

Introduction to Matrix Analysis

Introduction

These notes provide an introduction to the use of matrices in engineering analysis. Matrix notation is used to simplify the representation of systems of linear algebraic equations. In addition, the matrix representation of systems of equations provides important properties regarding the system of equations. The discussion here presents many results without proof. You can refer to a general advanced engineering math text¹ or a text on linear algebra for such proofs.

Parts of these notes have been prepared for use in a variety of courses to provide background information on the use of matrices in engineering problems. Consequently, some of the material may not be used in this course and different sections from these notes may be assigned at different times in the course.

Basic matrix definitions

A matrix is represented as a two-dimensional array of elements, a_{ij} , where i is the row index and j is the column index. The entire matrix is represented by the single boldface symbol \mathbf{A} . In general we speak of a matrix as having n rows and m columns. Such a matrix is called an $(n$ by $m)$ or $(n \times m)$ matrix. Equation [1] shows the representation of a typical $(n \times m)$ matrix.

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & \cdots & a_{1m} \\ a_{21} & a_{22} & a_{23} & \cdots & \cdots & a_{2m} \\ a_{31} & a_{32} & a_{33} & \cdots & \cdots & a_{3m} \\ \vdots & \vdots & \vdots & \ddots & & \vdots \\ \vdots & \vdots & \vdots & & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & \cdots & a_{nm} \end{bmatrix} \quad [1]$$

In general the number of rows may be different from the number of columns. Sometimes the matrix is written as $\mathbf{A}_{(n \times m)}$ to show its size. (Size is defined as the number of rows and the number of columns.) A matrix that has the number of rows equal to the number of columns is called a **square matrix**.

Matrices are used to represent physical quantities that have more than one number. These are usually used for engineering systems such as structures or networks in which we represent a collection of numbers, such as the individual stiffness of the members of a structure, as a single symbol known as a stiffness matrix. Networks of pipes, circuits, traffic streets, and the like may be represented by a connectivity matrix which indicates which pair of nodes in the matrix are directly joined to each other. The use of matrix notation and formulae for matrices leads to important analytical results. Students taking a vibrations course learn that a matrix property

¹ Kreyszig, *Advanced Engineering Mathematics* (9th edition), Wiley, 2006, Chapter 7.

knows as its eigenvalues represents the fundamental vibration frequencies in a mechanical system.

Two matrices can be added or subtracted if both matrices have the same size. If we define a matrix, \mathbf{C} , as the sum (or difference) of two matrices, \mathbf{A} and \mathbf{B} , we can write this sum (or difference) in terms of the matrices as follows.

$$\mathbf{C} = \mathbf{A} \pm \mathbf{B} \quad (\text{possible only if } \mathbf{A} \text{ and } \mathbf{B} \text{ have the same size}) \quad [2]$$

The components of the \mathbf{C} matrix are simply the sum (or difference) of the components of the two matrices being added (or subtracted). Thus for the matrix sum (or difference) shown in equation [2], the components of \mathbf{C} are give by the following equation.

$$\mathbf{C} = \mathbf{A} \pm \mathbf{B} \quad \Rightarrow \quad c_{ij} = a_{ij} \pm b_{ij} \quad (i = 1, n; j = 1, m) \quad [3]$$

The product of a matrix, \mathbf{A} , with a single number, x , yields a second matrix whose size is the same as that of matrix \mathbf{A} . Each component of the new matrix is the component of the original matrix, a_{ij} , multiplied by the number x . The number x in this case is usually called a scalar to distinguish it from a matrix or a matrix component.

$$\mathbf{B} = x\mathbf{A} \quad \text{if } b_{ij} = xa_{ij} \quad (i = 1, n; j = 1, m) \quad [4]$$

We define two special matrices, the null matrix, $\mathbf{0}$, and the identity matrix, \mathbf{I} . The null matrix is an arbitrary size matrix in which all the elements are zero. The identity matrix is a square matrix in which all the diagonal terms are 1 and the off-diagonal terms are zero. These matrices are sometimes written as $\mathbf{0}_{(m \times n)}$ or \mathbf{I}_n to specify a particular size for the null or identity matrix. The null matrix and the identity matrix are shown below.

$$\mathbf{0} = \begin{bmatrix} 0 & 0 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & 0 & & & 0 \\ \vdots & \vdots & \vdots & \ddots & & \vdots \\ \vdots & \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \cdots & 0 \end{bmatrix} \quad \mathbf{I} = \begin{bmatrix} 1 & 0 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & 1 & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & & \vdots \\ \vdots & \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \cdots & 1 \end{bmatrix} \quad [5]$$

A matrix that has the same pattern as the identity matrix, but has terms other than ones on its principal diagonal is called a **diagonal matrix**. The general term for such a matrix is $d_i \delta_{ij}$, where d_i is the diagonal term for row i and δ_{ij} is the Kronecker delta; the latter is defined such that $\delta_{ij} = 0$ unless $i = j$, in which case $\delta_{ij} = 1$. A diagonal matrix is sometimes represented in the following form: $\mathbf{D} = \mathbf{diag}(d_1, d_2, d_3, \dots, d_n)$; this says that \mathbf{D} is a diagonal matrix whose diagonal components are given by d_i

We call the diagonal for which the row index is the same as the column index, the main or principal diagonal. Algorithms in the numerical analysis of differential equations lead to matrices whose nonzero terms lie along diagonals. For such a matrix, all the nonzero terms may be represented by symbols like $a_{i,i-k}$ or $a_{i,i+k}$. Diagonals with subscripts $a_{i,i-k}$ or $a_{i,i+k}$ are said to lie, respectively, below or above the main diagonal.

If the n rows and m columns in a matrix, \mathbf{A} , are interchanged, we will have a new matrix, \mathbf{B} , with m rows and n columns. The matrix \mathbf{B} is said to be the **transpose** of \mathbf{A} , written as \mathbf{A}^T .

$$\mathbf{B} = \mathbf{A}^T \quad \text{if } b_{ij} = a_{ji} \quad [i = 1, n; j = 1, m; \mathbf{A} \text{ is } (n \times m); \mathbf{B} \text{ is } (m \times n).] \quad [6]$$

An example of an original \mathbf{A} matrix and its transpose is shown below.

$$\mathbf{A} = \begin{bmatrix} 3 & 12 & -6 \\ 14 & -2 & 0 \end{bmatrix} \quad \mathbf{A}^T = \begin{bmatrix} 3 & 14 \\ 12 & -2 \\ -6 & 0 \end{bmatrix} \quad [7]$$

The transpose of a product of matrices equals the product of the transposes of individual matrices, with the order reversed. That is,

$$(\mathbf{AB})^T = \mathbf{B}^T \mathbf{A}^T \quad (\mathbf{ABC})^T = \mathbf{C}^T \mathbf{B}^T \mathbf{A}^T \quad (\mathbf{ABCD})^T = \dots \quad [8]$$

Matrices with only one row are called row matrices; matrices with only one column are called column matrices.² Although we can write the elements of such matrices with two subscripts, the subscript of one for the single row or the single column is usually not included. The examples below for the row matrix, \mathbf{r} , and the column matrix, \mathbf{c} , show two possible forms for the subscripts. In each case, the second matrix has the commonly used notation. When row and column matrices are used in formulas that have two matrix subscripts, the first form of the matrices shown below are implicitly used to give the second subscript for the equation.

$$\mathbf{r} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & \cdots & \cdots & r_{1m} \\ r_1 & r_2 & r_3 & \cdots & \cdots & r_m \end{bmatrix} \quad \mathbf{c} = \begin{bmatrix} c_{11} \\ c_{21} \\ c_{31} \\ \vdots \\ \vdots \\ c_{n1} \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ \vdots \\ \vdots \\ c_n \end{bmatrix} \quad [9]$$

The transpose of a column matrix is a row matrix; the transpose of a row matrix is a column matrix. This is sometimes used to write a column matrix in the middle of text by saying, for example, that $\mathbf{c} = [1 \ 3 \ -4 \ 5]^T$.

