CHAPTER 4

Linear Transformations from a Geometric Viewpoint

4.1 INTRODUCTION

In Chapter 3 the reader was exposed briefly to one type of transformation, namely, that involving the rotation of either a point or a set of basis vectors by an orthogonal matrix. A large part of multivariate analysis is concerned with linear transformations, of which rotation represents only one type—albeit an important special case.

The purpose of this chapter is to describe the geometric aspects of various kinds of matrix transformations. In particular, we shall be interested in how general linear transformations can be viewed as composites of simple kinds of matrices that, individually, are more easily portrayed from a geometric viewpoint.

In the course of describing matrix transformations geometrically, two additional concepts of major importance to multivariate analysis—matrix inversion and the rank of a matrix—are described and illustrated numerically. These concepts are useful in the solution of sets of linear equations and play important roles in multivariate analysis.

The chapter starts out with an overview description of matrix transformations and their representation as sets of linear equations. Then the topic of orthogonal transformations, first introduced in Chapter 3, is reviewed and expanded upon. The distinction between point and basis vector transformations is also emphasized. Discussion then proceeds to more extensive cases involving general linear transformations. In the case of basis vector changes, these kinds of transformations require the use of matrix inverses; hence, the basic ideas of matrix inversion are introduced at this point.

The next major section of the chapter is devoted to the geometric representation of various types of matrix transformations, such as rotations, stretches, central dilations and reflections. The geometric character of combinations of various matrix transformations is also illustrated so that the reader can see how simple geometric changes, when taken in combination, lead to complex representations.

The remainder of the chapter focuses on the solution of simultaneous equations and the central roles that matrix inversion and matrix rank play in this activity. In particular, we discuss the solution of linear equations in multivariate analysis and, in the process, tie in the present topic with material presented in earlier chapters on determinants and the pivotal method for solving sets of linear equations.

4.2 SIMULTANEOUS EQUATIONS AND MATRIX TRANSFORMATIONS

The concept of a function or mapping is fundamental to all mathematics. By a mapping we mean an operation by which elements of one set of mathematical entities are transformed into elements of another. In scalar algebra we recall that functions like the following are often employed:

$$y = f(x) = bx; \qquad y = f(x) = e^x$$

$$y = f(x) = ax^b; \qquad y = f(x) = ab^x$$

For example, for a specific value of x, and a value for the parameter b, we can find a value of y from the linear equation $y = bx$. The possible values that x can assume are called the *domain* of the function. The possible values that y can assume are called the *range* of the function.

In scalar algebra our interest centers on the description of rules (i.e., the functions) by which pairs of numbers are related. In vector algebra we are interested in the rules by which pairs of vectors or points are related.

In our discussion of vector mappings we consider only single-valued, linear mappings of one vector space onto another, which may, of course, be the same space. Thus, we are interested in cases where both y and x are vectors.

By restricting ourselves to linear transformations–illustrated by $y = bx$ above–we can state three conditions of interest:

1. Every vector of a vector space is transformed into a uniquely determined vector of the space.
2. If $\mathbf{a}$ is transformed to $\mathbf{a}^*$ by a linear transformation $\mathbf{T}$, then $k\mathbf{a}$ is transformed (by $\mathbf{T}$) to $k\mathbf{a}^*$ for any scalar k.
3. If $\mathbf{a}$ is transformed to $\mathbf{a}^*$ and $\mathbf{b}$ to $\mathbf{b}^*$ by a linear transformation $\mathbf{T}$, then $\mathbf{a} + \mathbf{b}$ is transformed (by $\mathbf{T}$) to $\mathbf{a}^* + \mathbf{b}^*$.

Linear transformations are those that satisfy the above conditions. Moreover, any linear transformation can be represented in matrix form (e.g., as the multiplication of a vector by a matrix).

Returning to the topic of mappings, values obtained by a mapping are often called *images,* while the values being transformed are often called *preimages* of the transformation. Here we are mainly concerned with mappings that transform vectors into vectors, that is, mappings that represent vector-valued functions. Hence, both the preimages and the images of the mapping are vectors. Moreover, we shall mostly be concerned with linear transformations that involve square matrices as representations of the transformation so that the two vector spaces, before and after the transformation, are of the same dimensionality.[1]

[1] Or, essentially the same space, before and after the transformation.

4.2.1 Simultaneous, Linear Equations Expressed in Matrix Form

The need for matrix transformation arises quite naturally in the solution of linear equations. Consider the following set of linear equations:

$$\begin{aligned} x_1^* &= a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ x_2^* &= a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ &\vdots \\ x_n^* &= a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n \end{aligned}$$

which can be written, in matrix form, as

$$\mathbf{x}^* = \mathbf{A}\mathbf{x}$$

As a simple case of expressing a matrix transformation as a set of simultaneous equations, consider the point $\mathbf{x} = \left[\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right]$ in Fig. 4.1. Also consider the second point $\mathbf{x}^* = \left[\begin{smallmatrix} x_1^* \\ x_2^* \end{smallmatrix}\right]$ in the same figure. A linear mapping of $\mathbf{x}$ onto $\mathbf{x}^*$ can always be expressed by a set of linear equations:

$$x_1^* = a_{11}x_1 + a_{12}x_2; \qquad x_2^* = a_{21}x_1 + a_{22}x_2$$

which relate the coordinates x_1^*, x_2^* to the coordinates x_1, x_2 in the standard basis $\mathbf{e}_i$.

But, as noted above, these equations can be expressed in matrix form as

$$\begin{bmatrix} x_1^* \\ x_2^* \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

where the matrix

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

represents the mapping of $\mathbf{x}$, the preimage, onto $\mathbf{x}^*$, the image.

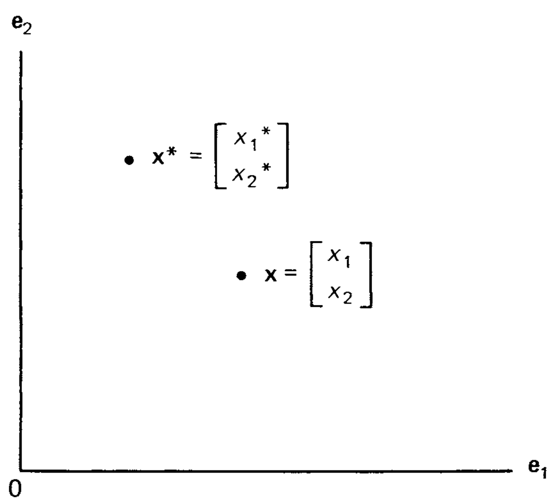

Fig. 4.1 A point transformation.

All of the mappings that we shall be considering are capable of being expressed in terms of a set of linear equations, similar to those illustrated above.

We review some simple examples of matrix transformations, first considered in Chapter 3, and then show how these cases can be extended to more general kinds of linear mappings.

4.2.2 Orthogonal Matrix Transformations

As recalled from Chapter 3, an orthogonal matrix $\mathbf{A}$ is one in which $\mathbf{A}'\mathbf{A} = \mathbf{A}\mathbf{A}' = \mathbf{I}$. That is, rows (and columns) of $\mathbf{A}$ are mutually orthogonal, and each is of unit length. This type of transformation is called a rotation, either proper or improper, depending upon the sign of its determinant.

In the preceding chapter, where we introduced the reader to orthogonal transformations, we recall that rotations were expressed as sets of direction cosines. Although the distinction was not emphasized, we also illustrated two types of rotations:

1. Point rotations, where the original basis vectors remained fixed and the point(s) moved, clockwise or counterclockwise, around the origin.
2. Basis vector rotations, where the original point(s) remained fixed and the basis vectors moved, clockwise or counterclockwise. In this latter case the fixed point was then expressed as a linear combination of the new basis vectors.

Since rotations deal with *relative* motion, either of the above approaches is equally appropriate in interpreting the nature of a rotation.

Figure 4.2 illustrates the four cases that are involved in rotating an arbitrary point $\mathbf{x}$. In Panel I $\mathbf{x}$ undergoes a counterclockwise rotation, mapping onto the point $\mathbf{y}$. Alternatively, we can rotate $\mathbf{x}$ clockwise to map it onto the point $\mathbf{z}$. Note that the standard basis vectors remain fixed throughout both of these rotations.

Panel II of Fig. 4.2 shows the set of basis vector rotations in which the $\mathbf{e}_i$ are mapped onto $\mathbf{f}_i$ via a counterclockwise rotation. Panel III shows the case of mapping the $\mathbf{e}_i$ onto $\mathbf{g}_i$ via a clockwise rotation. In each of these latter two cases, the coordinates of $\mathbf{x}$ are in terms of the new basis vectors. As will be shown, all four types of rotations are variations on a common theme and are related to each other in a straightforward way.

Now, however, we shall want to distinguish more carefully between point transformations and basis vector transformations by adopting a specific notation for each. An image obtained by a point transformation, in which the original basis vectors remain fixed, is denoted by $\mathbf{x}^*$. If obtained by a change in basis vectors—and, hence, the fixed point is referred to the *new* basis—the image is denoted by $\mathbf{x}^\circ$.

Point transformations are much less complicated than basis vector transformations. However, it is important to study both kinds since many multivariate techniques involve situations in which the change of basis vectors simplifies the character of the transformation quite markedly.

As a quick review of the two types of rotations, suppose we choose a standard basis $\mathbf{e}_1 = \left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]$ and $\mathbf{e}_2 = \left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]$ and a point in that space $\mathbf{x} = \left[\begin{smallmatrix}1\\2\end{smallmatrix}\right] = 1\mathbf{e}_1 + 2\mathbf{e}_2$. This linear combination of vectors can also be written as

$$\mathbf{x} = \begin{bmatrix}1\\2\end{bmatrix} = 1\begin{bmatrix}1\\0\end{bmatrix} + 2\begin{bmatrix}0\\1\end{bmatrix}$$

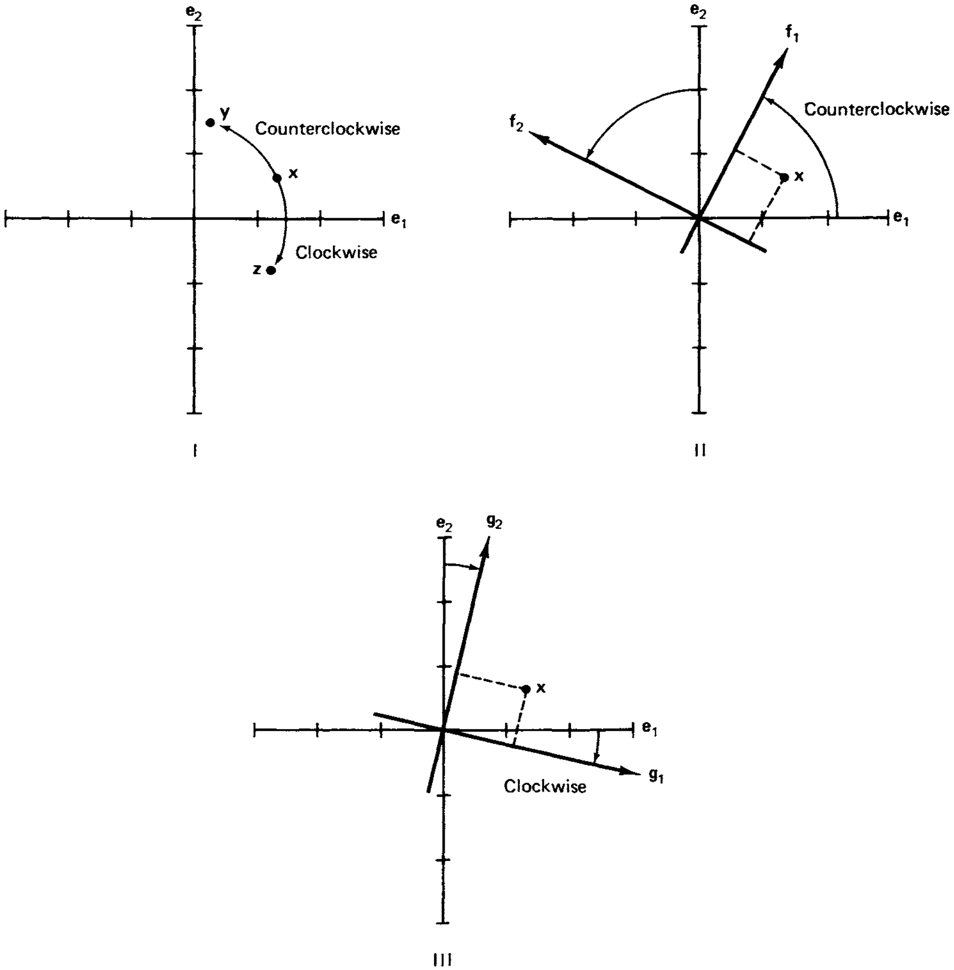

Fig. 4.2 Point and basis vector transformations.

where the standard basis vectors are shown explicitly as column vectors. Let us first consider point rotations and then basis vector rotations.

4.2.2.1 Point Rotations Suppose we examine the rotation of $\mathbf{x} = \left[\begin{smallmatrix}1\\2\end{smallmatrix}\right]$ in a counterclockwise direction through an angle Ψ of 30°. In the case of two dimensions, we have, in general, the equations

$$x_1{}^* = a_{11}x_1 + a_{12}x_2; \qquad x_2{}^* = a_{21}x_1 + c_{22}x_2$$

Here we use $\mathbf{x}^*$ to denote the image of the vector $\mathbf{x}$ under the matrix transformation $\mathbf{A}$.

We assume $\mathbf{A}$ to be orthogonal. The system of equations can be written as

$$\mathbf{x}^* = \mathbf{A}\mathbf{x}$$

Since only two dimensions are involved, we can simplify the problem a bit by relating all direction cosines to the single angle of $\Psi = 30°$, as was illustrated in Chapter 3. Hence for this specific example we have

$$\begin{bmatrix} x_1^* \\ x_2^* \end{bmatrix} = \begin{bmatrix} \cos\Psi & -\sin\Psi \\ \sin\Psi & \cos\Psi \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$\begin{bmatrix} -0.13 \\ 2.23 \end{bmatrix} = \begin{bmatrix} 0.87 & -0.50 \\ 0.50 & 0.87 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

Notice in this illustration that we made use of the trigonometric facts (involving complementary angles) that

1. $\cos(90° + \Psi) = -\sin\Psi = \cos 120° = -\sin 30° = -0.5$
2. $\cos(90° - \Psi) = \sin\Psi = \cos\ 60° = \sin 30° = 0.5$

This device allows us to avoid the more complex application of direction cosines involving pairs of angles θ_{ij}, as discussed in Chapter 3, although its use, as recalled, is restricted to two dimensions.

Figure 4.3 shows the results of this mapping. We observe specifically that the new coordinates $\mathbf{x}^* = \left[\begin{smallmatrix} -0.13 \\ 2.23 \end{smallmatrix}\right]$ of the point are still expressed in terms of the old basis, namely, $\mathbf{e}_1$ and $\mathbf{e}_2$. It is the *point*, that is rotated counterclockwise, while the axes maintain their original orientation.

Now let us see what happens when we rotate the point $\mathbf{x} = \left[\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right]$ *clockwise* through an angle of $\Psi = 30°$, as given by

$$\begin{bmatrix} x_1^* \\ x_2^* \end{bmatrix} = \begin{bmatrix} \cos\Psi & \sin\Psi \\ -\sin\Psi & \cos\Psi \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$\begin{bmatrix} 1.87 \\ 1.23 \end{bmatrix} = \begin{bmatrix} 0.87 & 0.50 \\ -0.50 & 0.87 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

In this case the new point coordinates are given, still in terms of the original basis vectors, in Fig. 4.4. Notice, in the case of a clockwise rotation of the point, the matrix $\mathbf{A}'$ is the *transpose* of that ($\mathbf{A}$) used to rotate the point counterclockwise. Finally, if we first rotate the point counterclockwise, given by $\mathbf{A}$, and then rotate it clockwise, given by $\mathbf{A}'$, we end up where we started, since in the case of orthogonal matrices:

$$\mathbf{A}'\mathbf{A} = \mathbf{I}$$

Similarly, had we started with the clockwise rotation and "undone" this by means of the counterclockwise rotation, we would have

$$\mathbf{A}\mathbf{A}' = \mathbf{I}$$

which, again in the special case of an orthogonal matrix, gets us back to where we started.

