

$$\begin{vmatrix} 3 & 1 \\ 1 & 3 \end{vmatrix} = (3 \times 3) - (1 \times 1) = 8.$$

With much larger matrices, however, computing the determinant directly is a time-consuming process. If the eigenvalues are known, property no. 3 makes this computation simple. Note also that if one of the eigenvalues of A is 0, then their product will be zero, and hence the determinant of A will be 0. In other words, if A has one or more eigenvalues equal to 0, then A must be singular—a result that also followed from property no. 2.

Property 4 *The trace of A equals the sum of its eigenvalues.* The trace of a matrix is equal to the sum of all elements on the main diagonal. So

$$\text{trace}\left(\begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}\right) = 3+3 = 6,$$

and note that the sum of the eigenvalues of this matrix is also 6 (i.e., 4 + 2).

Property 5 Diagonal Representation of a Symmetric Matrix. Suppose A is a symmetric $n \times n$ matrix. Construct the $n \times n$ diagonal matrix D that has the eigenvalues of A on its main diagonal (e.g., in descending order of magnitude). Construct the $n \times n$ square matrix V whose columns are the eigenvectors of A (so that the i^{th} column of V is the eigenvector that corresponds with the eigenvalue in row i and column i of the matrix D). Then A = VDV′.

In our numerical example, the diagonal representation of A is given by

$$A = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

$$= \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}.$$

A matrix in which all eigenvalues are positive is said to be *positive definite*, and a matrix in which all eigenvalues are non-negative (e.g., some may be zero) is said to be *positive semi-definite*. Note that, in this example, A is positive definite. PCA works on the eigenvalues and eigenvectors of the sample variance-covariance matrix. All variance-covariance matrices are symmetric and positive semidefinite.

MATLAB computes eigenvectors (and eigenvalues) using a similar form. Specifically, the command

```
[V,D] = eig(A)
```

returns a matrix V whose columns are the eigenvectors of A and a diagonal matrix D containing the eigenvalues of A. For example, the commands

```
>> A = [3 1;1 3];
>> [V,D] = eig(A)
```

produce the output

```
V =
-0.7071    0.7071
 0.7071    0.7071

D =
 2     0
 0     4
```

Appendix B: Multivariate Probability Distributions

A *random vector* is a vector in which every entry is a random variable. For example, consider the vector

$$\underline{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_r \end{bmatrix}.$$

If $\underline{x}$ is a random vector, then each x_i is a random variable. Let $f_i(x_i)$ denote the *probability density function* (pdf) of x_i. This function specifies the likelihood that a random sample drawn from the x_i population exactly equals any specific numerical value.[1] If x_i is normally distributed, then $f_i(x_i)$ is the familiar bell-shaped curve. With respect to the random vector $\underline{x}$, the pdfs $f_i(x_i)$ are known as the *marginal distributions*.

The marginal distributions of $\underline{x}$ provide much information about the sampling behavior of $\underline{x}$, but they do not tell us everything. In particular, they provide no information about any statistical relationships that might exist among the various x_i. Complete information about $\underline{x}$ is catalogued in the *joint probability density function* (or joint distribution)

$$f(x_1, x_2, \ldots, x_r) = f(\underline{x}).$$

This function specifies the likelihood that a random sample from the $\underline{x}$ population will produce any specific $r \times 1$ numerical vector.

If there is no statistical relationship among any of the x_i, then all information in the joint pdf is specified by the marginal pdfs. More specifically, the random variables $x_1, x_2, \ldots, x_r$ are *statistically independent* if and only if

1. This notation is sloppy because it does not discriminate between the name of the random variable or random vector and the specific numerical values that the random variable or vector can take. The current notation is simpler, and hopefully it is obvious from the context which interpretation is intended.

$$f(x_1, x_2, \ldots, x_r) = f_1(x_1) \times f_2(x_2) \times \cdots \times f_r(x_r), \tag{B.1}$$

for all possible values of $x_1, x_2, \ldots, x_r$. If equation B.1 fails for any combination of $x_1, x_2, \ldots, x_r$, then a statistical dependence exists among these random variables.