Matrix Multiplication

The definition of matrix multiplication seems unusual when encountered for the first time. However, it has its origins in the treatment of linear equations. For a simple example, we consider three two-dimensional coordinate systems. The coordinates in the first system are x_1

² Row and column matrices are called row vectors or column vectors when they are used to represent the components of a vector. In these notes we will use upper case boldface letters such as \mathbf{A} and \mathbf{B} to represent matrices with more than one row or more than one column; we will use lower case boldface letters such as \mathbf{a} or \mathbf{b} to represent matrices with only one row or only one column. We will generally refer to these matrices as vectors.

and x_2 . The coordinates for the second system are y_1 and y_2 . The third system has coordinates z_1 and z_2 . Each coordinate system is related by a coordinate transformation given by the following relations.

$$\begin{aligned} y_1 &= a_{11}x_1 + a_{12}x_2 & z_1 &= b_{11}y_1 + b_{12}y_2 \\ y_2 &= a_{21}x_1 + a_{22}x_2 & z_2 &= b_{21}y_1 + b_{22}y_2 \end{aligned} \quad [10]$$

We can obtain a relationship between the z coordinate system and the x coordinate system by combining the various components of equation [10] to eliminate the y_i coordinates as follows.

$$\begin{aligned} z_1 &= b_{11}[a_{11}x_1 + a_{12}x_2] + b_{12}[a_{21}x_1 + a_{22}x_2] \\ z_2 &= b_{21}[a_{11}x_1 + a_{12}x_2] + b_{22}[a_{21}x_1 + a_{22}x_2] \end{aligned} \quad [11]$$

We can rearrange these terms to obtain a set of equations similar to those in equation [10] that relates the z coordinate system to the x coordinate system.

$$\begin{aligned} z_1 &= [b_{11}a_{11} + b_{12}a_{21}]x_1 + [b_{11}a_{12} + b_{12}a_{22}]x_2 = c_{11}x_1 + c_{12}x_2 \\ z_2 &= [b_{21}a_{11} + b_{22}a_{21}]x_1 + [b_{21}a_{12} + b_{22}a_{22}]x_2 = c_{21}x_1 + c_{22}x_2 \end{aligned} \quad [12]$$

We see that the coefficients c_{ij} , for the new transformation are related to the coefficients for the previous transformations as follows.

$$\begin{aligned} c_{11} &= [b_{11}a_{11} + b_{12}a_{21}] & c_{12} &= [b_{11}a_{12} + b_{12}a_{22}] \\ c_{21} &= [b_{21}a_{11} + b_{22}a_{21}] & c_{22} &= [b_{21}a_{12} + b_{22}a_{22}] \end{aligned} \quad [13]$$

There is a general form for each c_{ij} coefficient in equation [13]. Each is a sum of products of two terms. The first term from each product is a b_{ik} value whose first subscript (i) is the same as the first subscript of the c_{ij} coefficient being computed. The second term in each product is an a_{kj} value whose second subscript (j) is the same as the second subscript of the c term being computed. In each $b_{ik}a_{kj}$ product, the second b subscript (k) is the same as the first a subscript. From these observations we can write a general equation for each of the four coefficients in equation [13] as follows.

$$c_{ij} = \sum_{k=1}^2 b_{ik}a_{kj} \quad (i = 1,2; j = 1,2) \quad [14]$$

The definition of matrix multiplication is a generalization of the simple example in equation [14] to any general sizes of matrices. In this general case, we define the product, $\mathbf{C} = \mathbf{AB}$, of two matrices, \mathbf{A} with n rows and p columns, and \mathbf{B} with p rows and m columns by the following equation.

$$\mathbf{C}_{(n \times m)} = \mathbf{A}_{(n \times p)} \mathbf{B}_{(p \times m)} \Rightarrow c_{ij} = \sum_{k=1}^p a_{ik}b_{kj} \quad (i = 1, \dots, n; j = 1, \dots, m) \quad [15]$$

There are two important items to consider in the formula for matrix multiplication. The first is that order is important. The product \mathbf{AB} is different from the product \mathbf{BA} . In fact, one of the products

may not be possible. The second item is the need for compatibility between the first and second matrix in the **AB** product.³ In order to obtain the product **AB** the number of columns in **A** must equal the number of rows in **B**. A simple example of matrix multiplication is shown below.

$$\mathbf{A} = \begin{bmatrix} 3 & 0 & -6 \\ 4 & -2 & 0 \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} 3 & 4 \\ 1 & 2 \\ 6 & 1 \end{bmatrix} \quad [16]$$

$$\mathbf{AB} = \begin{bmatrix} 3(3) + 0(1) - 6(6) & 3(4) + 0(2) - 6(1) \\ 4(3) - 2(1) + 0(6) & 4(4) - 2(2) + 0(1) \end{bmatrix} = \begin{bmatrix} -27 & 6 \\ 10 & 12 \end{bmatrix}$$

Matrix multiplication is simple to program. The VBA code for multiplying two matrices is shown below.⁴ This code assumes that all variables have been properly declared and initialized. The code uses the obvious notation to implement equation [15]. The array components are denoted as a[i][k], b[k][j] and c[i][j]. The product matrix, **C**, has the same number of rows, n, as in matrix **A** and the same number of columns, m, as in matrix **B**. The number of columns in **A** is equal to p, which must also equal the number of rows in **B**.

```

For i = 1 To n
  For j = 1 To m
    c(i,j) = 0.0
    For k = 1 To p
      c(i,j) = c(i,j) + a(i,k) * b(k,j)
    Next k
  Next j
Next i

```

We now examine how the coordinate transformations that we used above to introduce matrix multiplication can be represented as matrix equations. We can define matrices, **A**, **B**, and **C** to represent the coefficients that we used in our coordinate transformation equations.

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \quad \mathbf{C} = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} \quad [17]$$

The various coordinate pairs can be represented as column matrices as shown below.

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_{11} \\ x_{21} \end{bmatrix} \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} y_{11} \\ y_{21} \end{bmatrix} \quad \mathbf{z} = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} z_{11} \\ z_{21} \end{bmatrix} \quad [18]$$

With these matrix definitions, the two sets of simultaneous linear equations shown in equation [10] can be represented by the following pair of matrix equations:

³ The terms **premultiply** and **postmultiply** are commonly used to indicate the order of the matrices involved in matrix multiplication. In the matrix product **AB**, we say that **B** is **premultiplied** by **A** or that **A** is **postmultiplied** by **B**. Alternatively, the terms **left multiplied** and **right multiplied** are used. In the **AB** product, **A** is right multiplied by **B** and **B** is left multiplied by **A**.

⁴ The basic code structure is the same in any language. There are three nested loops. The two outer loops cover all possible combinations of i and j to ensure that all the c_{ij} components are computed. The inner loop code is the typical code for summing a number of items.

$$\mathbf{y} = \mathbf{A}\mathbf{x} \quad \text{and} \quad \mathbf{z} = \mathbf{B}\mathbf{y} \quad [19]$$

You can verify that the equations above are correct by applying the general formula for matrix multiplication in equation [15] to the matrix equations in [19]. To do this, you should use the definitions of \mathbf{A} , \mathbf{B} , \mathbf{x} , \mathbf{y} , and \mathbf{z} , provided in equations [17] and [18]. If we combine the matrix equations in [19] to eliminate the \mathbf{y} matrix, we get the following result.

$$\mathbf{z} = \mathbf{B}\mathbf{y} = \mathbf{B}\mathbf{A}\mathbf{x} \quad \text{or} \quad \mathbf{z} = \mathbf{C}\mathbf{x} \quad \text{with} \quad \mathbf{C} = \mathbf{B}\mathbf{A} \quad [20]$$

Note the importance of the order of multiplication. In general, $\mathbf{B}\mathbf{A}$; it is not equal to $\mathbf{A}\mathbf{B}$.