Point transformations—either orthogonal or more general linear transformations—present relatively few problems and can all be represented simply by

$$\mathbf{x}^* = \mathbf{A}\mathbf{x}$$

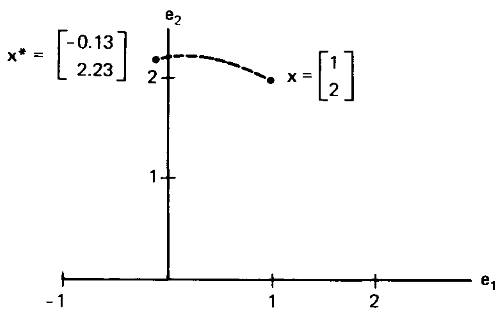

Fig. 4.3

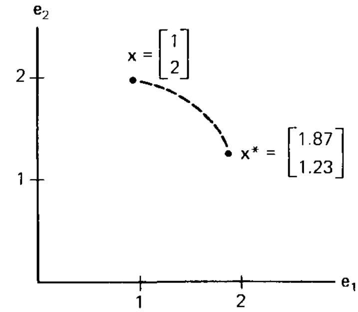

Fig. 4.4

Fig. 4.3 Counterclockwise rotation of point–new coordinates given in terms of old basis vectors.
Fig. 4.4 Clockwise rotation of point–new coordinates given in terms of old basis vectors.

where we should remember, of course, that it is the point(s) or vector(s) that moves relative to a *fixed basis.* Moreover, each row of the transformation matrix represents a linear combination of the *original* point coordinates.

In Chapter 3, however, we also considered the case of rotations which involved *basis vector* changes. In this case the point(s) remains fixed, but is then referred to a set of new (rotated) basis vectors.

*4.2.2.2 **Basis Vector Transformations*** In discussing rotations of basis vectors, we shall want to review a few fundamentals and introduce some new features as well. To be specific, suppose we wish to rotate a set of standard basis vectors clockwise through an angle of 30°. As noted above, the orthogonal matrix that effects this type of rotation is

$$\mathbf{A}' = \begin{bmatrix} \cos\Psi & \sin\Psi \\ -\sin\Psi & \cos\Psi \end{bmatrix} = \begin{bmatrix} 0.87 & 0.50 \\ -0.50 & 0.87 \end{bmatrix}$$

Let us start out with the standard basis vectors $\mathbf{e}_1$ and $\mathbf{e}_2$. Based on our earlier remarks–and as verified by Fig. 4.5–we see that a clockwise rotation of the basis vectors results in a new basis in which the fixed point $\mathbf{x} = \left[\begin{smallmatrix}1\\2\end{smallmatrix}\right]$ is now $\mathbf{x}^\circ = \left[\begin{smallmatrix}-0.13\\2.23\end{smallmatrix}\right]$ in terms of the new basis.[2]

To distinguish the two types of transformations, we let $\mathbf{x}^\circ$ denote the image of $\mathbf{x}$ under a basis vector transformation, while $\mathbf{x}^*$ denotes its image under a point transformation. Note further that the coordinates of $\mathbf{x}^\circ$ are the same as those found when the point was rotated counterclockwise. This is as it should be inasmuch as we are concerned with only relative motion.

The new basis vectors $\mathbf{f}_1$ and $\mathbf{f}_2$ can be expressed in terms of the original as

$$\begin{array}{ccc} \begin{array}{cc} \mathbf{f}_1 & \mathbf{f}_2 \end{array} & \mathbf{E} & \mathbf{A}' \\ \mathbf{F} = \begin{bmatrix} 0.87 & 0.50 \\ -0.50 & 0.87 \end{bmatrix} = & \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} & \begin{bmatrix} 0.87 & 0.50 \\ -0.50 & 0.87 \end{bmatrix} \end{array}$$

[2] It is well to remember that the vector x exists independently of its coordinates. However, the specific *coordinate* values that it assumes *are* dependent upon the basis vectors to which it is referred.

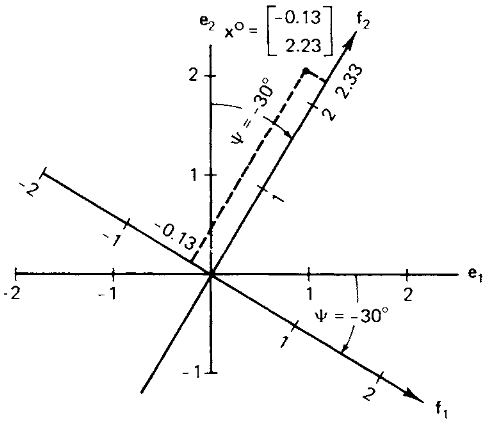

Fig. 4.5

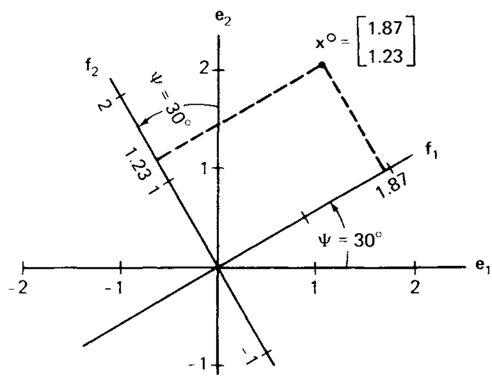

Fig. 4.6

Fig. 4.5 Clockwise rotation of axes—old coordinates given in terms of new basis vectors.

Fig. 4.6 Counterclockwise rotation of axes—old coordinates given in terms of new basis vectors.

Here we employ $\mathbf{E}$ to denote the matrix of the original (standard) basis and $\mathbf{F}$ to denote the matrix of the new basis. Also, note that the columns of $\mathbf{F}$ represent linear combinations of the (column) basis vectors of $\mathbf{E}$.

As can be seen in Fig. 4.5, $\mathbf{f}_1$, the first column of $\mathbf{F}$, passes through the point (0.87, −0.5) positioned in the original $\mathbf{e}_i$ basis, while $\mathbf{f}_2$ passes through the point (0.5, 0.87) in the original $\mathbf{e}_i$ basis.

In this case the original vector $\mathbf{x} = \begin{bmatrix}1\\2\end{bmatrix}$ is now $\mathbf{x}^\circ = \begin{bmatrix}-0.13\\2.23\end{bmatrix}$ when referred to the $\mathbf{f}_1, \mathbf{f}_2$ basis. Recall that these are the coordinates found by a 30° *counterclockwise* rotation of the point in the original basis.

Similarly, if the original basis vectors are rotated counterclockwise, this gives us the same result as found by rotating the point clockwise. A picture of this change in basis vectors is shown in Fig. 4.6.

At this point it might sound confusing and redundant to have two ways, point and basis vector changes, for expressing linear transformations. As we shall try to show later, however, there are advantages to considering basis vector as well as point transformations. This is particularly true when the transformation matrix is not orthogonal; that is, when the transformation does not entail a simple rotation.

To sum up, if we wish to find the coordinates of transformed $\mathbf{x}^*$ relative to the standard $\mathbf{e}_1, \mathbf{e}_2$ basis (i.e., to move the point relative to the fixed basis), we use the point transformation

$$\mathbf{x}^* = \mathbf{A}\mathbf{x}$$

where $\mathbf{x}$ is originally referred to the $\mathbf{e}_1, \mathbf{e}_2$ basis. In this case the coordinates of the point $\mathbf{x}^*$ are also expressed directly in terms of $\mathbf{e}_1, \mathbf{e}_2$. (See Figs. 4.3 and 4.4.)

Alternatively,we may care to transform the $\mathbf{e}_i$ *basis itself* to a new set of basis vectors $\mathbf{f}_i$. In this case it is the new basis vectors that are expressed in terms of $\mathbf{e}_i$, while the old coordinates of the point $\mathbf{x}$ are now expressed as $\mathbf{x}^\circ$ in terms of the new basis $\mathbf{f}_i$. (See Figs. 4.5 and 4.6.) Notice, then, that the values of the new coordinates *depend on which method we use to carry out the transformation.*

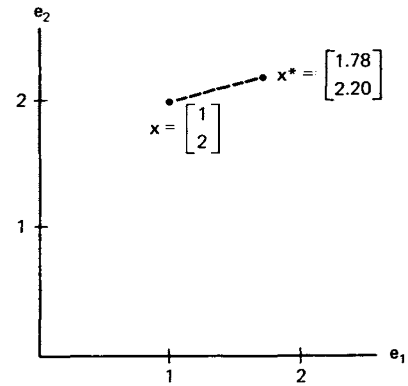

Fig. 4.7 Point transformation of x with fixed basis vectors.

4.2.3 Generalizing the Results

It is now of interest to describe what goes on when we do *not* restrict ourselves to rotations. Consider the more general transformation matrix

$$\mathbf{T} = \begin{bmatrix} 0.90 & 0.44 \\ 0.60 & 0.80 \end{bmatrix}$$

and the vector $\mathbf{x} = \left[\begin{smallmatrix}1\\2\end{smallmatrix}\right]$, again relative to the standard basis vectors

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{and} \quad \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Application of a point transformation to x (retaining the original basis) is quite straightforward and is shown in Fig. 4.7. Here we see that x moves to the position

$$\mathbf{x}^* = \begin{bmatrix} x_1{}^* \\ x_2{}^* \end{bmatrix} = \begin{bmatrix} 1.78 \\ 2.20 \end{bmatrix} = \begin{bmatrix} 0.90 & 0.44 \\ 0.60 & 0.80 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

However, notice that this point movement does *not* involve a simple rotation. Still, the process of finding the image vector x* presents no new problems. That is, we still have a straightforward task of premultiplying the original point by the matrix of the transformation **T**. Moreover, x* is still referred to the original $\mathbf{e}_i$ basis.

Unfortunately, things are not so simple when we attempt to construct the counterpart basis vector transformation. Unlike the special case of rotation, the present transformation matrix **T** presents complications in referring x to a new basis $\mathbf{f}_i$. Rather than try to solve this problem now, we can address ourselves to a related problem in point transformations. Solution of this related problem will pave the way for handling the basis vector transformation in the case of the linear transformation matrix **T**.

The related problem can be expressed as follows: Suppose we were given the point $\mathbf{x}^* = \left[\begin{smallmatrix}1.78\\2.20\end{smallmatrix}\right]$ in Fig. 4.7 to begin with and wanted to find x, knowing only the transformation matrix **T** and the original basis **E**. In other words, we now wish to find a transformation, call it $\mathbf{T}^{-1}$, that will get us *back* to x, given the transformed coordinates

$\mathbf{x}^*$. For this task we shall need to describe the nature of $\mathbf{T}_r^{-1}$, the *inverse* of T. Discussion of matrix inversion will require a brief digression, after which we can return to the particular problem at hand.

4.3 MATRIX INVERSION

In Chapter 2 we briefly described a special diagonal matrix, referred to as the identity matrix **I**. As discussed there, **I** consists of a square matrix with 1's along the main diagonal and zeros elsewhere. As will be shown, the identity matrix— and the concept of matrix inverse—play special roles in the matrix algebra "equivalent" of division (or, more appropriately, multiplication by a reciprocal). We recall from Chapters 2 and 3 that although we have discussed addition, subtraction, and multiplication as operations in matrix algebra, we have not discussed, as yet, the companion operation of division. As we shall see, matrix inversion in linear algebra is analogous to the operation of division in scalar algebra.

The identity matrix **I** plays a special role in matrix algebra. The effect of pre- or postmultiplying any conformable matrix by **I** is to leave the original matrix unchanged:

$$\mathbf{IA} = \mathbf{AI} = \mathbf{A}$$

That is, **I** in matrix algebra plays the role of the number 1 in ordinary arithmetic. We now ask if there is a type of matrix that is analogous to the reciprocal of a number, that is, a number, $a \neq 0$, for which the relation $ax = 1$ is true. If so, we should be able to develop the concept of "division" in the context of matrix algebra.

In the case of *square matrices* there is an analogue, in some instances, to the notion of a reciprocal in scalar arithmetic. This is called an *inverse.* As a matter of fact, matrix inverses—in a very generalized sense—can be obtained for rectangular matrices too. We discuss this more advanced topic in Appendix B while here we confine ourselves to *regular* inverses of *square* matrices.

A regular inverse of the (square) matrix **A** *is denoted by* $\mathbf{A}^{-1}$ *and, when it exists, it is unique and satisfies the following relations:*

$$\boxed{\mathbf{AA}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}}$$

For example, let us take the matrix

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

Assume, now, that we have gone on to solve for $\mathbf{A}^{-1}$ and have found it to be

$$\mathbf{A}^{-1} = \begin{bmatrix} 10/14 & -8/14 & 2/14 \\ 2/14 & 4/14 & -1/14 \\ -6/14 & 2/14 & 3/14 \end{bmatrix}$$

Then the relation $\mathbf{AA}^{-1}$ (or $\mathbf{A}^{-1}\mathbf{A}$) = **I** should hold. The reader can verify that it does:

$$\overset{\mathbf{A}}{\begin{bmatrix} 1 & 2 & 0 \\ 0 & 3 & 1 \\ 2 & 2 & 4 \end{bmatrix}} \overset{\mathbf{A}^{-1}}{\begin{bmatrix} 10/14 & -8/14 & 2/14 \\ 2/14 & 4/14 & -1/14 \\ -6/14 & 2/14 & 3/14 \end{bmatrix}} = \overset{\mathbf{I}}{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}}$$

The problem, of course, is to find $\mathbf{A}^{-1}$ when it exists.[3] As mentioned above, if $\mathbf{A}^{-1}$ exists, it will be unique for a given **A**.

4.3.1 The Determinant and the Adjoint of a Matrix

In Chapter 2 we also discussed the concept of the determinant of a (square) matrix. The determinant of **A**, denoted as $|\mathbf{A}|$, is merely a scalar or number that is computed in a certain way. We also discussed cofactors of a matrix and defined them to be determinants of order $n-1$ by $n-1$ obtained from a matrix **A** of order n by n by omitting the ith row and $j$the column of **A** and affixing the sign $(-1)^{i+j}$ to the determinant of the $n-1$ by $n-1$ submatrix. We further recall that a matrix **A** will have as many cofactors as there are entries in **A**.

The cofactors themselves can be, in turn, transformed in a way that possesses some special characteristics relative to **A** and the identity matrix. This matrix, called the adjoint of **A**, possesses the useful property that[4]

$$\boxed{\mathbf{A}\left[\frac{\operatorname{adj}(\mathbf{A})}{|\mathbf{A}|}\right] = \mathbf{I}}$$

We now need to define the adjoint of **A** in terms of the cofactors of **A**.

The adjoint of a (square) matrix **A**, *denoted as* adj(**A**), *is defined as the transpose of the matrix of cofactors obtained from* **A**.

[3] If each column of $\mathbf{A}^{-1}$ is considered originally as a set of unknowns, then all that is involved is solving a set of linear equations. For example,

$$\begin{aligned} 1x_1 + 2x_2 + 0x_3 &= 1 \\ 0x_1 + 3x_2 + 1x_3 &= 0 \\ 2x_1 + 2x_2 + 4x_3 &= 0 \end{aligned}$$

has the solution $x_1 = 10/14$, $x_2 = 2/14$, $x_3 = -6/14$, which is the first column of $\mathbf{A}^{-1}$. Similar procedures lead to the second and third columns of $\mathbf{A}^{-1}$. That is, $\mathbf{A}^{-1}$ could be found by solving three sets of linear equations each, in which the right-hand side of the equation is, respectively, the column vectors

$$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \text{and} \quad \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Details of the method appear in Section 4.7.5.

[4] As the reader might surmise, this property is also displayed by $\mathbf{A}^{-1}$; as will be shown, the inverse $\mathbf{A}^{-1}$ *does* equal the adjoint of **A** divided by the determinant of **A**.

In the simple case of a 2 x 2 matrix

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

we recall that the determinant $|\mathbf{A}|$ is

$$a_{11}a_{22} - a_{12}a_{21}$$

and the cofactors—consisting of single elements—are

$$A_{11} = a_{22}; \qquad A_{12} = -a_{21}; \qquad A_{21} = -a_{12}; \qquad A_{22} = a_{11}$$

Let us now place these cofactors in a square (2 x 2) matrix. Next, let us take the transpose of this matrix.