Multivariate Normal Distributions

The *multivariate normal distribution* is, by far, the most widely used multivariate distribution in statistics. For example, it serves as the error model in the GLM and as the model that underlies PCA. A multivariate normal distribution has three assumptions: (1) the marginal distributions are all normal; (2) the only possible relationships among the x_i are linear; and (3) all dependencies among the x_i can be expressed as a function of the dependencies between all possible pairs of x_i (i.e., there are no dependencies that depend on three-way or higher interactions). Thus, even if the x_i are each normally distributed, the random vector $\underline{x}$ is not necessarily multivariate normally distributed. In addition, it must also be true that the only possible statistical dependencies that exist among the x_i are pairwise linear relationships.

The well-known Pearson correlation coefficient (i.e., the Pearson's r) measures linear relationships between pairs of variables. This is the model of statistical dependence that underlies the multivariate normal distribution. Uncorrelated random variables have no linear relationship, but they could have a nonlinear relationship, in which case they would not be statistically independent (i.e., equation B.1 would not hold). Statistical independence implies zero correlation, but uncorrelated random variables are not necessarily independent. In a multivariate normal distribution, however, the only possible relationships are linear, so uncorrelated is equivalent to independent.

In the multivariate normal distribution, there is a mean and variance associated with each (random) variable and a correlation associated with each pair of variables. Let μ_i and σ_i^2 denote the mean and variance of x_i, respectively. The correlation between random variables x_1 and x_2 is defined as the standardized covariance:

$$\rho_{12} = \frac{\text{cov}_{12}}{\sigma_1 \sigma_2} = \frac{E[(x_1 - \mu_1)(x_2 - \mu_2)]}{\sigma_1 \sigma_2}. \tag{B.2}$$

If the means and variances are known, then note that it makes no difference whether we characterize the associations of a multivariate normal distribution in terms of correlations or covariances. From either one, equation B.2 allows us to solve for the other. The standard convention is to record the covariances.

The parameters of any multivariate normal distribution are catalogued in two structures: a mean vector $\underline{\mu}$ and a variance-covariance matrix Σ. The mean vector is a record of the mean of each marginal distribution,

$$\underline{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_r \end{bmatrix}, \tag{B.3}$$

and the variance-covariance matrix is a record of all variances and covariances,

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \text{cov}_{12} & \cdots & \text{cov}_{1r} \\ \text{cov}_{21} & \sigma_2^2 & \cdots & \text{cov}_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}_{r1} & \text{cov}_{r2} & \cdots & \sigma_r^2 \end{bmatrix}. \tag{B.4}$$

Because $\text{cov}_{ij} = \text{cov}_{ji}$, note that this is a symmetric matrix. It is also positive semidefinite (i.e., no eigenvalues can be negative; see appendix A).

Once numerical values are specified for the mean vector and the variance-covariance matrix, then the likelihood of any vector $\underline{x}$ can be computed from the multivariate normal pdf:

$$f(\underline{x}) = \frac{1}{(2\pi)^{r/2} |\Sigma|^{1/2}} \exp[-\tfrac{1}{2}(\underline{x} - \underline{\mu})' \Sigma^{-1} (\underline{x} - \underline{\mu})]. \tag{B.5}$$

Figure B.1 shows an example of this pdf for a bivariate normal distribution where the correlation between x_1 and x_2 is positive. The bottom panel shows some contours of equal likelihood from this distribution, which are created by slicing through the pdf shown in the top panel from different heights above the (x_1, x_2) plane and looking down at the results from above. Note that these contours all have the same shape and differ only in size. A scatterplot of random samples from the distribution shown in the top panel would have the same overall shape as these contours. The positive correlation causes the major axis of the contours to have a positive slope. Note that random samples from the distribution that have a large x_1 value will also tend to have a large x_2 value.