There are two cases where the order is not important. These are multiplication by a null matrix, which produces a null matrix, and multiplication by an identity matrix, which produces the original matrix.

$$\mathbf{0}\mathbf{A} = \mathbf{A}\mathbf{0} = \mathbf{0} \quad \text{and} \quad \mathbf{A}\mathbf{I} = \mathbf{I}\mathbf{A} = \mathbf{A} \quad [21]$$

Although the order is not important here, the actual identity and null matrices used may be different. We can rewrite equations [21] to explicitly show the rows and columns in each matrix.

$$\begin{aligned} \mathbf{0}_{(p \times n)} \mathbf{A}_{(n \times m)} &= \mathbf{0}_{(p \times m)} & \mathbf{A}_{(n \times m)} \mathbf{0}_{(m \times q)} &= \mathbf{0}_{(n \times q)} \\ \mathbf{A}_{(n \times m)} \mathbf{I}_{(m \times m)} &= \mathbf{I}_{(n \times n)} \mathbf{A}_{(n \times m)} & &= \mathbf{A}_{(n \times m)} \end{aligned} \quad [22]$$

By definition the identity matrix is a square matrix. One size specification for the identity matrix, the number of rows or the number of columns, is set by the compatibility condition for matrix multiplication. Once this is done, the other size is set by the requirement that \mathbf{I} is square. For the null matrices in equation [22], the size specifications, n or m , must match the sizes for the \mathbf{A} matrix. Although the size specifications p and q , for the null matrices in equation [22] are arbitrary, they are usually taken as $p = m$ and $q = n$ to give a square null matrix as the $\mathbf{0}\mathbf{A}$ product.

Simultaneous Linear Algebraic Equations

The coordinate transformation equations are simple examples of a more general case for simultaneous linear algebraic equations. In the general case we can have a set of simultaneous equations that is written as follows

$$\sum_{j=1}^m a_{ij} x_j = b_i \quad i = 1, \dots, n \quad [23]$$

We expect that a well-determined problem will have the number of equations, n , equal to the number of unknowns, m , but in the general case, shown above, these n may be different from m . We see that this system of equations in equation [23] can be represented by matrices $\mathbf{A}_{(n \times m)}$, $\mathbf{x}_{(m \times 1)}$, and $\mathbf{b}_{(n \times 1)}$, where \mathbf{A} has the form shown in equation [1], \mathbf{x} is the column matrix $[x_1, x_2, x_3, \dots, x_m]^T$ and \mathbf{b} is the column matrix $[b_1, b_2, b_3, \dots, b_n]^T$. With these definitions, equation [23] is the same as the general equation for matrix multiplication shown in equation [15]. (Recall that we have omitted the second subscript, which is one, on the components of \mathbf{x} and \mathbf{b} .) The system of equations shown in equation [23] is written, in matrix form, in equation [24], below.

$$\mathbf{A}\mathbf{x} = \mathbf{b} \quad [24]$$

These matrices are written out in detail below. Here the column matrix, \mathbf{x} , appears to have more rows than the coefficient matrix, \mathbf{A} . This is done to emphasize the notion that m may be different from n in general. Of course, m may be equal to or less than n rather than greater than n as implied in the matrices below.

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & \cdots & a_{1m} \\ a_{21} & a_{22} & a_{23} & \cdots & \cdots & a_{2m} \\ a_{31} & a_{32} & a_{33} & \cdots & \cdots & a_{3m} \\ \vdots & \vdots & \vdots & \ddots & & \vdots \\ \vdots & \vdots & \vdots & & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & \cdots & a_{nm} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ \vdots \\ \vdots \\ x_m \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ \vdots \\ b_n \end{bmatrix} \quad [25]$$

In order to solve a set of simultaneous linear equations we use a process that replaces equations in the set by equivalent equations. We can replace any equation in the set by a linear combination of other equations without changing the solution of the system of equations. For example, consider the simple set of two equations with two unknowns

$$\begin{aligned} 3x_1 + 5x_2 &= 13 \\ 7x_1 - 4x_2 &= -1 \end{aligned} \quad [26]$$

You can confirm that $x_1 = 1$ and $x_2 = 2$ is a solution to this set of equations. To solve this set of equations we can replace the second equation by a new equation, which is a linear combination of the two equations without changing the solution. The particular combination we seek is one that will eliminate x_1 . We can do this by subtracting the first equation, multiplied by $7/3$, from the second equation to obtain the following pair of equations, which is equivalent to the original set in equation [26].

$$\begin{aligned} 3x_1 + 5x_2 &= 13 \\ -\frac{47}{3}x_2 &= -\frac{94}{3} \end{aligned} \quad [27]$$

We can readily solve the second equation to find $x_2 = 2$, and substitute this value of x_2 into the first equation to find that $x_1 = [13 - 5(2)]/3 = 1$. The general process for solving the system of equations represented by equations [23], [24], or [25], known as Gauss elimination, is similar to the one just shown. It requires a series of operations on the coefficients a_{ij} and b_i to produce a set of equations with the form shown in equation [28], below, without changing the solution of the initial problem.

The basic rule in the Gauss elimination process is that we can use a linear combination of two equations to replace one of those equations, without changing the solution to the problem. This is the process that we used above in going from the set of equations in [26] to the set of equations in [27]. Both sets of equations are equivalent in the sense that both sets of equations give the same answers for x_1 and x_2 . However, the second set of equations can be directly solved for all the unknowns.

$$\begin{bmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} & \cdots & \cdots & \alpha_{1n-1} & \alpha_{1n} \\ 0 & \alpha_{22} & \alpha_{23} & \cdots & \cdots & \alpha_{2n-1} & \alpha_{2n} \\ 0 & 0 & \alpha_{33} & \cdots & \cdots & \alpha_{3n-1} & \alpha_{3n} \\ \vdots & \vdots & \vdots & \ddots & & & \\ \vdots & \vdots & \vdots & & \ddots & & \\ 0 & 0 & 0 & \cdots & \cdots & \alpha_{n-1n-1} & \alpha_{n-1n} \\ 0 & 0 & 0 & \cdots & \cdots & 0 & \alpha_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ \vdots \\ x_{n-1} \\ x_n \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \vdots \\ \vdots \\ \beta_{n-1} \\ \beta_n \end{bmatrix} \quad [28]$$

The revised coefficient matrix in equation [28] is called an upper triangular matrix. The only nonzero terms are on or above the principal diagonal. The same operations that are used to obtain the revised coefficient matrix are used to obtain the revised right-hand-side matrix.

The revised **A** and **b** matrices are obtained in a series of steps. In the first step, the x_1 coefficients are eliminated from all equations except the first one. This is done by the following replacement operations on the coefficients in equations 2 to n . The replacement notation (\leftarrow) from computer programming is used here to indicate that an old value of a_{ij} is being replaced by the results of a calculation. This avoids the need to use mathematical notation that would require separate symbols for the two values.

$$\left(a_{ij} \leftarrow a_{ij} - \frac{a_{i1}}{a_{11}} a_{1j} \quad j=1, \dots, n \quad \text{and} \quad b_i \leftarrow b_i - \frac{a_{i1}}{a_{11}} b_1 \right) \quad i=2, \dots, n \quad [29]$$

After equation [29] is applied to all rows below the first row, the only nonzero x_1 coefficient is in the first equation (represented by the first row of the matrix.) You can confirm that this will set $a_{i1} = 0$ for $i > 1$. You can also apply the formulae in [29] to equation [26] to see that the result is equation [27]. The elimination process is next applied to make the x_2 coefficients on all equations below the second equation zero.

$$\left(a_{ij} \leftarrow a_{ij} - \frac{a_{i2}}{a_{22}} a_{2j} \quad j=2, \dots, n \quad \text{and} \quad b_i \leftarrow b_i - \frac{a_{i2}}{a_{22}} b_2 \right) \quad i=3, \dots, n \quad [30]$$

Equation [30] has the same form as equation [29]; only the starting points for the row and column operations are different. The process described by equations [29] and [30] continues until the form shown in equation [28] is obtained. From equation [28], the various values of x can be found by back substitution. We can simply find x_n as β_n/α_{nn} . The remaining values of x are found in reverse order by the following equation.

$$x_i = \frac{\beta_i - \sum_{j=i+1}^n \alpha_{ij} x_j}{\alpha_{ii}} \quad i = n-1, n-2, \dots, 1 \quad [31]$$

When we are solving for x_i , all previous values of x_j required in the summation are known.