Then, the adjoint of **A**, in the 2 x 2 case, is just

$$\boxed{\operatorname{adj}(\mathbf{A}) = \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}}$$

In words, this says that in the 2 x 2 case, the adjoint of the matrix **A** involves (a) switching the entries along the main diagonal of the original matrix and (b) changing signs of the off-diagonal entries.

If the $n \times n$ matrix **A** is of higher order than 2 x 2, computation of adj(**A**) is a bit more complicated but proceeds in the same manner as stated above:

1. Find the minors of **A** via the procedure described in Chapter 2.
2. Find the cofactors, or signed minors of **A**, again via the procedure described in Chapter 2.
3. Place these cofactors in an $n \times n$ matrix.
4. Find the transpose of this matrix and call this transpose adj(**A**).

From here it is but a short step to finding the inverse $\mathbf{A}^{-1}$

4.3.2 The Matrix Inverse

Both the determinant and adjoint of **A** figure prominently in the computation of its inverse. *If the square matrix* **A** *has an inverse* $\mathbf{A}^{-1}$, *this inverse, defined such that* $\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$, *can be computed from*

$$\boxed{\mathbf{A}^{-1} = \frac{1}{|\mathbf{A}|} \operatorname{adj}(\mathbf{A}); \qquad |\mathbf{A}| \neq 0}$$

That is, the inverse is found by scalar multiplication of a matrix. In the simple 2 x 2 case described above, this is

$$\mathbf{A}^{-1} = \frac{1}{|\mathbf{A}|}\begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}; \qquad |\mathbf{A}| \neq 0$$

We now have an operational way, based on the transposed matrix of cofactors divided by the determinant of the matrix, to find the inverse of a square matrix **A**. We next consider this problem in the earlier context of point transformations.

4.3.3 Applying the Concept of an Inverse to Point Transformations

Suppose we now return to the problem of finding the preimage vector **x**, given the image vector **x*** and the transformation matrix

$$\mathbf{T} = \begin{bmatrix} 0.90 & 0.44 \\ 0.60 & 0.80 \end{bmatrix}$$

As recalled from Section 4.2.3, $\mathbf{x}^* = \left[\begin{smallmatrix} 1.78 \\ 2.20 \end{smallmatrix}\right]$ is shown in Fig. 4.7. Now we wish to find **x** which, of course, we already know to be $\mathbf{x} = \left[\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right]$. The starting equation is $\mathbf{x}^* = \mathbf{T}\mathbf{x}$, and we wish to solve for **x**.

By taking advantage of the fact that $\mathbf{T}^{-1}\mathbf{T} = \mathbf{I}$, we can find **x** from a matrix transformation that premultiplies both sides of $\mathbf{x}^* = \mathbf{T}\mathbf{x}$ by $\mathbf{T}^{-1}$. Thus

$$\mathbf{T}^{-1}\mathbf{x}^* = \mathbf{T}^{-1}\mathbf{T}\mathbf{x} = \mathbf{I}\mathbf{x} = \mathbf{x}$$

In terms of the specific problem of interest, we need to perform the following calculations. First, we obtain the determinant of **T**:

$$|\mathbf{T}| = 0.9(0.8) - 0.6(0.44) = 0.46$$

Based on the simple definition of the adjoint in the 2 x 2 case, we find

$$\operatorname{adj}(\mathbf{T}) = \begin{bmatrix} t_{22} & -t_{12} \\ -t_{21} & t_{11} \end{bmatrix} = \begin{bmatrix} 0.80 & -0.44 \\ -0.60 & 0.90 \end{bmatrix}$$

Having found both $|\mathbf{T}|$ and adj(**T**), we compute $\mathbf{T}^{-1}$ as follows:

$$\mathbf{T}^{-1} = \frac{1}{0.46}\begin{bmatrix} 0.80 & -0.44 \\ -0.60 & 0.90 \end{bmatrix} = \begin{bmatrix} 1.74 & -0.96 \\ -1.30 & 1.96 \end{bmatrix}$$

It now remains to show that

$$\mathbf{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \underset{\mathbf{T}^{-1}}{\begin{bmatrix} 1.74 & -0.96 \\ -1.30 & 1.96 \end{bmatrix}} \underset{\mathbf{x}^*}{\begin{bmatrix} 1.78 \\ 2.20 \end{bmatrix}}$$

which is, indeed, the case.

We can also verify that, within rounding error, $\mathbf{T}\mathbf{T}^{-1} = \mathbf{T}^{-1}\mathbf{T} = \mathbf{I}$. Finally, we should state that if $|\mathbf{T}|$ is zero, then $1/|\mathbf{T}|$ is not defined, and the (regular) inverse of **T** does not exist. In this case the matrix **T** is said to be *singular*. Otherwise, as is the case here, it is called nonsingular.

A nonsingular matrix **A**, *then, is one in which*

$$\boxed{|\mathbf{A}| \neq 0}$$

Nonsingularity is very important to the topic of matrix inversion since every nonsingular matrix has an inverse; moreover, only nonsingular matrices have (regular) inverses.

Now that we have found out how to compute a matrix inverse and solve the equation

$$\mathbf{x} = \mathbf{T}^{-1}\mathbf{x}^*$$

we should also state a property involving the inverse of the product of two (or more) matrices.

Given the product of two or more conformable matrices, $\mathbf{T}_1 \mathbf{T}_2 \cdots \mathbf{T}_s$, *the inverse of that product equals the product of the separate inverses in reverse order:*

$$\boxed{(\mathbf{T}_1\mathbf{T}_2 \cdots \mathbf{T}_s)^{-1} = \mathbf{T}_s^{-1} \cdots \mathbf{T}_2^{-1}\,\mathbf{T}_1^{-1}}$$

Notice that this property is similar to the property involving the transpose of the product of two or more matrices.

Having discussed some introductory aspects of matrix inversion, we return to the topic of vector transformation, but now in the context of changing basis vectors for the case of general linear transformations. As it turns out, the concept of matrix inverse is also needed here.

4.3.4 Transformation by Basis Vector Changes

As recalled in our earlier discussion of basis vector changes in the context of orthogonal transformations, a second way to examine transformations is in terms of referring some vector $\mathbf{x}$ to a new set of basis vectors $\mathbf{f}_i$. Let us return to the discussion involving transformations using the matrix $\mathbf{T}$.

$$\mathbf{T} = \begin{bmatrix} 0.9 & 0.44 \\ 0.6 & 0.8 \end{bmatrix}$$

Our interest now centers on the case of transformation via changed basis vectors where, as we know, $\mathbf{T}$ is *not* orthogonal. To find the new basis vectors $\mathbf{f}_i$ in the current problem of interest, we make note of the fact that linear combinations of the standard basis vectors $\mathbf{e}_1$ and $\mathbf{e}_2$ are found by

$$\mathbf{f}_1 = \begin{bmatrix} 0.90 \\ 0.44 \end{bmatrix} = 0.9\begin{bmatrix} 1 \\ 0 \end{bmatrix} + 0.44\begin{bmatrix} 0 \\ 1 \end{bmatrix}; \qquad \mathbf{f}_2 = \begin{bmatrix} 0.6 \\ 0.8 \end{bmatrix} = 0.6\begin{bmatrix} 1 \\ 0 \end{bmatrix} + 0.8\begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

To express this in matrix form, where the $\mathbf{f}_i$ also appear as column vectors, we have

$$\mathbf{F} = \begin{bmatrix} 0.90 & 0.60 \\ 0.44 & 0.80 \end{bmatrix} = \overset{\mathbf{E}}{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}} \overset{\mathbf{T}'}{\begin{bmatrix} 0.90 & 0.60 \\ 0.44 & 0.80 \end{bmatrix}}$$

Notice, then, that the new basis vectors are given by the transpose of the matrix $\mathbf{T}$, where $\mathbf{T}$ itself was used in the point transformation in which $\mathbf{x}$ moved relative to the fixed $\mathbf{e}_i$ basis.

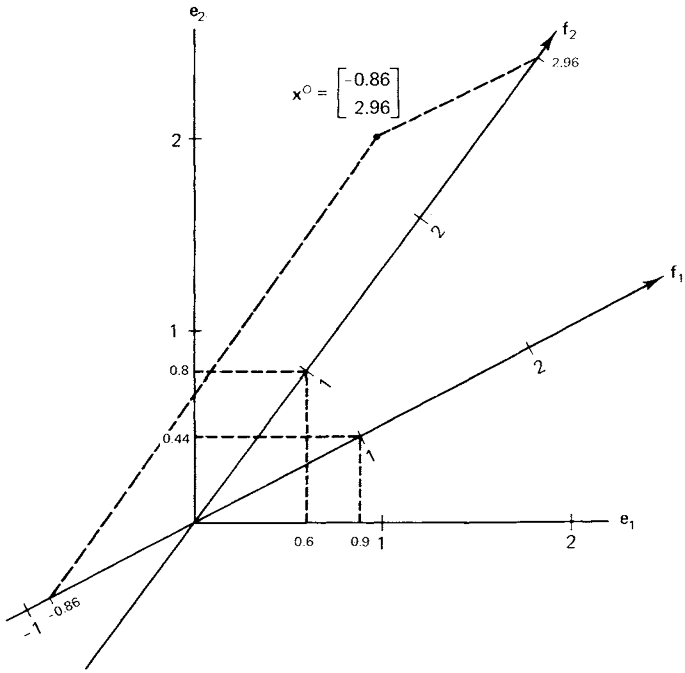

Fig. 4.8 Transformation of x via change in basis vectors.

Fig. 4.8 shows the new basis vectors $\mathbf{f}_1$ and $\mathbf{f}_2$ plotted in terms of the old basis. We note from the figure that the new basis vectors $\mathbf{f}_1$, $\mathbf{f}_2$ are oblique. Our problem now is to refer the original $\mathbf{x} = [\begin{smallmatrix}1\\2\end{smallmatrix}]$, as portrayed in **E**, to the new (oblique) basis in **F**. As we know, the point transformation **T** yields the point $\mathbf{x}^* = [\begin{smallmatrix}1.78\\2.20\end{smallmatrix}]$, shown in the original $\mathbf{e}_i$ basis of Fig. 4.7.

But we want to find the coordinates of the points, as referred to **F**. To do this we note that a point can either be expressed as **x** in **E** or as $\mathbf{x}^{\circ}$ in **F**.

$$\mathbf{F}\mathbf{x}^{\circ} = \mathbf{E}\mathbf{x}$$

But, as we know from the basis change shown above,

$$\mathbf{F} = \mathbf{E}\mathbf{T}'$$

Hence, to solve for $\mathbf{x}^{\circ}$ in $\mathbf{F}\mathbf{x}^{\circ} = \mathbf{E}\mathbf{x}$, we can premultiply both sides by $\mathbf{F}^{-1}$ to get

$$\mathbf{F}^{-1}\mathbf{F}\mathbf{x}^{\circ} = \mathbf{F}^{-1}\mathbf{E}\mathbf{x}$$

Noting that $\mathbf{F}^{-1}\mathbf{F} = \mathbf{I}$ on the left-hand side and substituting $\mathbf{E}\mathbf{T}'$ for **F** on the right-hand side, we recall the relationship involving the inverse of the product of two or more matrices shown in the preceding section to get

$$\mathbf{x}^{\circ} = (\mathbf{E}\mathbf{T}')^{-1}\mathbf{E}\mathbf{x} = (\mathbf{T}')^{-1}\mathbf{E}^{-1}\mathbf{E}\mathbf{x}$$

But, $\mathbf{E}^{-1}\mathbf{E} = \mathbf{I}$, so that

$$\boxed{\mathbf{x}^{\circ} = (\mathbf{T}')^{-1}\mathbf{x}}$$

What this all says is that to find $\mathbf{x}^\circ$ in terms of the new basis vectors $\mathbf{f}_i$ we are going to have to find $(\mathbf{T}')^{-1}$, the *inverse* of $\mathbf{T}'$.

We have already found $\mathbf{T}^{-1}$ in the case of moving $\mathbf{x}^* = \left[\begin{smallmatrix}1.78\\2.20\end{smallmatrix}\right]$ via a point transformation back to $\mathbf{x} = \left[\begin{smallmatrix}1\\2\end{smallmatrix}\right]$. We use the same type of matrix inversion procedure to find $(\mathbf{T}')^{-1}$, and this turns out to be

$$(\mathbf{T}')^{-1} = \begin{bmatrix} 1.74 & -1.30 \\ -0.96 & 1.96 \end{bmatrix}$$

Having found $(\mathbf{T}')^{-1}$, we then solve for $\mathbf{x}^\circ$ as

$$\mathbf{x}^\circ = \begin{bmatrix} -0.86 \\ 2.96 \end{bmatrix} = \begin{bmatrix} 1.74 & -1.30 \\ -0.96 & 1.96 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

Figure 4.8 shows the transformed coordinates, $\mathbf{x}^\circ = \left[\begin{smallmatrix}-0.86\\2.96\end{smallmatrix}\right]$ of the original and fixed point $\mathbf{x}$, now expressed in terms of the new (oblique) basis vectors $\mathbf{f}_i$.

How does this result relate to our earlier discussion involving basis vector changes that are given by orthogonal matrices? *In this case it turns out that for orthogonal matrices the following relationship holds:*

$$\boxed{\mathbf{A}^{-1} = \mathbf{A}'}$$

Hence, in the special case of orthogonal matrices we find the relationship

$$\mathbf{x}^\circ = (\mathbf{A}')^{-1}\,\mathbf{x} = \mathbf{A}\mathbf{x}$$

in solving the general equation

$$\mathbf{F}\mathbf{x}^\circ = \mathbf{E}\mathbf{x}$$

for $\mathbf{x}^\circ$. This, of course, is what we expect to find in the case of rotations. In other cases involving basis vector changes that employ general linear transformations such as $\mathbf{T}$, the appropriate expression, as we now know, is

$$\mathbf{x}^\circ = (\mathbf{T}')^{-1}\,\mathbf{x}$$

So, in summary, we can recapitulate the following points:

1. For point transformations, the simpler type of transformation, we have

   a. $\mathbf{x}^* = \mathbf{T}\mathbf{x}$ with $\mathbf{x}^*$ and $\mathbf{x}$ in terms of $\mathbf{E}$

   b. $\mathbf{x} = \mathbf{T}^{-1}\mathbf{x}^*$ with $\mathbf{x}$ and $\mathbf{x}^*$ in terms of $\mathbf{E}$

2. For transformations involving basis vector changes, the new basis vectors are given in terms of the old by $\mathbf{F} = \mathbf{E}\mathbf{T}'$. We then have the cases

   a. $\mathbf{x}^\circ = (\mathbf{T}')^{-1}\mathbf{x}$ with $\mathbf{x}^\circ$ in terms of $\mathbf{F}$ and $\mathbf{x}$ in terms of $\mathbf{E}$

   b. $\mathbf{x} = \mathbf{T}'\mathbf{x}^\circ$ with $\mathbf{x}$ in terms of $\mathbf{E}$ and $\mathbf{x}^\circ$ in terms of $\mathbf{F}$

Finally, in the special case of orthogonal matrices, denoted by $\mathbf{A}$, we have

1. For point transformations:

   a. $\mathbf{x}^* = \mathbf{A}\mathbf{x}$ with $\mathbf{x}^*$ and $\mathbf{x}$ in terms of $\mathbf{E}$

   b. $\mathbf{x} = \mathbf{A}'\mathbf{x}^*$ with $\mathbf{x}$ and $\mathbf{x}^*$ in terms of $\mathbf{E}$

2. For transformations involving basis vector changes, the new basis vectors are still given by the matrix $\mathbf{F} = \mathbf{E}\mathbf{A}'$. We then have the special cases:

   a. $\mathbf{x}^\circ = \mathbf{A}\mathbf{x}$ with $\mathbf{x}^\circ$ in terms of $\mathbf{F}$ and $\mathbf{x}$ in terms of $\mathbf{E}$

   b. $\mathbf{x} = \mathbf{A}'\mathbf{x}^\circ$ with $\mathbf{x}$ in terms of $\mathbf{E}$ and $\mathbf{x}^\circ$ in terms of $\mathbf{F}$

As the reader has probably gathered by now, point transformation, in which the basis vectors remain fixed, is considerably easier to follow intuitively. However, instances arise in multivariate analyses where the selection of an appropriate basis to which the vectors can be referred results in a significant degree of simplification in characterizing the nature of the transformation that the researcher is employing.