A special case of the multivariate normal distribution that is widely used throughout this book assumes that all variables are independent and all variances are equal. In this case, note that

$$\Sigma = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix} = \sigma^2 \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} = \sigma^2 \text{I}.$$

The multivariate z-distribution is a special case of this in which the mean vector equals 0 and the variance-covariance matrix equals I.

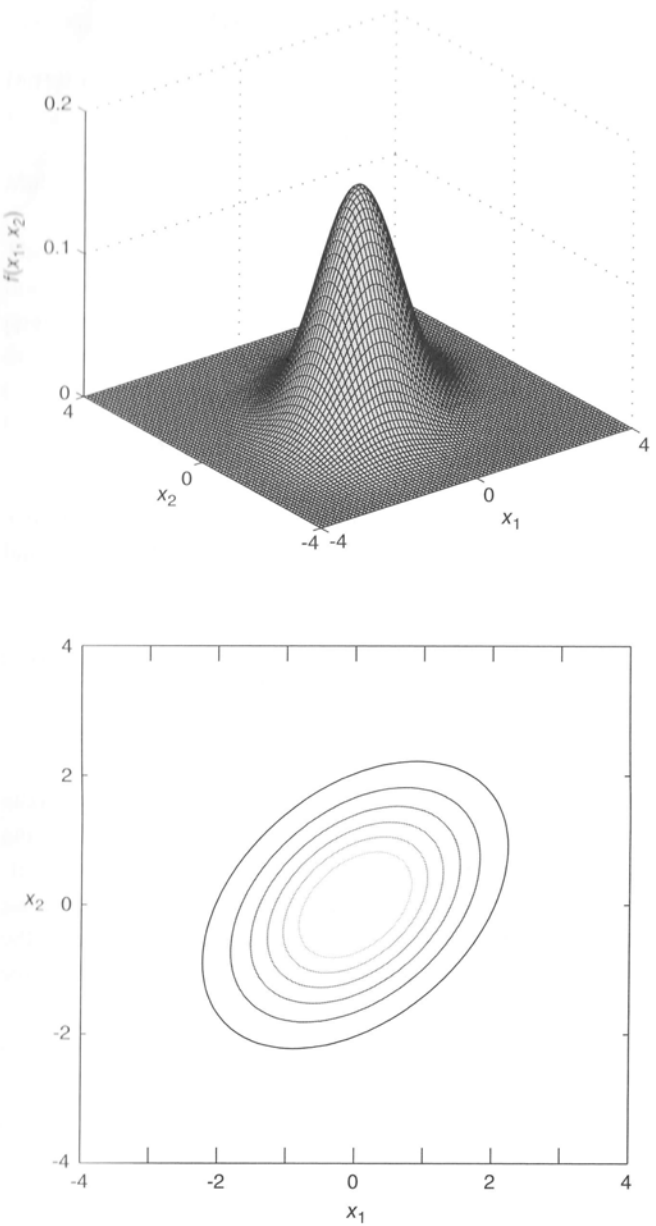

Figure B.1
(Top) The pdf of a bivariate normal distribution. (Bottom) Contours of equal likelihood from the pdf shown in the top panel. Note the positive correlation between x_1 and x_2.

Frequently in probability and statistics, and also in this book, we are interested in the distribution of a linear transformation of a random vector. More specifically, suppose $\underline{x}$ is an $r \times 1$ random vector, A is an $m \times r$ matrix of constants, and $\underline{b}$ is an $m \times 1$ vector of constants. Now consider the $m \times 1$ random vector

$$\underline{y} = A\underline{x} + \underline{b}.$$

Then regardless of the distribution of $\underline{x}$, the mean vector and variance-covariance matrix of $\underline{y}$ are equal to

$$\underline{\mu}_{\underline{y}} = A\underline{\mu}_{\underline{x}} + \underline{b}$$

and

$$\Sigma_{\underline{y}} = A\Sigma_{\underline{x}}A'.$$

Furthermore, if $\underline{x}$ has a multivariate normal distribution then $\underline{y}$ will also have a multivariate normal distribution (because linear transformations of normal random variables are normal).