The VBA code below shows a simplified version⁵ of how the Gauss elimination method is applied to the solution of equations. As in previous code examples, all data values are assumed to be properly declared and initialized. The number of equations is equal to the number of unknowns, n . The row that is subtracted from all rows below it is called the pivot row. The main outer loop in the first part of the code uses the variable, `pivot`, to represent this row. The code execution is simplified by augmenting the \mathbf{a} matrix so that $a_{i,n+1} = b_i$. This allows the code to proceed without separate consideration of similar operations on the \mathbf{A} and \mathbf{b} matrix components. All operations are performed on the original \mathbf{A} matrix so that the original data are lost in this routine.

```

    'augment a matrix with b values
For row = 1 to n
    a(row,n+1) = b(row)
Next row

    'get upper triangular array
For pivot = 1 to n
    For row = pivot+1 to n; row++ )
        'Code here to find row with larger pivot element
        For ( column = row+1 to n+1
            a(row,column) = a(row,column) - a(row,pivot)
                * a(pivot,column) / a(pivot,pivot)
        Next column
    Next row
Next pivot

    'Upper triangular matrix complete; get x values
For row = n To 1 Step -1
    x(row) = a(row,n+1);
    For column = n To row-1
        x(row) = x(row) - a(row,column) * x(column)
    Next column
    x(row) /= x(row) / a(row,row)
Next row

```

The process outlined above for the solution of a set of simultaneous equations is known as the Gaussian elimination procedure. Alternative procedures such as the Gauss-Jordan method and LU decomposition, work in a similar manner. They produce an upper triangular matrix or diagonal matrix that is then used to solve for the values of x_i in reverse order.

Matrix Rank Determines Existence and Uniqueness of Solutions

If the solution process outlined above is used on certain matrices, it may not be possible to obtain a solution. Consider the two sets of equations shown below.

$$\begin{array}{rcl}
 3x_1 + 5x_2 = 13 & & 3x_1 + 5x_2 = 13 \\
 6x_1 + 10x_2 = 26 & \text{and} & 6x_1 + 10x_2 = 27
 \end{array} \quad [32]$$

⁵ Actual code would have to account for the possibility that the system of equations might not have a solution. Typically the code switches rows of the augmented matrix (equivalent to exchanging the order of the equations so that no change is made in the solution). Before a row is considered as a pivot row, the value in the pivot column in each equation below the pivot row is examined. The row with the largest value in the pivot column (in absolute value) is exchanged with the original pivot row before subtracting the pivot row from other rows. If the pivot column does not have a nonzero element (nonzero to within numerical accuracy), the problem has a singular matrix.

In the set of equations on the left, the second equation is simply twice the first equation. If we multiply the first equation by two and subtract it from the second equation, we get the result that $0 = 0$. Thus, the second equation gives us no new information on the relationship between x_1 and x_2 . We say that this system of equations has an infinite number of solutions. Any value of $x_2 = 2.6 - 0.6x_1$ will satisfy both equations. The second set of equations has no solutions. If we multiply the first equation by two and subtract it from the second equation, we have the result that $0 = 1$! Thus, this second set of equations is incompatible and does not have a solution.⁶

This simple example can be generalized to discuss the existence and uniqueness of solutions for the general set of equations. If we carry out the solution process outlined above to form an upper triangular matrix, we may have the result that one (or more) of the final rows in the coefficient matrix is all zero. This means that we cannot obtain a unique solution. Such a case is called a singular matrix. The **rank** of a matrix is formally defined as the number of linearly independent rows in a matrix. (This can be shown to be equal to the number of linearly independent columns.) The practical determination of rank is based on the Gauss elimination process outlined above. If in the final matrix in the elimination process is a matrix with n rows of which n_{zero} rows contain all zeros, the rank of the matrix is $n - n_{\text{zero}}$. (This rank is the same for both the original matrix and the upper-triangular matrix because the Gauss elimination operations do not change the matrix rank.) The **A** matrix for both sets of equations in equation [32] has only one linearly independent row, thus its rank is one. The upper triangular form that results when a matrix is tested for rank is sometimes called the **row-echelon** form. Sometimes in this form each row is divided by the diagonal element on that row so that all the diagonal elements are one.

The two matrices in equation [33] below have been placed in row-echelon form by using Gauss elimination on the original matrices. Can you determine the rank of the original matrices before looking at the answers below?

The matrix on the left of equation [33] has four rows that are not all zero; thus its rank is four. The one on the right has six rows that are not all zero, thus its rank is six. This rank-six matrix has eight columns. Because the number of linearly independent columns and the number of linearly independent rows is the same as the rank of six, we know that these eight columns will be related by two different linearly equations.

$$\begin{bmatrix} 6 & 0 & 2 & 0 & 0 & 0 \\ 0 & 1 & 7 & 8 & 6 & 2 \\ 0 & 0 & 3 & 4 & 0 & 0 \\ 0 & 0 & 0 & 9 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad \begin{bmatrix} 6 & 0 & 2 & 0 & 0 & 0 & 1 & 6 \\ 0 & 1 & 7 & 8 & 6 & 2 & 8 & 4 \\ 0 & 0 & 2 & 0 & 3 & 5 & 8 & 0 \\ 0 & 0 & 0 & 4 & 1 & 0 & 7 & 3 \\ 0 & 0 & 0 & 0 & 6 & 0 & 3 & 5 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 & 1 \end{bmatrix} \quad [33]$$

The existence and uniqueness of solutions are defined in terms of the rank of the augmented matrix, **[A,b]**. This is the matrix in which the right hand side column matrix, **b**, is added as the final column in the **A** matrix. This augmented matrix is shown below for the general case of n equations and m unknowns. The n equations mean that there are n rows in the matrix. The m unknowns give $m + 1$ columns to the augmented matrix.

⁶ This result has a geometric interpretation. When we have two simultaneous linear algebraic equations, we can plot each equation in x_1 - x_2 space. The solution to the pair of equations is located at the point where both equations intersect. If we did this for the left set of equations in [32], we would only have a single line.

$$[\mathbf{A}, \mathbf{b}] = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & \cdots & a_{1m} & b_1 \\ a_{21} & a_{22} & a_{23} & \cdots & \cdots & a_{2m} & b_2 \\ a_{31} & a_{32} & a_{33} & \cdots & \cdots & a_{3m} & b_3 \\ \vdots & \vdots & \vdots & \ddots & & \vdots & \vdots \\ \vdots & \vdots & \vdots & & \ddots & \vdots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & \cdots & a_{nm} & b_n \end{bmatrix} \quad [34]$$

The existence and uniqueness of solutions to $\mathbf{Ax} = \mathbf{b}$ is stated below without proof.

If the rank of the original matrix, \mathbf{A} , equals the rank of the augmented matrix, $[\mathbf{A}, \mathbf{b}]$, and both equal the number of unknowns, m , there is a unique solution to the matrix equation, $\mathbf{Ax} = \mathbf{b}$.

If the rank of the original matrix, \mathbf{A} , equals the rank of the augmented matrix, $[\mathbf{A}, \mathbf{b}]$, but the common rank is less than the number of unknowns, m , there are an infinite number of solutions to the matrix equation, $\mathbf{Ax} = \mathbf{b}$.

If the rank of the original matrix, \mathbf{A} , is not equal the rank of the augmented matrix, $[\mathbf{A}, \mathbf{b}]$, there is no solution to the matrix equation, $\mathbf{Ax} = \mathbf{b}$.

We can see that these statements are consistent with the examples in equation [32]. A formal proof of these statements is given in linear algebra texts.

These guidelines for the existence and uniqueness of solutions to simultaneous linear equations are illustrated in the three sets of equations shown below. Each equation set has three equations in three unknowns. The original equation set, shown in the first column, is converted to an upper triangular form in the second column. We see that the first set has a unique solution. The second and third sets do not have a unique solution; however, there is a difference between these two. The second set has an infinite number of solutions. For any value, α that we pick for x_3 we can determine a value of x_1 and x_2 that is consistent with the original set of equations. However, for the third set of equations, the upper triangular form gives an inconsistent third equation. Thus this set of equations has no solution.