For this reason, we carry our analysis one more step, albeit the most complex one so far. We can pose the problem as one of starting with a point transformation of $\mathbf{x}$, as given by the matrix $\mathbf{T}$, relative to the standard basis vectors $\mathbf{e}_i$.

However, suppose the $\mathbf{e}_i$ basis is related, in turn, to a new basis $\mathbf{f}_i$ via a matrix $\mathbf{L}$. If so, how is the point transformation given by $\mathbf{T}$ in the context of $\mathbf{e}_i$ represented in the new basis $\mathbf{f}_i$? We shall call the new transformation matrix $\mathbf{T}^\circ$ to take into consideration the fact that it transforms points in the new basis $\mathbf{f}_i$.

The practical import of all of this is that in applied multivariate problems we often seek special sets of basis vectors in which the transformation matrix $\mathbf{T}^\circ$ displays a particularly simple character. This pragmatic aspect of basis vector transformations is deferred until Chapter 5. However, here we can at least go through the mechanics of the process of relating a transformation represented by $\mathbf{T}$ in the standard basis $\mathbf{e}_i$ to its counterpart $\mathbf{T}^\circ$ in the derived basis $\mathbf{f}_i$.

4.3.5 Transformations under Arbitraiy Changes of Basis Vectors

Suppose we continue to consider the matrix

$$\mathbf{T} = \begin{bmatrix} 0.90 & 0.44 \\ 0.60 & 0.80 \end{bmatrix}$$

but let us also consider a second matrix

$$\mathbf{L} = \begin{bmatrix} 0.83 & 0.55 \\ 0.20 & 0.98 \end{bmatrix}$$

that can be used to transform the vectors in the standard basis $\mathbf{e}_i$ to a new basis $\mathbf{f}_i$. Note that the sums of squares of the row elements of $\mathbf{L}$ equal unity. Hence, the rows of $\mathbf{L}$ can be considered as direction cosines. However, since the scalar product of row 1 with row 2 does not equal zero, the new $\mathbf{f}_i$ basis is oblique.

As before, our first job is to find the oblique basis $\mathbf{f}_i$ in terms of $\mathbf{e}_i$, the standard basis. Again, we define the column vectors $\mathbf{f}_i$ as follows:

$$\mathbf{f}_1 = 0.83\mathbf{e}_1 + 0.55\mathbf{e}_2; \qquad \mathbf{f}_2 = 0.20\mathbf{e}_1 + 0.98\mathbf{e}_2$$

which, in matrix multiplication form, are obtained as column vectors from

$$\mathbf{F} = \begin{bmatrix} 0.83 & 0.20 \\ 0.55 & 0.98 \end{bmatrix} = \overset{\mathbf{E}}{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}} \overset{\mathbf{L}'}{\begin{bmatrix} 0.83 & 0.20 \\ 0.55 & 0.98 \end{bmatrix}}$$

where, of course, we have the linear combinations

$$\mathbf{f}_1 = 0.83\begin{bmatrix} 1 \\ 0 \end{bmatrix} + 0.55\begin{bmatrix} 0 \\ 1 \end{bmatrix}; \qquad \mathbf{f}_2 = 0.20\begin{bmatrix} 1 \\ 0 \end{bmatrix} + 0.98\begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Panel II of Fig. 4.9 shows the oblique basis vectors $\mathbf{f}_1$ and $\mathbf{f}_2$ plotted in terms of the original basis $\mathbf{e}_i$.

From earlier discussion we know how to find the point transformation:

$$\mathbf{x}^* = \mathbf{T}\mathbf{x}$$

From Fig. 4.7 we note that $\mathbf{x}^* = \left[\begin{smallmatrix} 1.78 \\ 2.20 \end{smallmatrix}\right]$. For ease of comparison this reappears in Panel I of Fig. 4 9 relative to the original basis $\mathbf{E}$.

Our objective now is to find the counterpart point transformation of $\mathbf{x}^\circ = \left[\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right]$, referred now to the oblique basis $\mathbf{f}_i$ in Panel II of Fig. 4.9.

Let us first present the solution to this problem and then examine it, piece by piece. First, we have the relationship

$$\boxed{\mathbf{x}^{*\circ} = \mathbf{T}^\circ\mathbf{x}^\circ}$$

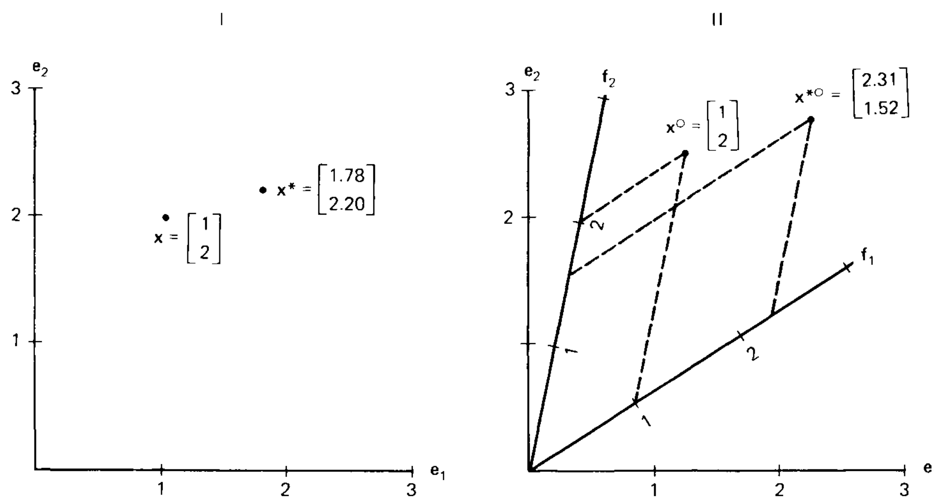

Fig. 4.9 Point transformations relative to different bases.

but $\mathbf{T}^\circ$ is found from the product

$$\boxed{\mathbf{T}^\circ = (\mathbf{L}')^{-1}\mathbf{T}\mathbf{L}'}$$

Since we already have $\mathbf{L}'$ and $\mathbf{T}$ above, the problem now involves the computation of $(\mathbf{L}')^{-1}$. The determinant of $\mathbf{L}'$ is

$$|\mathbf{L}'| = (0.83 \times 0.98) - (0.20 \times 0.55) = 0.703$$

We then need to find adj($\mathbf{L}'$) and, finally, $(\mathbf{L}')^{-1}$. This is carried out as follows:

$$(\mathbf{L}')^{-1} = \frac{1}{0.703} \underbrace{\begin{bmatrix} 0.98 & -0.20 \\ -0.55 & 0.83 \end{bmatrix}}_{\text{adj}(\mathbf{L}')} = \begin{bmatrix} 1.39 & -0.28 \\ -0.78 & 1.18 \end{bmatrix}$$

Hence

$$\mathbf{T}^\circ = \underset{(\mathbf{L}')^{-1}}{\begin{bmatrix} 1.39 & -0.28 \\ -0.78 & 1.18 \end{bmatrix}} \underset{\mathbf{T}}{\begin{bmatrix} 0.90 & 0.44 \\ 0.60 & 0.80 \end{bmatrix}} \underset{\mathbf{L}'}{\begin{bmatrix} 0.83 & 0.20 \\ 0.55 & 0.98 \end{bmatrix}} = \begin{bmatrix} 1.11 & 0.60 \\ 0.34 & 0.59 \end{bmatrix}$$

Returning to the vector $\mathbf{x} = \left[\begin{smallmatrix}1\\2\end{smallmatrix}\right]$, we recall under the fixed basis vectors $\mathbf{e}_1$, $\mathbf{e}_2$, that $\mathbf{x}$ is transformed by $\mathbf{T}$ onto $\mathbf{x}^* = \left[\begin{smallmatrix}1.78\\2.20\end{smallmatrix}\right]$, as noted in Panel I of Fig. 4.9. However, if we use the basis $\mathbf{f}_1$, $\mathbf{f}_2$, the point transformation of $\mathbf{x}^\circ = \left[\begin{smallmatrix}1\\2\end{smallmatrix}\right]$ in the basis $\mathbf{F}$ utilizes $\mathbf{T}^\circ$ and is

$$\mathbf{x}^{*\circ} = \begin{bmatrix} 2.31 \\ 1.52 \end{bmatrix} = \begin{bmatrix} 1.11 & 0.60 \\ 0.34 & 0.59 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

Notice, then, that $\mathbf{x}^{*\circ}$ is a *point* transformation of $\mathbf{x}^\circ = \left[\begin{smallmatrix}1\\2\end{smallmatrix}\right]$ in the $\mathbf{F}$ basis.

Panel II of Fig. 4.9 shows the nature of the transformation. First, we plot $\mathbf{x}^\circ = \left[\begin{smallmatrix}1\\2\end{smallmatrix}\right]$ in terms of the oblique basis $\mathbf{F}$. Note that $\mathbf{f}_1 = \left[\begin{smallmatrix}0.83\\0.55\end{smallmatrix}\right]$ and $\mathbf{f}_2 = \left[\begin{smallmatrix}0.20\\0.98\end{smallmatrix}\right]$ are, in turn, plotted in terms of the original $\mathbf{E}$ basis, while $\mathbf{x}^\circ = \left[\begin{smallmatrix}1\\2\end{smallmatrix}\right]$ is positioned with respect to the new basis $\mathbf{F}$. The process of finding the point transformation $\mathbf{x}^{*\circ} = \left[\begin{smallmatrix}2.31\\1.52\end{smallmatrix}\right]$ in $\mathbf{F}$ proceeds by decomposing

$$\mathbf{T}^\circ = (\mathbf{L}')^{-1}\mathbf{T}\mathbf{L}'$$

as follows, starting from the far right of the expression to the right of the equals sign:

1. $\mathbf{L}'$ maps $\mathbf{x}^\circ$ onto $\mathbf{x}$; that is,

$$\underset{\mathbf{L}'}{\begin{bmatrix} 0.83 & 0.20 \\ 0.55 & 0.98 \end{bmatrix}} \underset{\mathbf{x}^\circ}{\begin{bmatrix} 1 \\ 2 \end{bmatrix}} = \underset{\mathbf{x}}{\begin{bmatrix} 1.23 \\ 2.51 \end{bmatrix}}$$

2. $\mathbf{T}$ maps $\mathbf{x} = \left[\begin{smallmatrix}1.23\\2.51\end{smallmatrix}\right]$ onto $\mathbf{x}^*$

$$\underset{\mathbf{T}}{\begin{bmatrix} 0.90 & 0.44 \\ 0.60 & 0.80 \end{bmatrix}} \underset{\mathbf{x}}{\begin{bmatrix} 1.23 \\ 2.51 \end{bmatrix}} = \underset{\mathbf{x}^*}{\begin{bmatrix} 2.21 \\ 2.75 \end{bmatrix}}$$

3. $(\mathbf{L}')^{-1}$ maps $\mathbf{x}^*$ onto $\mathbf{x}^{*\circ}$

$$\begin{array}{ccc} (\mathbf{L}')^{-1} & \mathbf{x}^* & = \mathbf{x}^{*\circ} \\ \begin{bmatrix} 1.39 & -0.28 \\ -0.78 & 1.18 \end{bmatrix} & \begin{bmatrix} 2.21 \\ 2.75 \end{bmatrix} & = \begin{bmatrix} 2.31 \\ 1.52 \end{bmatrix} \end{array}$$

We can follow through each of these steps from Panel II of Fig. 4.9. First, in step one we note that the coordinates of $\mathbf{x}^\circ$ after mapping onto $\mathbf{x}$ are, indeed, $\mathbf{x} = \left[\begin{smallmatrix}1.23\\2.51\end{smallmatrix}\right]$ with respect to the $\mathbf{E}$ basis. In step two $\mathbf{x}$ is mapped onto $\mathbf{x}^* = \left[\begin{smallmatrix}2.21\\2.75\end{smallmatrix}\right]$, again with respect to the $\mathbf{E}$ basis. However, to refer the point to the $\mathbf{F}$ basis we employ step three, giving us $\mathbf{x}^{*\circ} = \left[\begin{smallmatrix}2.31\\1.52\end{smallmatrix}\right]$ with respect to the oblique basis $\mathbf{F}$.

As will be pointed out in Chapter 5, the practical matter is to find a suitable basis $\mathbf{F}$ such that the matrix of the transformation with respect to this new basis takes on a particularly simple form, such as a diagonal matrix. Finally, it is worth noting that if $\mathbf{L}$ is orthogonal we have the simplification

$$\mathbf{T}^\circ = \mathbf{LTL}'$$

since, as noted earlier, if $\mathbf{L}$ is orthogonal, then

$$(\mathbf{L}')^{-1} = \mathbf{L}$$

4.3.6 Recapitulation

The concept of vector transformation is central to matrix algebra and multivariate analysis. The simplest type of transformation is represented by a point transformation, relative to a fixed basis:

$$\mathbf{x}^* = \mathbf{Tx}$$

Usually, we choose the fixed basis to be $\mathbf{E}$, the standard basis. Then the axes are mutually orthogonal and of unit length. In point transformations $\mathbf{x}$ moves according to $\mathbf{T}$ while the basis stays fixed. Figure 4.7 shows the geometric character of this type of transformation.

Alternatively, we can allow the point to stay fixed and, instead, transform the basis vectors to some new, and possibly oblique, orientation. This type of transformation, called a basis vector transformation, is exemplified by

$$\mathbf{x}^\circ = (\mathbf{T}')^{-1}\,\mathbf{x}$$

and is illustrated in Fig. 4.8.

As also pointed out, if $\mathbf{T}$ is orthogonal, various simplifications result that make the geometric interpretation easier. Finally, the specific character of some particular mapping, denoted generally by τ, depends on the reference basis. We showed how one matrix of the transformation, represented by $\mathbf{T}$ with respect to $\mathbf{e}_i$, the original basis, can be represented by $\mathbf{T}^\circ$ if we know the transformation that connects $\mathbf{f}_i$, the basis for $\mathbf{T}^\circ$ with $\mathbf{e}_i$, the basis for $\mathbf{T}$.

While not illustrated in the cases that were covered, it should be mentioned in passing that $|\mathbf{T}| = |\mathbf{T}^\circ|$. That is, the determinant of a linear transformation is independent of the basis to which the transformation is referred.

Inasmuch as matrix transformations are so central to the subject, we continue our discussion of the geometric character of various types of special matrices. While we have described transformations represented by orthogonal matrices (i.e., rotations), it turns out that many other kinds of matrices have intuitively simple geometric representations as well.

4.4 GEOMETRIC RELATIONSHIPS INVOLVING MATRIX TRANSFORMATIONS

As we have illustrated, many matrix operations can be usefully represented geometrically if two or three dimensions are involved. At this point it seems useful to extend this line of geometric reasoning to other aspects of matrix transformations. In so doing the reader may get some intuitive understanding of what is involved in higher dimensionalities where geometrical representation is no longer feasible.

In order to motivate the discussion, let us consider the small 9 x 2 matrix of synthetic data shown in Table 4.1. This matrix consists of nine points, positioned in two dimensions, as diagrammed in Panel I of Fig. 4.10. Note that the points involve a square lattice arrangement.

TABLE 4.1

Original Matrix **X** *of Nine Points in Two Dimensions*

Point (code letter)	Dimension 1	Dimension 2
a	1	0
b	1	1
c	1	2
d	2	0
e	2	1
f	2	2
g	3	0
h	3	1
i	3	2
	—	—
$\bar{X}_j$	2	1

We shall now describe a variety of operations on this synthetic data matrix and show their effects geometrically. Whereas our earlier discussions of matrix transformations involved transforming a single vector **x** into another vector **x***, here we shall transform a *set* of vectors, which can be represented as a matrix **X**. The basic principles remain the same.