	Original Equation Set	Upper Triangular Form	Solutions
Set I	$x_1 - 4x_2 - 26x_3 = 2$ $2x_2 + 9x_3 = -5$ $7x_1 + 3x_2 + 8x_3 = -13$	$x_1 - 4x_2 - 26x_3 = 2$ $2x_2 + 9x_3 = -5$ $50.5x_3 = 50.5$	$x_1 = 0$ $x_2 = -7$ $x_3 = 1$
Set II	$x_1 - 4x_2 - 26x_3 = 2$ $2x_2 + 9x_3 = -5$ $-2x_1 + 10x_2 + 61x_3 = -9$	$x_1 - 4x_2 - 26x_3 = 2$ $2x_2 + 9x_3 = -5$ $0 = 0$	$x_1 = 12 - 8\alpha$ $x_2 = -2.5 - 4.5\alpha$ $x_3 = \alpha$
Set III	$x_1 - 4x_2 - 26x_3 = 2$ $2x_2 + 9x_3 = -5$ $-2x_1 + 10x_2 + 61x_3 = -8$	$x_1 - 4x_2 - 26x_3 = 2$ $2x_2 + 9x_3 = -5$ $0 = 1$	No Solution

These three sets of equations are shown in terms of their \mathbf{A} and augmented $[\mathbf{A} \ \mathbf{b}]$ matrices in the table below. We see that the set of equations in the table above corresponds to the data in the augmented matrix. The first set of equations has rank $\mathbf{A} = \text{rank} [\mathbf{A} \ \mathbf{b}] = 3$, the number of

unknowns. We have already seen that this provides the unique solution above. The second set of equations has rank $\mathbf{A} = \text{rank} [\mathbf{A} \ \mathbf{b}] = 2$, less than the number of unknowns. This means that we have an infinite number of solutions. Again, this corresponds to the result above. Finally, the third case below has rank $\mathbf{A} = 2$, but rank $[\mathbf{A} \ \mathbf{b}] = 3$. This difference in rank shows that there are no solutions.

	Original Matrices		Row-Echelon Form		Rank	
	\mathbf{A}	$[\mathbf{A} \ \mathbf{b}]$	\mathbf{A}	$[\mathbf{A} \ \mathbf{b}]$	\mathbf{A}	$[\mathbf{A} \ \mathbf{b}]$
Set I	$\begin{bmatrix} 1 & -4 & 2 \\ 0 & 2 & 9 \\ 7 & 3 & 8 \end{bmatrix}$	$\begin{bmatrix} 1 & -4 & 2 & 2 \\ 0 & 2 & 9 & -5 \\ 7 & 3 & 8 & -13 \end{bmatrix}$	$\begin{bmatrix} 1 & -4 & 2 \\ 0 & 2 & 9 \\ 0 & 0 & 50.5 \end{bmatrix}$	$\begin{bmatrix} 1 & -4 & 2 & 2 \\ 0 & 2 & 9 & -5 \\ 0 & 0 & 50.5 & 50.5 \end{bmatrix}$	3	3
Set II	$\begin{bmatrix} 1 & -4 & 2 \\ 0 & 2 & 9 \\ -2 & 10 & 61 \end{bmatrix}$	$\begin{bmatrix} 1 & -4 & 2 & 2 \\ 0 & 2 & 9 & -5 \\ -2 & 10 & 61 & -9 \end{bmatrix}$	$\begin{bmatrix} 1 & -4 & 2 \\ 0 & 2 & 9 \\ 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 1 & -4 & 2 & 2 \\ 0 & 2 & 9 & -5 \\ 0 & 0 & 0 & 0 \end{bmatrix}$	2	2
Set III	$\begin{bmatrix} 1 & -4 & 2 \\ 0 & 2 & 9 \\ -2 & 10 & 61 \end{bmatrix}$	$\begin{bmatrix} 1 & -4 & 2 & 2 \\ 0 & 2 & 9 & -5 \\ -2 & 10 & 61 & -8 \end{bmatrix}$	$\begin{bmatrix} 1 & -4 & 2 \\ 0 & 2 & 9 \\ 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 1 & -4 & 2 & 2 \\ 0 & 2 & 9 & -5 \\ 0 & 0 & 0 & 1 \end{bmatrix}$	2	3

There is one final case to consider; that is the case of **homogenous equations**, where the \mathbf{b} matrix is all zeros. If there are n equations and the rank of the coefficient matrix is n then the only solution to the set of equations is that all $x_i = 0$. (This is called the trivial solution.) However, if the rank is less than n , it is possible to have a solution in which all the x_i are not zero. However, such a solution is not unique.

Consider the two sets of homogenous equations shown below. Each set of equations has a right-hand side that is all zeros. (The two equation sets are identical except for the coefficient of the x_1 term in the first equation.)

$$\begin{array}{lcl}
 -x_1 - 4x_2 + 3x_3 = 0 & & x_1 - 4x_2 + 3x_3 = 0 \\
 -4x_1 + 11x_2 - 6x_3 = 0 & \text{and} & -4x_1 + 11x_2 - 6x_3 = 0 \\
 x_1 - 8x_2 + 5x_3 = 0 & & x_1 - 8x_2 + 5x_3 = 0
 \end{array} \quad [35]$$

If we carry out the usual solution process to create an upper triangular matrix for these two sets of equations, we obtain the following results.

$$\begin{array}{lcl}
 -x_1 - 4x_2 + 3x_3 = 0 & & x_1 - 4x_2 + 3x_3 = 0 \\
 27x_2 - 18x_3 = 0 & \text{and} & -6x_2 + 6x_3 = 0 \\
 0 = 0 & & -2.8x_3 = 0
 \end{array} \quad [36]$$

For the set of equations on the right, the rank of the coefficient matrix is the same as the number of equations. Here we have a unique solution in which all of the $x_i = 0$. The rank of the coefficient matrix for the equations on the left is less than the number of equations. In this case, we have an infinite number of solutions. If we pick $x_3 = \alpha$, an arbitrary constant, we can satisfy all three

equations if $x_2 = 2\alpha/3$ and $x_1 = 3\alpha - 4(2\alpha/3) = \alpha/3$. One of the infinite solutions, with $\alpha = 0$, is the trivial solution where all $x_i = 0$.⁷

Determinants

A determinant is a single numerical value that can be computed for a square array. The values of determinants play a theoretical role in matrix analysis and can be used for calculations on small matrices. For matrices greater than 3×3 or 4×4 , alternative calculation methods are used in place of determinants.

Various notations are available for a determinant. If \mathbf{A} is a matrix, then $\text{Det } \mathbf{A}$ is the determinant for the coefficients in a matrix. The determinant for an array of numbers that looks like a matrix can be written using the absolute value sign, $|\langle \text{array} \rangle|$, instead of the brackets, $\langle \text{array} \rangle$, that we have been using for matrix coefficients. The various notations are shown below for a 2×2 array. For this array, the formula for the determinant is particularly simple.

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \quad \text{Det } \mathbf{A} = \text{Det} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{21}a_{12} \quad [37]$$

For a 3×3 array, the determinant is a bit more complex.

$$\text{Det} \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}a_{22}a_{33} + a_{21}a_{32}a_{13} + a_{31}a_{12}a_{23} - a_{11}a_{32}a_{23} - a_{21}a_{12}a_{33} - a_{31}a_{22}a_{13} \quad [38]$$

The general equation for computing a determinant is given in terms of minors (or cofactors) of a determinant. The minor, M_{ij} of a determinant is the smaller determinant that results if row i and column j are eliminated from the original determinant. The cofactor, A_{ij} , equals $(-1)^{i+j}M_{ij}$. For example, if we start with a 3×3 determinant, such as the one shown in equation [38] we can define nine possible minors (and cofactors). Four of these are shown below:

$$\begin{aligned} A_{33} &= (-1)^{3+3} M_{33} = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} & A_{22} &= (-1)^{2+2} M_{22} = \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} \\ A_{32} &= (-1)^{3+2} M_{32} = - \begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix} & A_{31} &= (-1)^{3+1} M_{31} = \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} \end{aligned} \quad [39]$$

The determinant of a matrix can be written in terms of its minors or cofactors as follows.

⁷ You should make sure that you can place the original sets of equations in [35] and the upper triangular forms in [36] into an \mathbf{A} and an augmented $[\mathbf{A} \ \mathbf{b}]$ matrix and show that both sets of equations have $\text{rank } \mathbf{A} = \text{rank } [\mathbf{A} \ \mathbf{b}]$. Do both sets of equations produce \mathbf{A} matrices with the same rank? What are the ranks of the \mathbf{A} and $[\mathbf{A} \ \mathbf{b}]$ matrices for the two sets of equations?

$$\begin{aligned}
 \text{Det } \mathbf{A}_{(n \times n)} &= \sum_{i=1}^n (-1)^{i+j} a_{ij} M_{ij} = \sum_{i=1}^n a_{ij} A_{ij} \\
 &= \sum_{j=1}^n (-1)^{i+j} a_{ij} M_{ij} = \sum_{j=1}^n a_{ij} A_{ij}
 \end{aligned}
 \tag{40}$$

Note that the sum is taken over any one row or over any one column. In applying this formula, one seeks rows or columns with a large number of zeros to simplify the calculation of the determinant. We can show that this equation is consistent with the results given previously for the determinants of 2x2 and 3x3 arrays. Applying equation [40] to the third row of a 3x3 array gives the following result.

$$\text{Det } A_{(3 \times 3)} = \sum_{j=1}^3 a_{3j} A_{3j} = a_{31} A_{31} + a_{32} A_{32} + a_{33} A_{33}
 \tag{41}$$

We could have applied equation [40] to any of the three rows or any of the three columns to compute the determinant. I chose to use the third row since the necessary cofactors can be found in equation [39]. If we use equation [37] to expand the (2 x 2) cofactors in [39] and apply those results to equation [41], we obtain the following result.

$$\begin{aligned}
 \text{Det } A_{(3 \times 3)} &= a_{31} M_{31} - a_{32} M_{32} + a_{33} M_{33} \\
 &= a_{31} \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} - a_{32} \begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix} + a_{33} \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \\
 &= a_{31}(a_{12}a_{23} - a_{22}a_{13}) - a_{32}(a_{11}a_{23} - a_{21}a_{13}) + a_{33}(a_{11}a_{22} - a_{12}a_{21}) \\
 &= a_{31}a_{12}a_{23} - a_{31}a_{22}a_{13} - a_{32}a_{11}a_{23} + a_{32}a_{21}a_{13} + a_{33}a_{11}a_{22} - a_{33}a_{12}a_{21}
 \end{aligned}
 \tag{42}$$

The final result, after some rearrangement, is the same as the one in equation [38].

Two rules about determinants are apparent from equation [41]:

- A determinant is zero if any row or any column contains all zeros.
- If one row or one column of a determinant is multiplied by a constant, k, the value of the determinant is multiplied by the same constant. Note the implication for matrices: if a matrix is multiplied by a constant, k, then each matrix element is multiplied by k. If \mathbf{A} is an $n \times n$ matrix, $\text{Det}(k\mathbf{A}) = k^n \text{Det}(\mathbf{A})$.

Additional rules for and properties of determinants are stated below without proof.

- If one row (or one column) of a determinant is replaced by a linear combination of that row (or column) with another row (or column), the value of the determinant is not changed. This means that the operations of the Gauss elimination process do not change the determinant of a matrix.
- If two rows (or two columns) of a determinant are linearly dependent, the value of the determinant is zero.
- The determinant of the product of two matrices, \mathbf{A} and \mathbf{B} is the product of the determinants of the individual matrices: $\text{Det}(\mathbf{AB}) = \text{Det}(\mathbf{A}) \text{Det}(\mathbf{B})$.
- The determinant of transposed matrix is the same as the determinant of the original matrix: $\text{Det}(\mathbf{A}^T) = \text{Det}(\mathbf{A})$.

If we apply the column expansion of equation [41] to an upper triangular matrix, \mathbf{A} , we find that $\text{Det } \mathbf{A} = a_{11}A_{11}$, since the a_{11} term is the only term in the first column. We can apply equation [41] repeatedly to the cofactors. Each application shows that the determinant is simply the new term in the upper left of the array times its cofactor. Continuing in this fashion we see that the determinant of an upper triangular matrix is simply the product of the diagonal terms. We can combine this result with the fact noted above that the operations of the Gauss elimination process do not change the determinant of a matrix to develop a practical for computing determinants of any matrix. Apply Gauss elimination to get the matrix in upper triangular form then the determinant (of both the original matrix and the one in upper triangular form) is simply the product of the diagonal elements.

As an example consider the matrices from "Set 1" in the table on page 12. The original matrix

$$\text{was } \begin{bmatrix} 1 & -4 & 2 \\ 0 & 2 & 9 \\ 7 & 3 & 8 \end{bmatrix}; \text{ its upper triangular form was } \begin{bmatrix} 1 & -4 & 2 \\ 0 & 2 & 9 \\ 0 & 0 & 50.5 \end{bmatrix}. \text{ We can readily compute the}$$

determinant as the product $(1)(2)(50.5) = 101$. You can show that the same value is obtained by the conventional formula for the evaluation of the original 3×3 determinant.

Determinants are not used in normal numerical calculations. However if you need to find the numerical value for a large determinant, the process outlined above is the most direct numerical approach.

Cramer's rule gives the solution to a system of linear equations in terms of determinants. This approach is never used except in some theoretical applications. According to Cramer's rule the solution for a particular unknown x_i is the ratio of two determinants. The determinant in the denominator uses the usual matrix coefficients, a_{ij} . The determinant in the numerator consists of the a_{ij} coefficients except in one column. When we are solving for x_j we replace column i in the a_{ij} coefficients by the right-hand-side matrix coefficients, b_i . For a set of three equations in three unknowns, Cramer's rule would give the solutions shown in equation [43].

Cramer's rule allows us to find an analytical expression for the solution of a set of equations, and it is sometimes used to solve small sets of equations (2×2 or 3×3). However, it is never used for numerical calculations of larger systems because it becomes extremely time consuming the number of equations increases.

$$x_1 = \frac{\begin{vmatrix} b_1 & a_{12} & a_{13} \\ b_2 & a_{22} & a_{23} \\ b_3 & a_{32} & a_{33} \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}}, \quad x_2 = \frac{\begin{vmatrix} a_{11} & b_1 & a_{13} \\ a_{21} & b_2 & a_{23} \\ a_{31} & b_3 & a_{33} \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}}, \quad x_3 = \frac{\begin{vmatrix} a_{11} & a_{12} & b_1 \\ a_{21} & a_{22} & b_2 \\ a_{31} & a_{32} & b_3 \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}} \quad [43]$$

Determinants are also related to rank. An array in which the rows are not linearly independent will have a zero determinant. As an example of this consider the left-hand set of equations from equation [35]. Recall that the coefficients for that set of three equations had a rank of two because the equations were not linearly independent. When we evaluate the determinant for this array below, using equation [38] for the determinant of a (3×3) array, we find that the determinant is zero.

$$\begin{vmatrix} -1 & -4 & 3 \\ -4 & 11 & -6 \\ 1 & -8 & 5 \end{vmatrix} = (-1)(11)(5) + (-4)(-8)(3) + (1)(-4)(-6) \\ - (-1)(-8)(-6) - (-4)(-4)(5) - (1)(11)(3) \\ = -55 + 96 + 24 - (-48) - 80 - 33 = 0 \quad [44]$$

This gives us another approach to determining when a set of equations with all zeros on the right-hand side has a solution other than the simple one that all x_i are zero. This condition is that the determinant of the coefficient matrix is zero. If $\text{Det}(\mathbf{A}) = 0$ then a solution to $\mathbf{Ax} = \mathbf{b}$, where \mathbf{b} contains all zeros, that does not have all $x_i = 0$ is possible. There are actually an infinite number of such solutions. These solutions differ by an arbitrary multiplier. We will use this idea below when considering matrix eigenvalues.