In all examples of this section we deal with the simpler case of point transformations, rather than basis vector transformations. However, so as to show the flexibility of matrix transformations, in this section we *postmultiply* the original matrix **X** by the transformation matrix, in order to obtain **X***. No new principles are involved in this

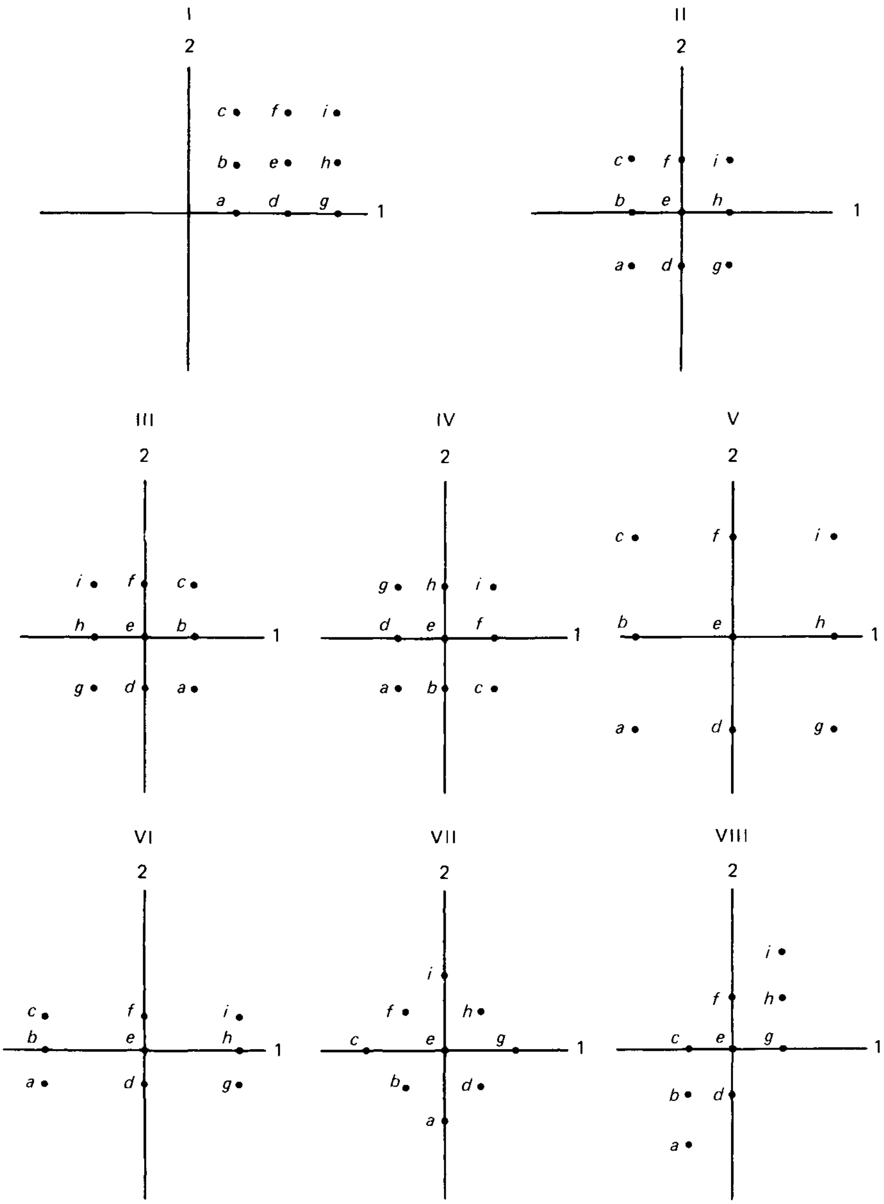

Fig. 4.10 Geometric effect of various simple transformations. Key: I, original data; II, translation; III, reflection; IV, permutation; V, central dilation; VI, stretch; VII, counterclockwise rotation of 45° ; VIII, shear.

change, as the reader will note in due course. Thus, the basic format to be followed consists of the transformation

$$\boxed{\mathbf{X}^* = \mathbf{XA}}$$

where **A** assumes various special forms that exhibit simple geometric patterns. These special forms are of particular relevance to multivariate analysis.

Notice, however, that each *row* vector in **X** is being postmultiplied by **A** to obtain each row vector in $\mathbf{X}^*$. As we shall see later, this viewpoint modifies the specific entries of the matrices, although all basic concepts remain unchanged from our earlier discussion in which **A** premultiplied column vectors.

4.4.1 Translation

Matrix translation has to do with the problem of relating a set of points to a particular *origin* in the space. Suppose, for example, that we wished to refer the nine points of Panel I of Fig. 4.10 to a centroid-centered origin. By "centroid" we mean, of course, the arithmetic mean of each set of coordinates on each dimension. As shown in Table 4.1, the mean of the points on the first dimension is 2; their mean on the second dimension is 1. By expressing each original value as a *deviation* from its mean, we arrive at matrix $\mathbf{X}_d$ as shown in Table 4.2. Note that this simply involves a subtraction of $\bar{\mathbf{X}}$, whose entries represent the mean of the points on each dimension, from the original matrix **X**. Panel II of Fig. 4.10 shows the effect of the translation geometrically.

TABLE 4.2

Translation of Matrix **X** *to Origin at Centroid*

$$\begin{array}{ccccc} \mathbf{X} & & \bar{\mathbf{X}} & & \mathbf{X}_d \\ \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 2 & 0 \\ 2 & 1 \\ 2 & 2 \\ 3 & 0 \\ 3 & 1 \\ 3 & 2 \end{bmatrix} & - & \begin{bmatrix} 2 & 1 \\ 2 & 1 \\ 2 & 1 \\ 2 & 1 \\ 2 & 1 \\ 2 & 1 \\ 2 & 1 \\ 2 & 1 \\ 2 & 1 \end{bmatrix} & = & \begin{bmatrix} -1 & -1 \\ -1 & 0 \\ -1 & 1 \\ 0 & -1 \\ 0 & 0 \\ 0 & 1 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix} \end{array}$$

A translation, then, involves a parallel displacement of every point to some new origin of interest. In this case the centroid of the points is the origin of interest. The particular nature of the matrix **X**, consisting of the column means of **X**, is

$$\boxed{\bar{\mathbf{X}} = 1/m\mathbf{1}\mathbf{1}'\mathbf{X}}$$

where **1** is the unit column vector and $m = 9$. We can then find $\mathbf{X}_d$ as

$$\boxed{\begin{aligned} \mathbf{X}_d &= \mathbf{X} - 1/m\mathbf{1}\mathbf{1}'\mathbf{X} \\ &= (\mathbf{I} - 1/m\mathbf{1}\mathbf{1}')\mathbf{X} \end{aligned}}$$

In previous sections of the chapter, all of our discussion of matrix transformations, involved a *multiplicative* form, such as

$$\mathbf{x}^* = \mathbf{A}\mathbf{x}$$

or, in the present format,

$$\mathbf{X}^* = \mathbf{X}\mathbf{A}$$

In translating a set of points in two dimensions, denoted by the matrix $\mathbf{X}$, we see that for each transformed point we have the coordinates

$$x_{i1}^* = x_{i1} + h; \qquad x_{i2}^* = x_{i2} + k$$

where h and k are constants. Hence, translation departs from the usual matrix multiplication format by involving the sum or difference of two matrices.

However, by using the following device:

$$(x_{i1}^*, x_{i2}^*, 1) = (x_{i1}, x_{i2}, 1)\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ h & k & 1 \end{bmatrix}$$

we can obtain

$$x_{i1}^* = x_{i1} + h; \qquad x_{i2}^* = x_{i2} + k; \qquad 1 = 1$$

Note that the last equation $(1 = 1)$ is trivial, but does enable us to express a translation in the multiplicative format used earlier.[5]

Translations are frequently used in multivariate analysis. In particular, the SSCP, covariance, and correlation matrices all utilize mean-corrected scores and involve, among other things, a translation of raw scores into deviation scores.

4.4.2 Reflection

Reflection of a set of points, as noted in the discussion of improper rotation in Chapter 3, entails multiplication of the coordinates of each point to be reflected by -1. For example, suppose we wished to reflect the nine points of Panel II in Fig. 4.10 "across" axis 2. This can be accomplished by multiplying each of the coordinates on axis 1 by a -1, as shown in Table 4.3 and illustrated graphically in Panel III of Fig. 4.10.

The matrix used for this purpose is represented by

$$\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$$

Similarly, if desired, one could reflect the nine points across axis 1. It should be reiterated, however, that reflection typically involves an odd number of dimensions. If we

[5] This particular computational trick can be useful in the preparation of computer routines for translating the origin of a set of points.

TABLE 4.3

Reflection of Matrix $\mathbf{X}_d$ *across Axis 2*

$$\overset{\mathbf{X}_d}{\begin{bmatrix} -1 & -1 \\ -1 & 0 \\ -1 & 1 \\ 0 & -1 \\ 0 & 0 \\ 0 & 1 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix}} \overset{\text{Reflection}}{\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}} = \begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 0 & -1 \\ 0 & 0 \\ 0 & 1 \\ -1 & -1 \\ -1 & 0 \\ -1 & 1 \end{bmatrix}$$

were to reflect the points across *both* axis 1 and axis 2, the overall effect—called a reflection through the origin—is equivalent to a rotation of the points about an angle of 180° from their original orientation.

As discussed earlier, rotation followed by reflection is often referrred to as an "improper" rotation in the case of orthogonal matrices whose entries consist of direction cosines. It is relevant to point out that improper rotation satisfies the same conditions as proper rotation:

1. All transformation vectors are of unit length.
2. All transformation vectors are mutually orthogonal.

The differentiating feature of improper rotation is that the determinant of this type of orthogonal matrix is -1, while the determinant of an orthogonal matrix constituting a proper rotation is $+1$.

Perhaps all of this can be summarized by saying:

1. A reflection can always be described as a proper rotation followed by reflection of *one* dimension.
2. Reflection of an even number of dimensions (e.g., two, four, six, etc.) is equivalent to a proper rotation in the 1–2 plane, the 3–4 plane, and so on.
3. Reflection of an odd number of dimensions is equivalent to a proper rotation followed by reflection of one dimension.

4.4.3 Axis Permutation

Permutation of a set of points, as the name suggests, involves a matrix transformation that carries each coordinate value on axis 1 into a corresponding coordinate on axis 2, and vice versa. This is illustrated in Table 4.4, and the effect is shown graphically in Panel IV of Fig. 4.10.

The permutation matrix

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

TABLE 4.4

Axis Permutation of Matrix $\mathbf{X}_d$

$$\overset{\mathbf{X}_d}{\begin{bmatrix} -1 & -1 \\ -1 & 0 \\ -1 & 1 \\ 0 & -1 \\ 0 & 0 \\ 0 & 1 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix}} \begin{array}{c}\text{Permutation} \\ \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\end{array} = \begin{bmatrix} -1 & -1 \\ 0 & -1 \\ 1 & -1 \\ -1 & 0 \\ 0 & 0 \\ 1 & 0 \\ -1 & 1 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}$$

TABLE 4.5

Central Dilation of Matrix $\mathbf{X}_d$

$$\overset{\mathbf{X}_d}{\begin{bmatrix} -1 & -1 \\ -1 & 0 \\ -1 & 1 \\ 0 & -1 \\ 0 & 0 \\ 0 & 1 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix}} \begin{array}{c}\text{Central} \\ \text{dilation} \\ \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}\end{array} = \begin{bmatrix} -2 & -2 \\ -2 & 0 \\ -2 & 2 \\ 0 & -2 \\ 0 & 0 \\ 0 & 2 \\ 2 & -2 \\ 2 & 0 \\ 2 & 2 \end{bmatrix}$$

that postmultiplies $\mathbf{X}_d$ produces an interchange of columns. However, if the permutation matrix premultiplies a matrix of preimages, then the rows of the preimages are interchanged.

In Table 4.4, however, we see that the first and second columns of $\mathbf{X}_d$ are interchanged by means of postmultiplication of $\mathbf{X}_d$ by the permutation matrix.

4.4.4 Central Dilation

Central dilation of a set of points entails scalar multiplication of the matrix of coordinates, which is equivalent to multiplication by a scalar matrix; that is, a diagonal matrix in which each diagonal entry involves the same positive constant λ. Central dilation leads to a uniform expansion, if $\lambda > 1$, or a uniform contraction, if $\lambda < 1$, of each dimension. If $\lambda = 1$, then the scalar matrix becomes an identity matrix, and the point positions remain as originally expressed.

Table 4.5 shows application of a central dilation where $\lambda = 2$. Panel V of Fig. 4.10 shows the results graphically. Scalar matrix transformations are particularly simple from a geometric standpoint since we see that *uniform* stretching or compressing of the dimensions takes place along the original axes of orientation. According to the present case, the scalar matrix is

$$\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

where each of the original axes is dilated to twice its original length.

4.4.5 Stretch

A stretch transformation of a set of points involves application of a *diagonal* matrix where, in general, the diagonal entries are such that $\lambda_{ii} \neq \lambda_{jj}$. In contrast to central dilation, a stretch involves *differential* stretching or contraction (rescaling) of points corresponding, again, to directions along the original axes.

TABLE 4.6

A Stretch of Matrix $\mathbf{X}_d$

$$\overset{\mathbf{X}_d}{\begin{bmatrix} -1 & -1 \\ -1 & 0 \\ -1 & 1 \\ 0 & -1 \\ 0 & 0 \\ 0 & 1 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix}} \; \overset{\text{Stretch}}{\begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}} = \begin{bmatrix} -2 & -1 \\ -2 & 0 \\ -2 & 1 \\ 0 & -1 \\ 0 & 0 \\ 0 & 1 \\ 2 & -1 \\ 2 & 0 \\ 2 & 1 \end{bmatrix}$$

For example, if $\lambda_{11} = 2$ and $\lambda_{22} = 1$, the effect of this transformation is to stretch axis 1 to twice its original length, thus producing a latticelike rectangle out of the original latticelike square. Table 4.6 illustrates the computations involved, and Panel VI of Fig. 4.10 shows the graphical results of the transformation

$$\begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$$

Here we shall restrict the term "stretch" to the case in which all $\lambda_{ii} > 0$. If some λ_{ii} were zero, those dimensions would be annihilated. A $\lambda_{ii} < 0$ would correspond to a stretching (or contraction) followed by reflection.

4.4.6 Rotation

As described earlier in the chapter, axis rotation involves application of a rather special kind of matrix—an orthogonal matrix. An orthogonal matrix is distinguished by the properties: (a) the sum of squares of each column (row) equals 1, and (b) the scalar product of each pair of columns (rows) equals zero. To illustrate, the matrix corresponding to a 45° rotation

$$\begin{bmatrix} 0.707 & 0.707 \\ -0.707 & 0.707 \end{bmatrix}$$

meets these conditions, since, by columns, for example,

$$\text{(a)} \quad (0.707)^2 + (-0.707)^2 = 1; \qquad (0.707)^2 + (0.707)^2 = 1$$

$$\text{(b)} \quad (0.707, -0.707) \begin{bmatrix} 0.707 \\ 0.707 \end{bmatrix} = 0$$

As we know, in two dimensions an orthogonal matrix entails a rigid rotation of the original configuration of points about some angle Ψ. In the preceding example we rotate the points counterclockwise about an angle, $\Psi = 45°$, and this involves the following:

$$\begin{bmatrix} \cos \Psi & \sin \Psi \\ -\sin \Psi & \cos \Psi \end{bmatrix}$$

That is, $\cos 45° = \sin 45° = 0.707$, while $-\sin 45° = -0.707$.

TABLE 4.7

Counterclockwise Rotation of Matrix $\mathbf{X}_d$ *through 45° Angle*

$$\overset{\mathbf{X}_d}{\begin{bmatrix} -1 & -1 \\ -1 & 0 \\ -1 & 1 \\ 0 & -1 \\ 0 & 0 \\ 0 & 1 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix}} \overset{\text{Rotation}}{\begin{bmatrix} 0.707 & 0.707 \\ -0.707 & 0.707 \end{bmatrix}} = \begin{bmatrix} 0 & -1.414 \\ -0.707 & -0.707 \\ -1.414 & 0 \\ 0.707 & -0.707 \\ 0 & 0 \\ -0.707 & 0.707 \\ 1.414 & 0 \\ 0.707 & 0.707 \\ 0 & 1.414 \end{bmatrix}$$

The reader should note that the above rotation matrix is postmultiplying each row vector of the matrix $\mathbf{X}_d$. This is in direct contrast to the rotation depicted in Fig. 4.3, in which the matrix of this transformation is *premultiplying* the column vector of coordinate values $\mathbf{x}$. As such, the orthogonal matrix, whose effect is depicted in Fig. 4.3, still represents a counterclockwise point rotation but now has the form

$$\begin{bmatrix} \cos\Psi & -\sin\Psi \\ \sin\Psi & \cos\Psi \end{bmatrix}$$

This is the transpose of the matrix form shown above.