Inverse of a Matrix

We have defined operations for adding, subtracting and multiplying matrices. Matrix inversion is the matrix analog of division. For a square matrix, \mathbf{A} , we define a matrix inverse, \mathbf{A}^{-1} , by the following equation.

$$\mathbf{AA}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I} \quad [45]$$

If we have a matrix equation, $\mathbf{Ax} = \mathbf{b}$, we can, in principle, solve this equation by premultiplying both sides of the equation by \mathbf{A}^{-1} . This gives the following result.

$$\text{If } \mathbf{Ax} = \mathbf{b}, \quad \mathbf{A}^{-1}\mathbf{Ax} = \mathbf{A}^{-1}\mathbf{b} \Rightarrow \mathbf{Ix} = \mathbf{A}^{-1}\mathbf{b} \Rightarrow \mathbf{x} = \mathbf{A}^{-1}\mathbf{b} \quad [46]$$

The various steps in equation [46] use the definition, in equation [45], that the product of a matrix and its inverse is the identity matrix and the definition, in equation [21] that the product of any matrix with the identity matrix is the original matrix. Although the result that $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$ may be written as the solution to the original equation, the actual solution of matrix equations like $\mathbf{Ax} = \mathbf{b}$ is done methods other than the direct calculation of the inverse. It is not always possible to find the inverse. A square matrix that has no inverse is called a **singular matrix**.

It is usually not necessary to find the inverse of a matrix. If necessary, you can find a numerical value of the inverse by the same process used to solve simultaneous linear algebraic equations. To understand how this is done, we define a second matrix, \mathbf{B} , as \mathbf{A}^{-1} . Then, by the definition of inverse we have the following equation.

$$\text{If } \mathbf{B} = \mathbf{A}^{-1}, \quad \mathbf{AB} = \mathbf{I} \quad [47]$$

Equation [48] shows the matrices involved in this equation.

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdots & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & & \vdots \\ \vdots & \vdots & \vdots & & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & b_{13} & \cdots & \cdots & b_{1n} \\ b_{21} & b_{22} & b_{23} & \cdots & \cdots & b_{2n} \\ b_{31} & b_{32} & b_{33} & \cdots & \cdots & b_{3n} \\ \vdots & \vdots & \vdots & \ddots & & \vdots \\ \vdots & \vdots & \vdots & & \ddots & \vdots \\ b_{n1} & b_{n2} & b_{n3} & \cdots & \cdots & b_{nn} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & 1 & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & & \vdots \\ \vdots & \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \cdots & 1 \end{bmatrix} \quad [48]$$

We have a form similar to the usual problem of solving a set of equations. The coefficient matrix, \mathbf{A} , is the same, but we have n right-hand side columns of known values. Each of these columns of known values corresponds to one column of unknowns in the \mathbf{B} matrix that is \mathbf{A}^{-1} . If we use our usual process for solving $\mathbf{Ax} = \mathbf{b}$, with, for example, $\mathbf{b} = [1 \ 0 \ 0 \ 0 \ \dots \ 0]^T$, we will obtain the first column of $\mathbf{B} = \mathbf{A}^{-1}$. Repeating the process for similar \mathbf{b} columns, which are all zeros except for a 1 in row k gives us column k of the inverse. For example, equation [49] shows the solution for the second column of $\mathbf{B} = \mathbf{A}^{-1}$.

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdots & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & & \vdots \\ \vdots & \vdots & \vdots & & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} b_{12} \\ b_{22} \\ b_{32} \\ \vdots \\ \vdots \\ b_{n2} \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ \vdots \\ 0 \end{bmatrix} \quad [49]$$

Because the operations for solving a set of simultaneous linear equations are based on the \mathbf{A} matrix only, the solution for the inverse is actually done simultaneously for all columns.

An analytical expression for the inverse can be obtained in terms of the cofactors discussed in the section on determinants. We continue to define $\mathbf{B} = \mathbf{A}^{-1}$; the components of the inverse, b_{ij} , are then given in terms of the minors or cofactors of the original \mathbf{A} matrix and its determinant.

$$\text{If } \mathbf{B} = \mathbf{A}^{-1}, \quad b_{ij} = \frac{A_{ji}}{\text{Det}(\mathbf{A})} = (-1)^{i+j} \frac{M_{ji}}{\text{Det}(\mathbf{A})} \quad [50]$$

The simplest application of this equation is to a 2x2 matrix. For such a matrix,

$$\begin{aligned} b_{11} &= (-1)^{1+1} \frac{M_{11}}{\text{Det}(\mathbf{A})} = \frac{a_{22}}{\text{Det}(\mathbf{A})} & b_{12} &= (-1)^{1+2} \frac{M_{21}}{\text{Det}(\mathbf{A})} = -\frac{a_{12}}{\text{Det}(\mathbf{A})} \\ b_{21} &= (-1)^{2+1} \frac{M_{12}}{\text{Det}(\mathbf{A})} = -\frac{a_{21}}{\text{Det}(\mathbf{A})} & b_{22} &= (-1)^{2+2} \frac{M_{22}}{\text{Det}(\mathbf{A})} = \frac{a_{11}}{\text{Det}(\mathbf{A})} \end{aligned} \quad [51]$$

Combining the results of equation [51] with equation [37] for a 2x2 determinant, gives the following result for the inverse of a 2x2 matrix.

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^{-1} = \frac{1}{a_{11}a_{22} - a_{21}a_{12}} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix} \quad [52]$$

You can easily show that this is correct by multiplying the original matrix by its inverse. You will obtain a unit matrix by either multiplication: \mathbf{AA}^{-1} or $\mathbf{A}^{-1}\mathbf{A}$. The same process can be used to find the inverse of a (3x3) matrix; the result is shown below:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}^{-1} = \frac{\begin{bmatrix} (a_{22}a_{33} - a_{32}a_{23}) & (a_{32}a_{13} - a_{33}a_{12}) & (a_{12}a_{23} - a_{22}a_{13}) \\ (a_{31}a_{23} - a_{33}a_{21}) & (a_{11}a_{33} - a_{31}a_{13}) & (a_{21}a_{13} - a_{11}a_{23}) \\ (a_{21}a_{32} - a_{31}a_{22}) & (a_{31}a_{12} - a_{11}a_{32}) & (a_{11}a_{22} - a_{21}a_{12}) \end{bmatrix}}{\begin{pmatrix} a_{11}a_{22}a_{33} + a_{21}a_{32}a_{13} + a_{31}a_{12}a_{23} \\ -a_{11}a_{32}a_{23} - a_{21}a_{12}a_{33} - a_{31}a_{22}a_{13} \end{pmatrix}} \quad [53]$$

Equations [52] and [53] show the value of determinants in providing analytical solutions to inverses. Although determinants are valuable in such cases any use of determinants should be avoided in numerical work.

The general rule for the inverse of a matrix product and the inverses of the individual matrices is similar to the same equation for the transpose of a matrix product and the product of the transposes of the individual matrices. This relation is shown below.

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1} \quad (\mathbf{ABC})^{-1} = \mathbf{C}^{-1}\mathbf{B}^{-1}\mathbf{A}^{-1} \quad (\mathbf{ABCD})^{-1} = \dots \quad [54]$$

Matrix Eigenvalues and Eigenvectors

If a square matrix can premultiply a column vector and return the original column vector multiplied by a scalar, the scalar is said to be an eigenvalue of the matrix and the column vector is called an eigenvector. In the following equation, the scalar, λ , is an eigenvalue of the matrix \mathbf{A} , and \mathbf{x} is an eigenvector.

$$\mathbf{A}_{(n \times n)}\mathbf{x}_{(n \times 1)} = \lambda\mathbf{x}_{(n \times 1)} \quad [55]$$

We can use the identity matrix to rewrite this equation as follows.

$$[\mathbf{A}_{(n \times n)} - \mathbf{I}_{(n \times n)}\lambda]\mathbf{x}_{(n \times 1)} = \mathbf{0}_{(n \times 1)} \quad [56]$$

$$\begin{bmatrix} a_{11} - \lambda & a_{12} & a_{13} & \cdots & \cdots & a_{1n} \\ a_{21} & a_{22} - \lambda & a_{23} & \cdots & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} - \lambda & \cdots & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & & \vdots \\ \vdots & \vdots & \vdots & & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & \cdots & a_{nn} - \lambda \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ \vdots \\ x_m \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ \vdots \\ 0 \end{bmatrix} \quad [57]$$

As discussed above in the sections on simultaneous linear equations and determinants, equation [57] has the solution that all values of x_i are zero. It may have a nonzero solution if the determinant of the coefficient matrix is zero. That is,

$$\text{Det}[\mathbf{A} - \mathbf{I}\lambda] = \text{Det} \begin{bmatrix} a_{11} - \lambda & a_{12} & a_{13} & \cdots & \cdots & a_{1n} \\ a_{21} & a_{22} - \lambda & a_{23} & \cdots & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} - \lambda & \cdots & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & & \vdots \\ \vdots & \vdots & \vdots & & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & \cdots & a_{nn} - \lambda \end{bmatrix} = 0 \quad [58]$$