Table 4.7 shows the application of the 45° counterclockwise point rotation, and Panel VII of Fig. 4.10 summarizes the results graphically. Note that the zero point (intersection of the axes) can be viewed as the hub of a wheel and remains fixed during the rotation.

As first discussed in Chapter 3, entries of the rotation matrix can all be expressed as direction cosines of the angles θ_{11}, θ_{12}, θ_{21}, θ_{22} made between old and new axes. The first subscript refers to the old axis, while the second refers to the new axis. And, as stated earlier, this type of generalization is important when more than two dimensions are involved.

A second point of interest is that the same kind of transformation noted above can be made by rotating the axes, rather than the points, around an angle of $-45°$, relative to the original orientation. This alternative view, of course, was covered earlier in the chapter.

4.4.7 Shear

A shear transformation is characterized by the following form:

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

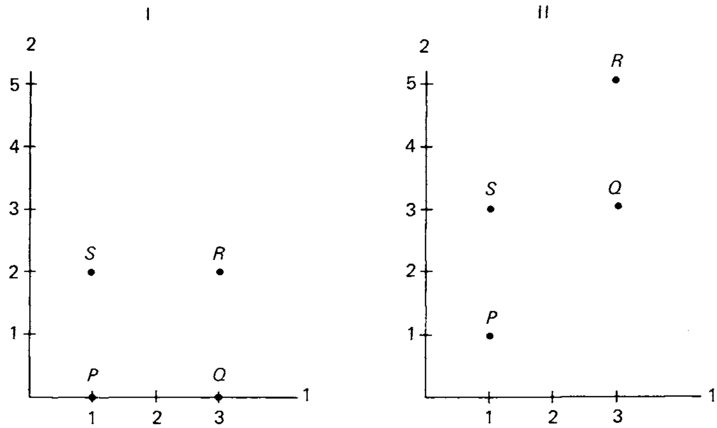

Fig. 4.11 A shear transformation. Key: I, before; II, after.

so that postmultiplication of some matrix by a shear has the effect of adding columns. For example,

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \underset{\text{Shear}}{\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}} = \begin{bmatrix} a & a+b \\ c & c+d \end{bmatrix}$$

while premultiplication has the effect of adding rows. For example,

$$\underset{\text{Shear}}{\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a+c & b+d \\ c & d \end{bmatrix}$$

The geometric effect of a shear is shown, illustratively, in Fig. 4.11 for the simple case involving the rectangle

$$P = (1,0); \qquad Q = (3,0); \qquad R = (3,2); \qquad S = (1,2)$$

When the shear transformation postmultiplies the vertices of the rectangle, we obtain:

$$\begin{array}{c} P \\ Q \\ R \\ S \end{array} \begin{bmatrix} 1 & 1 \\ 3 & 3 \\ 3 & 5 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 3 & 0 \\ 3 & 2 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

as shown in Panel II of Fig. 4.11.

TABLE 4.8

A Shear Transformation of Matrix $\mathbf{X}_d$

$$\overset{\mathbf{X}_d}{\begin{bmatrix} -1 & -1 \\ -1 & 0 \\ -1 & 1 \\ 0 & -1 \\ 0 & 0 \\ 0 & 1 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix}} \overset{\text{Shear}}{\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}} = \begin{bmatrix} -1 & -2 \\ -1 & -1 \\ -1 & 0 \\ 0 & -1 \\ 0 & 0 \\ 0 & 1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}$$

If we apply the shear to the matrix $\mathbf{X}_d$, we obtain the coordinates shown in Table 4.8 and plotted in Panel VIII of Fig. 4.10.

4.5 COMPOSITE TRANSFORMATIONS

The transformations described in the preceding section have each been applied singly. It is instructive to see what happens when some of these transformations are applied on a composite basis. For illustrative purposes we consider the following:

1. a rotation followed by reflection of the first axis
2. a stretch followed by a rotation
3. a rotation followed by a stretch
4. a rotation followed by a stretch followed by another rotation
5. an arbitrary linear transformation.

The theoretical rationale for applying successive matrix transformation is based on the idea that if $\mathbf{T}$ is the matrix of one linear transformation and $\mathbf{S}$ is the matrix of a transformation that maps images obtained from $\mathbf{T}$, then the matrix to be transformed can be mapped by a *composite* transformation that involes the matrix product $\mathbf{TS}$.[6] This same idea can be extended in the same manner, to more than two transformation matrices.

The *order* in which successive matrix transformations are applied is quite important. That is, in general the results of the composite mapping involving $\mathbf{TS}$ are not the same as the images that would be obtained from the composite mapping $\mathbf{ST}$, as will be demonstrated shortly.

Moreover, it should be reiterated that the matrix of a transformation is uniquely defined only relative to a set of *specific* basis vectors.[7] All along we have been using the standard basis vectors $\mathbf{e}_i$, and we shall continue to do so here.[8]

[6] We are continuing to assume that $\mathbf{T}$ (and $\mathbf{S}$) are *postmultiplying* (say) $\mathbf{X}_d$, the initial configuration of points.

[7] Here we continue to distinguish between τ, the transformation (e.g., a stretch or a rotation), and $\mathbf{T}$, its characterization with respect to a specific set of basis vectors.

[8] We shall continue to refer to the transformation of $\mathbf{X}_d$, the configuration of points in Panel II of Fig. 4.10.

4.5.1 Rotation Followed by Reflection

If we multiply the following matrices:

$$\begin{bmatrix} 0.707 & 0.707 \\ -0.707 & 0.707 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$$

we obtain

$$\begin{bmatrix} -0.707 & 0.707 \\ 0.707 & 0.707 \end{bmatrix}$$

As can be easily verified, the resultant matrix has a determinant of −1. And as indicated earlier, this type of matrix is called an improper rotation. Although it meets the conditions of an orthogonal matrix, its application actually involves a "proper" rotation, in which the determinant is +1, *followed by a reflection* of axis 1. Table 4.9 shows the

TABLE 4.9

An Improper Rotation of Matrix $\mathbf{X_d}$

$$\begin{array}{c} \mathbf{X_d} \\ \begin{bmatrix} -1 & -1 \\ -1 & 0 \\ -1 & 1 \\ 0 & -1 \\ 0 & 0 \\ 0 & 1 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix} \end{array} \begin{array}{c} \text{Improper rotation} \\ \begin{bmatrix} -0.707 & 0.707 \\ 0.707 & 0.707 \end{bmatrix} \end{array} = \begin{bmatrix} 0 & -1.414 \\ 0.707 & -0.707 \\ 1.414 & 0 \\ -0.707 & -0.707 \\ 0 & 0 \\ 0.707 & 0.707 \\ -1.414 & 0 \\ -0.707 & 0.707 \\ 0 & 1.414 \end{bmatrix}$$

results of this composite transformation, while Panel I of Fig. 4.12 shows the geometric results. That is, the points in Panel II of Fig. 4.10 are first rotated counterclockwise about an angle of 45°, and then the (implied) $\mathbf{e}_1{}^*$ axis is reflected, as observed in Panel I of Fig. 4.12.

4.5.2 Stretch Followed by Rotation

If we multiply the following matrices:

$$\begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0.707 & 0.707 \\ -0.707 & 0.707 \end{bmatrix}$$

we obtain

$$\begin{bmatrix} 1.414 & 1.414 \\ 0.707 & 0.707 \end{bmatrix}$$

This latter composite matrix first involves a stretch of the configuration followed by a counterclockwise rotation of 45°. Table 4.10 shows the computations, while Panel II of

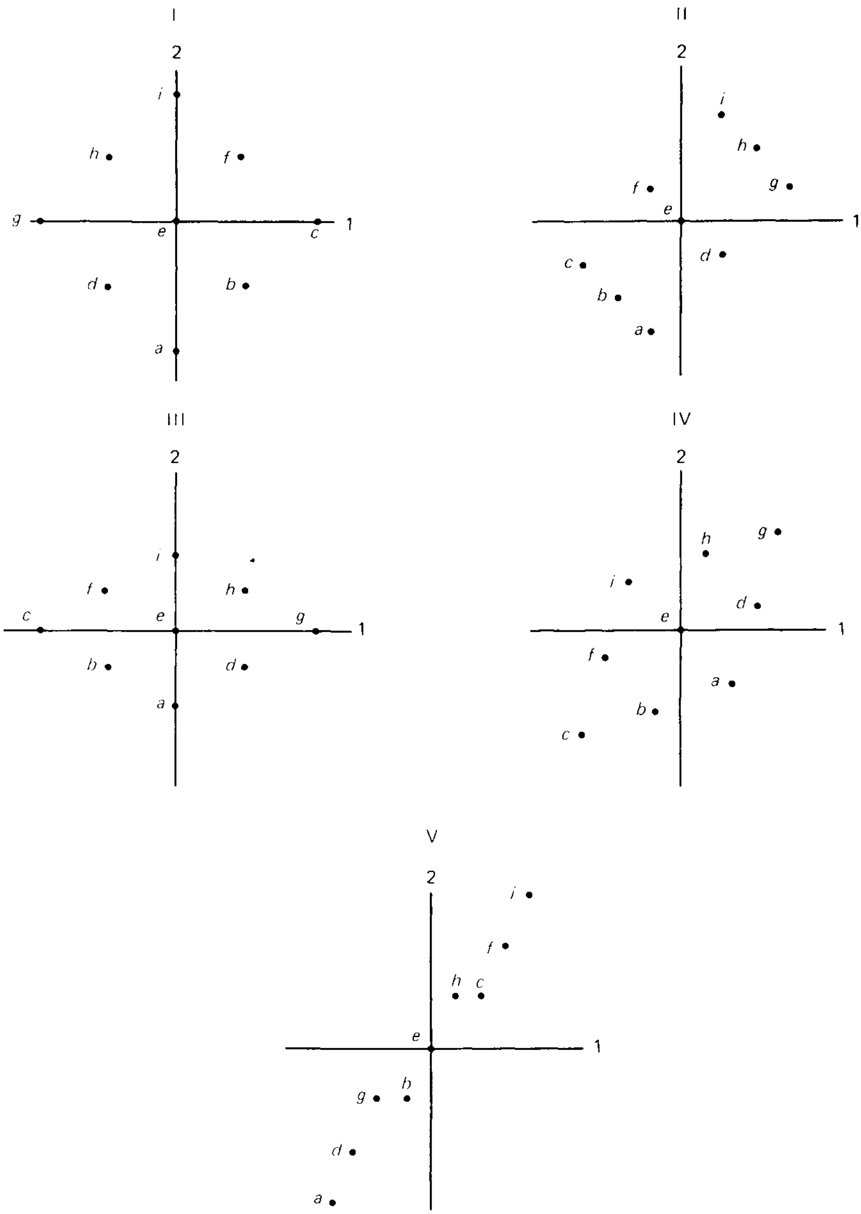

Fig. 4.12 Geometric effect of various composite transformations. Key: I, rotation followed by reflection; II, stretch followed by rotation; III, rotation followed by stretch; IV, rotation–stretch–rotation; V, arbitrary linear transformation.

TABLE 4.10

A Stretch of Matrix $\mathbf{X}_d$ *Followed by Counterclockwise Rotation through an Angle of 45°*

$$\overset{\mathbf{X}_d}{\begin{bmatrix} -1 & -1 \\ -1 & 0 \\ -1 & 1 \\ 0 & -1 \\ 0 & 0 \\ 0 & 1 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix}} \overset{\text{Stretch–rotation}}{\begin{bmatrix} 1.414 & 1.414 \\ -0.707 & 0.707 \end{bmatrix}} = \begin{bmatrix} -0.707 & -2.121 \\ -1.414 & -1.414 \\ -2.121 & -0.707 \\ 0.707 & -0.707 \\ 0 & 0 \\ -0.707 & 0.707 \\ -2.121 & 0.707 \\ 1.414 & 1.414 \\ 0.707 & 2.121 \end{bmatrix}$$

Fig. 4.12 shows the geometric results. The first matrix maps the square lattice into a rectangular lattice, while the second transformation rotates this rectangular lattice 45°, counterclockwise.

4.5.3 Rotation Followed by Stretch

In general, the matrix product $\mathbf{AB} \neq \mathbf{BA}$. That is, usually the multiplication of matrices (even if conformable) is not commutative. This can be illustrated rather dramatically by considering the following matrix product:

$$\begin{bmatrix} 0.707 & 0.707 \\ -0.707 & 0.707 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1.414 & 0.707 \\ -1.414 & 0.707 \end{bmatrix}$$

While the same two matrices as those used in the preceding section are employed here, we see that their matrix product differs markedly. In the present case we have a counterclockwise rotation of $\mathbf{X}_d$ through an angle of 45° *followed* by a stretch. The result of this is that even the original "shape" of the points (a square lattice) is deformed

TABLE 4.11

A Counterclockwise Rotation of Matrix $\mathbf{X}_d$ *through an Angle of 45° Followed by a Stretch*

$$\overset{\mathbf{X}_d}{\begin{bmatrix} -1 & -1 \\ -1 & 0 \\ -1 & 1 \\ 0 & -1 \\ 0 & 0 \\ 0 & 1 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix}} \overset{\text{Rotation–stretch}}{\begin{bmatrix} 1.414 & 0.707 \\ -1.414 & 0.707 \end{bmatrix}} = \begin{bmatrix} 0 & -1.414 \\ -1.414 & -0.707 \\ -2.829 & 0 \\ 1.414 & -0.707 \\ 0 & 0 \\ -1.414 & 0.707 \\ 2.829 & 0 \\ 1.414 & 0.707 \\ 0 & 1.414 \end{bmatrix}$$

TABLE 4.12

Rotation–Stretch–Rotation Composite Transformation of Matrix $\mathbf{X}_d$

$$\overset{\mathbf{X}_d}{\begin{bmatrix} -1 & -1 \\ -1 & 0 \\ -1 & 1 \\ 0 & -1 \\ 0 & 0 \\ 0 & 1 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix}} \; \overset{\text{Rotation–stretch–rotation}}{\begin{bmatrix} 0.87 & 1.32 \\ -1.58 & -0.10 \end{bmatrix}} = \begin{bmatrix} 0.71 & -1.22 \\ -0.87 & -1.32 \\ -2.45 & -1.42 \\ 1.58 & 0.10 \\ 0 & 0 \\ -1.58 & -0.10 \\ 2.45 & 1.42 \\ 0.87 & 1.32 \\ -0.71 & 1.22 \end{bmatrix}$$

TABLE 4.13

Application of an Arbitrary Linear Transformation to Matrix $\mathbf{X}_d$

$$\overset{\mathbf{X}_d}{\begin{bmatrix} -1 & -1 \\ -1 & 0 \\ -1 & 1 \\ 0 & -1 \\ 0 & 0 \\ 0 & 1 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix}} \; \overset{\mathbf{V}}{\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}} = \begin{bmatrix} -4 & -6 \\ -1 & -2 \\ 2 & 2 \\ -3 & -4 \\ 0 & 0 \\ 3 & 4 \\ -2 & -2 \\ 1 & 2 \\ 4 & 6 \end{bmatrix}$$

into a rhomboidlike figure. The computations appear in Table 4.11, and the geometric results appear in Panel III of Fig. 4.12.

4.5.4 A Rotation–Stretch–Rotation Composite

As an extended case, let us now consider a 45° counterclockwise rotation followed by a stretch followed by a 30° counterclockwise rotation. This combination can be illustrated by the following matrix product:

$$\begin{bmatrix} 0.707 & 0.707 \\ -0.707 & 0.707 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0.866 & 0.500 \\ -0.500 & 0.866 \end{bmatrix} = \begin{bmatrix} 0.87 & 1.32 \\ -1.58 & -0.10 \end{bmatrix}$$

Table 4.12 shows the results of applying this composite transformation, while Panel IV of Fig. 4.12 portrays the results graphically.