From the general expression for a determinant, we see that one component of the final expression for a determinant of any size is the product of all elements on the principal diagonal. In equation [58] this term will give an n^{th} order polynomial in λ (for our $n \times n$ matrix). This n^{th} order polynomial is known as the **characteristic equation** of the matrix. This characteristic equation can be solved for n values of λ , not all of which may be distinct. For a two-by-two matrix, setting $\text{Det}[\mathbf{A} - \mathbf{I}\lambda]=0$ gives the following quadratic equation.

$$\begin{vmatrix} a_{11} - \lambda & a_{12} \\ a_{21} & a_{22} - \lambda \end{vmatrix} = (a_{11} - \lambda)(a_{22} - \lambda) - a_{21}a_{12} = \lambda^2 - (a_{11} + a_{22})\lambda + a_{11}a_{22} - a_{21}a_{12} = 0 \quad [59]$$

We can solve the quadratic equation in [59] to get two roots that give us the two possible eigenvalues:

$$\lambda = \frac{(a_{11} + a_{22}) \pm \sqrt{(a_{11} + a_{22})^2 + 4(a_{21}a_{12} - a_{11}a_{22})}}{2} \quad [60]$$

Each eigenvalue will have its own eigenvector. Each eigenvector is found by the solution of equation [53]. If we denote the eigenvectors as $\mathbf{x}_{(1)}$ and $\mathbf{x}_{(2)}$, the components of eigenvector j may be written as $x_{(j)1}$ and $x_{(j)2}$. Accordingly, we have to solve the set of equations shown below two times once for λ_1 and once for λ_2 .

$$\begin{aligned} (a_{11} - \lambda_j)x_{(j)1} + a_{12}x_{(j)2} &= 0 \\ a_{21}x_{(j)1} + (a_{22} - \lambda_j)x_{(j)2} &= 0 \end{aligned} \quad [61]$$

Again, the solution is not unique. Any set of \mathbf{x} values, multiplied by an arbitrary constant, will satisfy this set of equations. For simplicity we pick $x_{(j)1} = \alpha$. We have two possible results for the eigenvector component, $x_{(j)2}$ depending on which equation we use.

$$x_{(j)1} = \alpha \quad x_{(j)2} = \frac{(\lambda_j - a_{11})}{a_{12}} \alpha = \frac{a_{21}}{(\lambda_j - a_{22})} \alpha \quad [62]$$

There appear to be two different solutions for $x_{(j)2}$, depending on the use of the first or second equation to get this eigenvector component. However, equating these two values for $x_{(j)2}$, will eliminate the arbitrary constant, α , and obtain equation [59] that we solved for λ . Thus the two possible expressions for $x_{(j)2}$ in equation [62] will result in the same value.

The approach outlined above for finding eigenvalues and eigenvectors can, in principle, be applied to any size of matrix. However, numerical methods are used to find the eigenvalues and eigenvectors of large matrices. The eig function of MATLAB can be used to find both eigenvalues and eigenvectors.

As a numerical example, consider the determination of the eigenvalues and eigenvectors for the

matrix, $\mathbf{A} = \begin{bmatrix} 1 & 5 \\ 0 & 2 \end{bmatrix}$. You can find the answer using equations [61] and [62]. However, we will

outline the entire solution process as an example of finding eigenvalues and eigenvectors for larger systems. Solving the equation $\text{Det}[\mathbf{A} - \mathbf{I}\lambda]$ for this matrix gives the following result.

$$\text{Det}[\mathbf{A} - \mathbf{I}\lambda] = \begin{vmatrix} 1-\lambda & 5 \\ 0 & 2-\lambda \end{vmatrix} = (1-\lambda)(2-\lambda) - (0)(5) = 0. \text{ The roots to this equation are } \lambda_2 =$$

2 and $\lambda_1 = 1$. We now substitute each eigenvalues into the equation $(\mathbf{A} - \mathbf{I}\lambda_k)\mathbf{x}_{(k)} = \mathbf{0}$, and solve for the components of each eigenvector. For the first eigenvector we obtain.

$$\begin{aligned} (1-2)x_{(1)1} + 5x_{(1)2} &= 0 \\ 0x_{(1)1} + (2-2)x_{(1)2} &= 0 \end{aligned} \quad [63]$$

We see that the last equation results in $0 = 0$, which gives us no useful information. Since we know that the homogenous equation set has an infinite number of solutions, we pick an arbitrary value, α , for $x_{(1)2}$. With this value, the first equation gives us the result that $x_{(1)1} = 5\alpha$. Thus our first eigenvector, $\mathbf{x}_{(1)} = [5\alpha \ \alpha]^T$. We can apply the same procedure to find the second eigenvector.

$$\begin{aligned} (1-1)x_{(2)1} + 5x_{(2)2} &= 0 \\ 0x_{(2)1} + (2-1)x_{(2)2} &= 0 \end{aligned} \quad [64]$$

Here both equations tell us that $x_{(2)2}$ must be zero. However, there is no information about $x_{(2)1}$. We conclude that this must be an arbitrary quantity that we will call β . This gives our second eigenvector, $\mathbf{x}_{(2)} = [\beta \ 0]^T$. We can verify our solution for eigenvalues and eigenvectors by showing that they satisfy the defining equation, $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$.

$$\mathbf{A}\mathbf{x}_{(1)} = \begin{bmatrix} 1 & 5 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 5\alpha \\ \alpha \end{bmatrix} = \begin{bmatrix} (1)(5\alpha) + (5)(\alpha) \\ (0)(5\alpha) + (2)(\alpha) \end{bmatrix} = \begin{bmatrix} 10\alpha \\ 2\alpha \end{bmatrix} = 2 \begin{bmatrix} 5\alpha \\ \alpha \end{bmatrix} = \lambda_1 \mathbf{x}_{(1)} \quad [65]$$

$$\mathbf{A}\mathbf{x}_{(2)} = \begin{bmatrix} 1 & 5 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} \beta \\ 0 \end{bmatrix} = \begin{bmatrix} (1)(\beta) + (5)(0) \\ (0)(\beta) + (2)(0) \end{bmatrix} = \begin{bmatrix} \beta \\ 0 \end{bmatrix} = 1 \begin{bmatrix} \beta \\ 0 \end{bmatrix} = \lambda_2 \mathbf{x}_{(2)} \quad [66]$$

The calculations above show that the definition of eigenvalues and eigenvectors is satisfied regardless of our choices for α and β . This is a general result. We are always free to choose one component of the eigenvector. However, the remaining components will be set. Typically the

eigenvector components are chosen to give a simple expression for the eigenvector (*i.e.*, one in which all the components are integers or simple fractions) or a unit vector.⁸

In the example of the two-by-two matrix used above, we could express the eigenvectors we found in any of the ways shown immediately below. The last expression shown for each eigenvector is a unit vector. Note that the two eigenvectors are not orthogonal in this example.

$$\mathbf{x}_{(1)} = \begin{bmatrix} 5\alpha \\ \alpha \end{bmatrix} \quad \mathbf{x}_{(1)} = \begin{bmatrix} 5 \\ 1 \end{bmatrix} \quad \mathbf{x}_{(1)} = \begin{bmatrix} 5/\sqrt{26} \\ 1/\sqrt{26} \end{bmatrix} \quad \mathbf{x}_{(2)} = \begin{bmatrix} \beta \\ 0 \end{bmatrix} \quad \mathbf{x}_{(2)} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad [67]$$

⁸ A unit vector, \mathbf{u} , is a row or column vector for which the two norm, $\|\mathbf{u}\|_2 = \sqrt{\sum u_i^2} = 1$. If we have a vector, \mathbf{v} , that is not a unit vector we can convert it into a unit vector by dividing each component by $\|\mathbf{v}\|_2$.

This would give the components of the new unit vector by the following equation: $u_i = \frac{v_i}{\sqrt{\sum v_k^2}}$