4.5.5 An Arbitrary Linear Transformation

To round out discussion, assume that we had the arbitrarily selected linear transformation

$$\mathbf{V} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$$

and wished to find out what would happen if this transformation were applied to the matrix $\mathbf{X}_d$. If $\mathbf{X}_d$ is postmultiplied by $\mathbf{V}$, we obtain the product shown in Table 4.13.

Panel V of Fig. 4.12 shows the effect graphically. Compared to the preceding cases the pattern of the transformed points may look a bit strange. As we shall show in the next chapter, however, even the arbitrary linear transformation $\mathbf{V}$ can be represented as a composite of more simple transformations, of the types illustrated in Fig. 4.10.

Anticipating material to be described in Chapter 5, Fig. 4.13 shows three configurations that successively portray the movement of the points of $\mathbf{X}_d$ from their original positions in Panel II of Fig. 4.10 to their positions shown in Panel V of Fig. 4.12. For the moment we shall do a bit of "hand waving" and present the results of decomposing $\mathbf{V}$ into the product of simpler transformations.

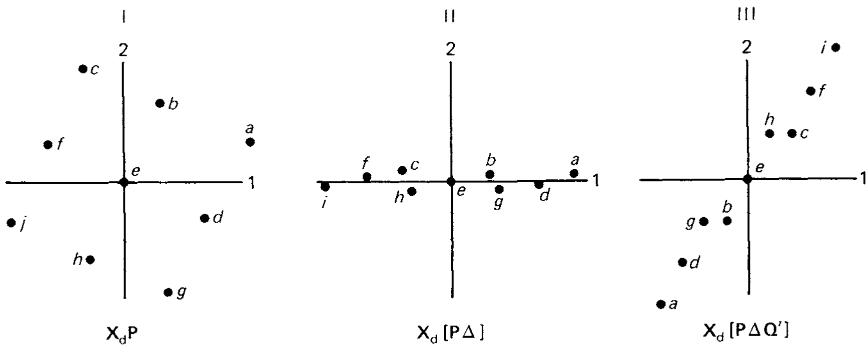

Fig. 4.13 Decomposition of general linear transformation $\mathbf{V} = \mathbf{P\Delta Q}'$.

First, suppose we postmultiply $\mathbf{X}_\mathrm{d}$ by the orthogonal transformation

$$\mathbf{P} = \begin{bmatrix} -0.41 & -0.91 \\ -0.91 & 0.41 \end{bmatrix}$$

In this case $\mathbf{P}$ produces a 66° clockwise rotation of points in $\mathbf{X}_\mathrm{d}$, followed by a reflection of the first axis. This preliminary result appears in Panel I of Fig. 4.13.

Next, let us assume that the configuration in Panel I is stretched in accordance with the diagonal matrix

$$\Delta = \begin{bmatrix} 5.47 & 0 \\ 0 & 0.37 \end{bmatrix}$$

This operation is shown in Panel II of Fig. 4.13

Finally, let us assume that the configuration in Panel II is further rotated by the orthogonal matrix

$$\mathbf{Q}' = \begin{bmatrix} -0.58 & -0.82 \\ 0.82 & -0.58 \end{bmatrix}$$

This rotation involves a clockwise movement of 125° from the orientation in Panel II. The results appear in Panel III of Fig. 4.13. Table 4.14 shows the accompanying numerical results.

The upshot of all of this is that the arbitrary linear transformation

$$\mathbf{V} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$$

has been decomposed into the product of an improper rotation, followed by a stretch, followed by another rotation. This triple product can be represented as

$$\boxed{\mathbf{V} = \mathbf{P\Delta Q}'}$$

TABLE 4.14

Decomposition of an Arbitrary Linear Transformation

$$\overset{\mathbf{X}_d}{\begin{bmatrix} -1 & -1 \\ -1 & 0 \\ -1 & 1 \\ 0 & -1 \\ 0 & 0 \\ 0 & 1 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix}}
\overset{\substack{\mathbf{P} \\ \text{Rotation–} \\ \text{reflection}}}{\begin{bmatrix} -0.41 & -0.91 \\ -0.91 & 0.41 \end{bmatrix}}
\Rightarrow
\overset{\mathbf{X}_d\mathbf{P}}{\begin{bmatrix} 1.32 & 0.50 \\ 0.41 & 0.91 \\ -0.50 & 1.32 \\ 0.91 & -0.41 \\ 0 & 0 \\ -0.91 & 0.41 \\ 0.50 & -1.32 \\ -0.41 & -0.91 \\ -1.32 & -0.50 \end{bmatrix}}
\overset{\substack{\Delta \\ \text{Stretch}}}{\begin{bmatrix} 5.47 & 0 \\ 0 & 0.37 \end{bmatrix}}
\Rightarrow
\overset{\mathbf{X}_d[\mathbf{P}\,\Delta]}{\begin{bmatrix} 7.22 & 0.19 \\ 2.24 & 0.34 \\ -2.74 & 0.49 \\ 4.98 & -0.15 \\ 0 & 0 \\ -4.98 & 0.15 \\ 2.74 & -0.49 \\ -2.24 & -3.34 \\ -7.22 & -0.19 \end{bmatrix}}
\overset{\substack{\mathbf{Q}' \\ \text{Rotation}}}{\begin{bmatrix} -0.58 & -0.82 \\ 0.82 & -0.58 \end{bmatrix}}
\Rightarrow
\overset{\mathbf{X}_d[\mathbf{P}\,\Delta\,\mathbf{Q}']}{\begin{bmatrix} -4 & -6 \\ -1 & -2 \\ 2 & 2 \\ -3 & -4 \\ 0 & 0 \\ 3 & 4 \\ -2 & -2 \\ 1 & 2 \\ 4 & 6 \end{bmatrix}}$$

where both **P** and **Q** are orthogonal and Δ is diagonal. This rather remarkable result is of extreme importance to multivariate analysis.[9] A major part of Chapter 5 is devoted to discussing this class of decompositions in some detail.

4.5.6 Observations on Composite Transformations

Clearly we could continue with various other kinds of composite transformations, although sufficient variety has been shown for the reader to get some idea of their geometric effect. As noted above, each simple matrix, such as a stretch, rotation, or reflection, is associated with a geometric analogue. It is when these operations are considered in a composite way that the overall transformation appears complex.

As it turns out, however, the value of this approach lies precisely in looking at the other side of the coin. That is, by *decomposing* seemingly complex-appearing matrices into the product of simpler ones, we can gain a geometric understanding of the transformation in a direct, intuitive way.

And, so it turns out, *any nonsingular matrix transformation with real-valued entries can be uniquely decomposed into the product of either (a) a rotation, followed by a stretch, followed by another rotation or (b) a rotation, followed by a reflection, followed by a stretch, followed by another rotation.*

This important and useful result will stand us in good stead in examining the geometric aspects of various multivariate techniques in Chapters 5 and 6. Its value lies in contributing to our understanding of what goes on under various matrix transformations. Indeed, still further generalizations are possible in cases where the matrix transformation is singular, as will be examined in Chapter 5.

4.6 INVERTIBLE TRANSFORMATIONS AND MATRIX RANK

As pointed out at the beginning of the chapter, all matrix transformations involve sets of linear equations; conversely, sets of linear equations can be compactly displayed in matrix form. The purpose of this section is to pull together material briefly presented earlier on the topics of matrix inversion and determinants, along with additional concepts as related to the general objective of solving sets of simultaneous linear equations.

As the reader may recall from basic algebra, in the general problem of attempting to solve m linear equations in n unknowns, three possibilities can arise:

1. The set of equations may have no solution; that is, they may form an inconsistent system.
2. The set of equations, while consistent, may have an infinite number of solutions.
3. The set of equations may be both consistent and have exactly one solution.

Solutions of simultaneous linear equations based on the application of inversion assume that the number of equations equals the number of unknowns. In this instance, the matrix of coefficients is square. If other conditions (to be described) are met, the matrix

[9] The representation $\mathbf{V} = \mathbf{P}\Delta\mathbf{Q}'$ is variously called decomposition to basic structure or singular value decomposition.

of coefficients has a *unique inverse,* and exactly one solution exists for the set of equations. Accordingly, we emphasize the last of the three cases above since this is the one that involves matrix inversion and, furthermore, is most relevant for multivariate analysis.

Appendix B discusses the topic of simultaneous linear equations from a much broader point of view, one that encompasses all three of the preceding cases and describes their characteristics in detail.

As recalled, a set of n linear equations in n variables can be compactly written in matrix form as

$$\mathbf{A}_{n \times n}\mathbf{x}_{n \times 1} = \mathbf{b}_{n \times 1}$$

where $\mathbf{A}$ is the $n \times n$ matrix of coefficients, $\mathbf{x}$ is the $n \times 1$ column vector of unknowns, and $\mathbf{b}$ is the $n \times 1$ column vector of constants. For example, suppose we had the following equations:

$$4x_1 - 10x_2 = -2$$

$$3x_1 + \;\; 7x_2 = 13$$

These can be written in matrix form as

$$\begin{array}{ccc} \mathbf{A} & \mathbf{x} & & \mathbf{b} \\ \begin{bmatrix} 4 & -10 \\ 3 & 7 \end{bmatrix} & \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} & = & \begin{bmatrix} -2 \\ 13 \end{bmatrix} \end{array}$$

If the inverse $\mathbf{A}^{-1}$ exists, then we also know that the following relationship holds:

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$$

where $\mathbf{A}^{-1}$ is unique. We can solve for $\mathbf{x}$, the vector of unknowns, as follows:

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

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

$$\mathbf{I}\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$$

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

Furthermore, from previous discussion we also know that $\mathbf{A}^{-1}$ exists, provided that $|\mathbf{A}| \neq 0$. In the current example, $|\mathbf{A}| = 58$. From Section 4.3.2, we can find $\mathbf{A}^{-1}$ by dividing each entry of the adjoint of $\mathbf{A}$, adj($\mathbf{A}$), by the determinant of $\mathbf{A}$:

$$\mathbf{A}^{-1} = \frac{1}{58} \overset{\text{adj}(\mathbf{A})}{\begin{bmatrix} 7 & 10 \\ -3 & 4 \end{bmatrix}} = \begin{bmatrix} 7/58 & 5/29 \\ -3/58 & 2/29 \end{bmatrix}$$

If $\mathbf{b}$ is then premultiplied by $\mathbf{A}^{-1}$, we obtain the solution vector:

$$\mathbf{x} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

In short, if the determinant of $|\mathbf{A}|$, the coefficient matrix, is not equal to zero, computing the inverse $\mathbf{A}^{-1}$ is a useful way to solve sets of linear equations in which the number of equations equals the number of unknowns.

Two major questions crop up in the discussion of general solution methods for sets of linear equations:

1. If the conditions are such that matrix inversion methods can be employed, what are some of the properties of matrix inverses?
2. Suppose $\mathbf{A}^{-1}$ does not exist, but we still want to say something about those aspects of the space that are preserved under the linear transformation $\mathbf{A}$. What is the connection between the number of linearly independent dimensions in the transformation and the number of dimensions that are preserved under that transformation?

Discussion of these two questions constitutes the primary focus of this section of the chapter.

4.6.1 Properties of Matrix Inverses

In Section 4.3.3 we described one important property of matrix inverses, namely, that the inverse of the product of two or more (conformable) matrices equals the product of the separate inverses in reverse order:

$$\boxed{(\mathbf{T}_1\mathbf{T}_2 \cdots \mathbf{T}_s)^{-1} = \mathbf{T}_s^{-1} \cdots \mathbf{T}_2^{-1}\mathbf{T}_1^{-1}}$$

Moreover, in Section 4.3.4 we made note of the fact that if a matrix $\mathbf{B}$ is orthogonal, then

$$\mathbf{B}' = \mathbf{B}^{-1}$$

Some other aspects of matrix inverses are also useful to point out:

1. The inverse of an inverse is the original matrix:

$$\boxed{(\mathbf{A}^{-1})^{-1} = \mathbf{A}}$$

2. The inverse of a scalar times a matrix equals the reciprocal of the scalar times the matrix inverse:

$$\boxed{(k\mathbf{A})^{-1} = 1/k\mathbf{A}^{-1}}$$

3. The inverse of the transpose of a matrix, equals the transpose of the inverse:

$$\boxed{(\mathbf{A}')^{-1} = (\mathbf{A}^{-1})'}$$

4. The inverse of the diagonal matrix $\mathbf{D}$ is obtained by simply finding the reciprocals of the entries on the main diagonal:

$$\boxed{(\mathbf{D}^{-1}) = \text{diag}(d_{ii}^{-1})}$$

5. If **A**, **B**, and **C** are each square and of order n by n, and if **A** is nonsingular, then

$$\boxed{\mathbf{AB} = \mathbf{AC} \qquad \text{implies} \qquad \mathbf{B} = \mathbf{C}}$$

Note, then, that the cancellation law of scalar algebra holds over the set of nonsingular matrix transformations. We shall use several of the above properties in Chapters 5 and 6 that employ inverses in various types of multivariate computations. In the present geometrically oriented context, however, we note that inverses relate to *invertible* functions in which for every vector in one space we have one and only one uniquely paired vector in another space. This "other" space may, of course, be the original space. The image vector is then another point in the same space as the preimage vector.

4.6.2 Characteristics of Invertible Transformations

If we consider the following transformation:

$$\mathbf{x}^* = \begin{bmatrix} 2.23 \\ 2.20 \end{bmatrix} = \overset{\mathbf{T}}{\begin{bmatrix} 0.45 & 0.89 \\ 0.60 & 0.80 \end{bmatrix}} \overset{\mathbf{x}}{\begin{bmatrix} 1 \\ 2 \end{bmatrix}}$$

we note that a point $\mathbf{x} = \left[\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right]$ in two dimensions is mapped by **T** onto $\mathbf{x}^*$, another point in that same space. Furthermore, by methods discussed earlier, we could find $\mathbf{T}^{-1}$ and observe the following:

$$\mathbf{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \overset{\mathbf{T}^{-1}}{\begin{bmatrix} -4.60 & 5.11 \\ 3.45 & -2.59 \end{bmatrix}} \overset{\mathbf{x}^*}{\begin{bmatrix} 2.23 \\ 2.20 \end{bmatrix}}$$

Here, the inverse $\mathbf{T}^{-1}$ maps $\mathbf{x}^*$ in the plane onto $\mathbf{x}$ in the plane. This is an illustration of a one-to-one, or *invertible,* transformation in which to each point in the (x_1, x_2) plane, we have one and only one point in the $(x_1{}^*, x_2{}^*)$ plane. Most of our discussion in Sections 4.4 and 4.5 centered around invertible transformations.

If $\mathbf{T}^{-1}$ exists, every $\mathbf{x}$ has a unique $\mathbf{x}^*$ and vice versa. Invertible transformations exhibit the important property of being nonsingular and, hence, square. Furthermore, as long as *all* (square) transformations $\mathbf{T}_1, \mathbf{T}_2, \ldots, \mathbf{T}_p$ are each nonsingular, their product is also nonsingular, and all information about $\mathbf{x}$ is preserved in the mapping in the sense that $(\mathbf{T}_1\mathbf{T}_2 \cdots \mathbf{T}_p)^{-1}$ could undo the original composite transformation by transforming $\mathbf{x}^*$ back to $\mathbf{x}$.

What happens if **T** is singular and, hence, $\mathbf{T}^{-1}$ does not exist? As might be surmised, if such is the case, we lose the one-to-one correspondence between $\mathbf{x}$ and $\mathbf{x}^*$. Moreover, it may be the case that additional information about $\mathbf{x}$ is irretrievably lost in the mapping process. What can be said about the dimensionality of the image space under these circumstances? A discussion of this question involves a topic of central importance to matrix algebra, namely, the *rank* of a matrix.

4.6.3 The Rank of a Matrix

The concept of matrix rank is related to two topics that have already been discussed in earlier chapters:

1. linear dependence of a set of row or column vectors,
2. the determinant of a matrix.

There are two basic, and compatible, ways of defining the rank of a matrix.

One definition takes a dimensional, or geometric, viewpoint. Assume that we have a matrix **A** that is not necessarily square. *The rank of* **A**, *denoted by* $r(\mathbf{A})$, *is defined as the maximum number of linearly independent rows (columns) of* **A**. While it may seem strange to say that the row rank of a matrix is always equal to its column rank, such is the case, as we shall illustrate subsequently. Any matrix, square or rectangular, has a unique rank, one that equals the maximum number of linearly independent vectors.

In Chapter 3 we discussed the concept of linear independence of row or column vectors. In fact, the Gram–Schmidt process was illustrated as a way of finding an orthonormal basis from a set of arbitrary basis vectors.

At that time we pointed out that n linearly independent vectors, each of n components, are sufficient to define a basis. If more that n vectors are present, then the set cannot be linearly independent. However, if less than n vectors are present, they are insufficient to span a space of n dimensions. That is, either the number of components or the number of vectors is sufficient to constrain the dimensionality.

What this boils down to is that if a matrix **A** is rectangular, its rank cannot exceed its smaller dimension (rows or columns as the case may be). Moreover, its row rank equals its column rank so that we can say without ambiguity that $r(\mathbf{A}) = k$, where k is some nonnegative integer. Of course, this in itself does not say that $r(\mathbf{A})$ must equal the lesser of m or n; its rank k may be *less* than the minimum of m or n. All that is being said is that one can talk just as appropriately about m points in n dimensions as n points in m dimensions and that the lesser of the two numbers denotes the maximum subspace in which the points are contained.

Notice that the definition of matrix rank refers to the *maximum* number of rows (columns) that are linearly independent. As suggested above, if a matrix **A** is of order $m \times n$, and if the matrix has rank $r(\mathbf{A}) = k$, then there exist k rows and k columns, where $k \leqslant \min(m, n)$ that are linearly independent. Furthermore, any set of $k+1$ rows (columns) is linearly dependent.

The reader will recall that we also discussed determinants in Chapter 2 and elsewhere. It turns out that a fully compatible definition of rank can be developed from the foundation of determinants.

The rank of a matrix **A**, *denoted* $r(\mathbf{A})$, *is the order of the largest square submatrix of* **A** *whose determinant is not zero.*

To illustrate the nature of the latter definition, suppose we have the following square matrix:

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

By inspection we note that the third column is a multiple of the second; clearly $|\mathbf{A}| = 0$ and $r(\mathbf{A})$ is not 3. However, consider one of the 2×2 submatrices, for example,

$$\mathbf{B} = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$$

We find that $|\mathbf{B}| = -3$; hence, $r(\mathbf{A}) = 2$.

Next, suppose we have the following rectangular matrix:

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 1 & 2 \end{bmatrix}$$

Clearly, there are no square submatrices of order 3×3 since the matrix is only 2×3; hence, $r(\mathbf{A})$ is at most 2. If we take one of the 2×2 submatrices

$$\mathbf{B} = \begin{bmatrix} 2 & 3 \\ 1 & 2 \end{bmatrix}$$

we find that $|\mathbf{B}| = 1$; hence $r(\mathbf{A}) = 2$.

It should be stated that we have to find only one square submatrix (of the desired order) with a nonzero determinant. Once we have done so, we can stop the search.

By the systematic examination of determinants of various submatrices of $\mathbf{A}$, we have a way to go about finding $r(\mathbf{A})$ in either the square or rectangular matrix cases. Also, we should note that the lowest rank of any matrix must be zero, and this would happen only if $\mathbf{A} = \phi$; that is, the matrix consisted of all zeros. Otherwise, there would be some nonzero element in the single-element minors, and $r(\mathbf{A})$ would at least be 1.

How do we relate the concept of matrix rank to the topics of matrix inverse and invertible transformations described earlier? Perhaps the most direct way is to state that if we have a square $(n \times n)$ transformation matrix $\mathbf{A}$, whose inverse $\mathbf{A}^{-1}$ exists, the following statements are all equivalent:

1. $\mathbf{A}$ is nonsingular; that is, $|\mathbf{A}| \neq 0$.
2. The rank $r(\mathbf{A}) = n$; that is, $\mathbf{A}$ is of full rank in which its rank equals its order.
3. The row vectors of $\mathbf{A}$ are linearly independent.
4. The column vectors of $\mathbf{A}$ are linearly independent.
5. The image space obtained from $\mathbf{A}$ fully preserves the preimage space in a one-to-one fashion.
6. One can obtain the unique preimages transformed by $\mathbf{A}$ from their counterpart images by means of $\mathbf{A}^{-1}$, the inverse.
7. The specific image points and preimage points are in one-to-one correspondence.

Of course, not all transformations of interest to multivariate analysis will involve cases in which the inverse $\mathbf{A}^{-1}$ exists. Accordingly, means are needed to find out what happens when the transformation is *not* fully invertible. Accordingly, we now consider some of the difficulties that arise when the transformation matrix is singular.

4.6.4 The Relationship of Rank to Linear Transformations

As might be surmised at this point, the rank of a matrix transformation is quite important in matrix algebra. Consider the transformation matrix

$$\mathbf{T} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}$$

and the vector $\mathbf{x} = \left[\begin{smallmatrix}1\\2\end{smallmatrix}\right]$. Via matrix multiplication, we can find the point transformation $\mathbf{x}^* = \mathbf{T}\mathbf{x}$ as follows:

$$\mathbf{x}^* = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$$

and, presumably $\mathbf{T}$ has taken $\mathbf{x}$ into three dimensions. However, as can be seen in Fig. 4.14, the transformation rotates the $\mathbf{e}_1$, $\mathbf{e}_2$ plane through a 45° angle. All of the points in the $\mathbf{e}_1$, $\mathbf{e}_2$ plane, including $\mathbf{x} = \left[\begin{smallmatrix}1\\2\end{smallmatrix}\right]$, undergo this rotation. The range of the transformation *is still a plane.* What should be remembered is that the range of any transformation that takes a point in n dimensions into a point in m dimensions (where $m > n$) cannot exceed n; that is, the higher space *cannot* be filled from a space of lower dimensionality.

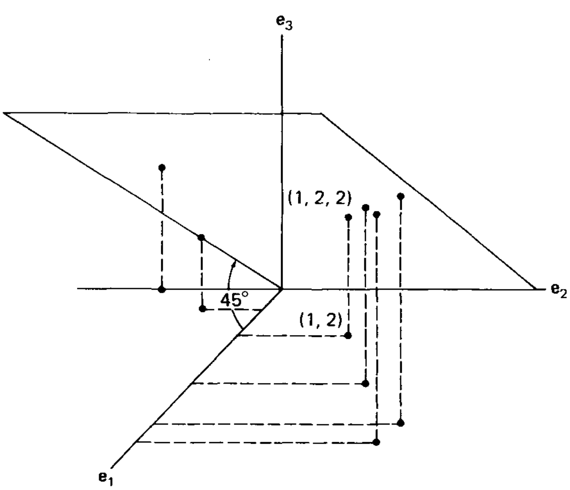

Fig. 4.14

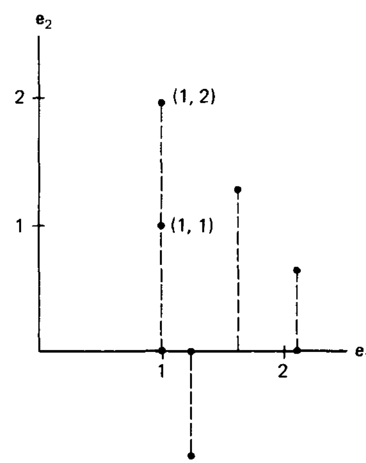

Fig. 4.15

Fig. 4.14 Geometric effect of a linear transformation involving a point in a lower dimensionality.
Fig. 4.15 Geometric effect of a linear transformation involving a point in a higher dimensionality.

Notice that $\mathbf{T}$ has only two columns, and $r(\mathbf{T})$ is at most equal to 2. In this case $r(\mathbf{T})$ *is* equal to 2, and we observe that the third row of $\mathbf{T}$ equals the second. Thus, the rank of $\mathbf{T}$ has placed restrictions on the dimensionality of the transformation.

Now consider the matrix

$$\mathbf{S} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

In this case if we desire to find $\mathbf{x}^* = \mathbf{S}\mathbf{x}$, we have

$$\mathbf{x}^* = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

and, indeed, all points in two dimensions will be mapped onto the (single) $\mathbf{e}_1$ axis. That is, $r(\mathbf{S}) = 1$ and, hence, the transformation maps all vectors, $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$, etc., onto a line, regardless of the original dimensionality of $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$. Fig. 4.15 shows this effect geometrically. Note in particular that points (1, 2) and (1, 1) are not distinguished after this transformation. Thus, the transformation preserves the first component of the vector but not the second; these all become zero, and this information is irretrievably lost.

The fact that the rank of a matrix and the number of linearly independent vectors of a matrix are equal is important in the understanding of matrix transformations generally. As noted above, knowledge of the rank of a transformation matrix provides information about the characteristics of the original dimensionality that are "preserved" under the mapping.

To round out the preceding comments, we can extend the discussion of Section 3.3.5 on vector projection to the more general case of projecting a vector in n dimensions onto some hyperplane of dimension k ($k < n$) passing through the origin of the space (see Panel IV of Fig. 3.15 for an illustrative case).

For example, suppose we have a three-dimensional space and a plane passing through the origin of that space and through the two points

$$\mathbf{b}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}; \qquad \mathbf{b}_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

Let us also assume that the vector of interest is represented by

$$\mathbf{a} = \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}$$

in the full space of three dimensions. If so, the projection $\mathbf{a}^*$ of $\mathbf{a}$ onto the plane defined by $\mathbf{0}$, $\mathbf{b}_1$, and $\mathbf{b}_2$ is given by

$$\boxed{\mathbf{a}^* = \mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\,\mathbf{B}'\mathbf{a}}$$

where **B** is the matrix whose columns are the vectors $\mathbf{b}_1$ and $\mathbf{b}_2$ in the plane of interest. In terms of this numerical example, we have

$$\mathbf{a}^* = \underset{\mathbf{B}}{\begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}} \underset{(\mathbf{B}'\mathbf{B})^{-1}}{\frac{1}{3}\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}} \underset{\mathbf{B}'}{\begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}} \underset{\mathbf{a}}{\begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}} = \begin{bmatrix} 8/3 \\ 7/3 \\ 2/3 \end{bmatrix}$$

and $\mathbf{a}^*$ is the vector representing the projection of **a** onto the plane defined by **0**, the origin, and $\mathbf{b}_1$ and $\mathbf{b}_2$.

Since the line through **0** and $\mathbf{b}_1$ is in the plane, an orthogonal projection implies that $(\mathbf{a}^* - \mathbf{a})$ be perpendicular to $\mathbf{b}_1$; similarly so for the line through **0** and $\mathbf{b}_2$, giving the equations

$$(\mathbf{a}^* - \mathbf{a})'\mathbf{b}_1 = 0; \qquad (\mathbf{a}^* - \mathbf{a})'\mathbf{b}_2 = 0$$

A brief sketch of the derivation of $\mathbf{a}^*$ may be in order. In more general terms, if the dimensionality of the full space is n and k denotes the dimensionality of the hyperplane defined by **0** and the $\mathbf{b}_i$, then the $n \times k$ matrix **B** has rank k, and we have

$$(\mathbf{a}^* - \mathbf{a})'\mathbf{b}_1 = 0; \quad (\mathbf{a}^* - \mathbf{a})'\mathbf{b}_2 = 0; \ldots; \quad (\mathbf{a}^* - \mathbf{a})'\mathbf{b}_k = 0$$

or

$$\mathbf{a}^{*\prime}\mathbf{b}_i = \mathbf{a}'\mathbf{b}_i \qquad \text{for} \quad i = 1, 2, \ldots, k$$

In matrix notation, this can be written as

$$\mathbf{a}^{*\prime}\mathbf{B} = \mathbf{a}'\mathbf{B}$$

However, since $\mathbf{a}^*$ is itself in the hyperplane, it represents a linear combination of the vectors $\mathbf{b}_i$:

$$\mathbf{a}^* = \sum_{i=1}^{k} p_i \mathbf{b}_i$$

or, equivalently,

$$\mathbf{a}^* = \mathbf{Bp}$$

where **p** is a vector of arbitrary scalars defining the linear combination. Substituting for $\mathbf{a}^{*\prime}$ above, we have

$$\mathbf{p}'\mathbf{B}'\mathbf{B} = \mathbf{a}'\mathbf{B}$$

B is $n \times k$ and of rank k. $\mathbf{B}'\mathbf{B}$ is $k \times k$ and nonsingular, and we can postmultiply both sides by $(\mathbf{B}'\mathbf{B})^{-1}$ to get

$$\mathbf{p}' = \mathbf{a}'\mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}$$

Next, we can find the transpose of $\mathbf{p}'$:

$$\mathbf{p} = (\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'\mathbf{a}$$

and recall that $(\mathbf{B}'\mathbf{B})^{-1}$ is symmetric. Finally, we substitute $\mathbf{B}^{-1}\mathbf{a}^*$ for $\mathbf{p}$ and then premultiply both sides by $\mathbf{B}$ to get

$$\mathbf{a}^* = \mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'\mathbf{a}$$

as desired.

The reason for introducing this generalization of vector projection at all is that the idea of (orthogonal) projection plays a central role in least squares, one of the cornerstone methods in multivariate analysis.

4.6.5 Transformations Involving Singular Matrices

To round out our discussion of matrix rank and its relationship to determinants and linear independence, let us consider the geometric effect of a singular matrix from another viewpoint. In Section 3.6.1 we showed how the determinant of a 2×2 transformation matrix $\mathbf{T}$ measures the ratio of areas of the transformed to original figure. As illustrated, in the case of vertices of the unit square

$$O = (0,0); \qquad I = (1,0); \qquad J = (1,1); \qquad K = (0,1)$$

Transformed by

$$\mathbf{T} = \begin{bmatrix} 2 & 1 \\ 3 & 4 \end{bmatrix}$$

we obtained the quadrilateral, reproduced in Panel I of Fig. 4.16. The ratio of areas between transformed and original figure is given by $|\mathbf{T}| = 5$.

Now suppose we transformed the vertices of the unit square by

$$\mathbf{U} = \begin{bmatrix} 2 & 4 \\ 3 & 6 \end{bmatrix}$$

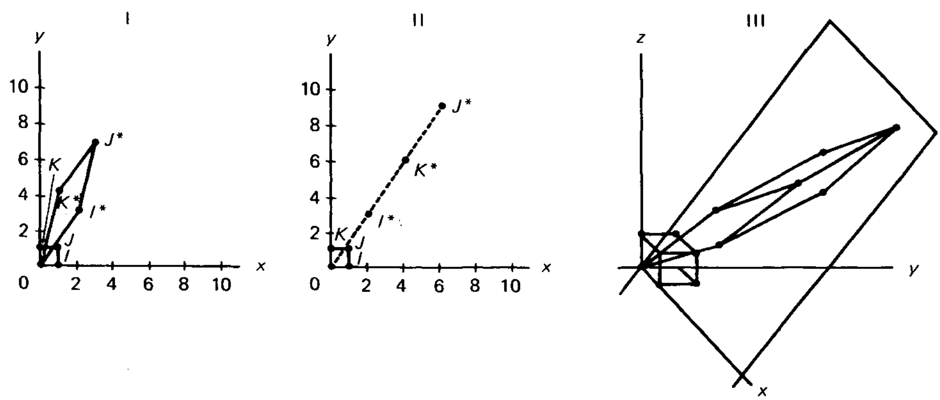

Fig. 4.16 Geometric relationships involving nonsingular and singular transformations.

leading to

$$\mathbf{V} = \begin{bmatrix} 2 & 4 \\ 3 & 6 \end{bmatrix} \begin{array}{c} \begin{array}{cccc} O & I & J & K \end{array} \\ \begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix} \end{array} = \begin{array}{c} \begin{array}{cccc} O^* & I^* & J^* & K^* \end{array} \\ \begin{bmatrix} 0 & 2 & 6 & 4 \\ 0 & 3 & 9 & 6 \end{bmatrix} \end{array}$$

When we plot the new points of **V** in Panel II of Fig. 4.16, the disconcerting outcome is that **V** becomes a straight line of zero area.

This result, of course, is not hard to understand when we note that $|\mathbf{U}| = 0$ and that the entries in the second column of **U** are twice those in the first. Under $\mathbf{T}^{-1}$, the inverse of **T**, we could reverse the mapping and get back to the unit square in Panel I of the figure. This is not possible in the case of the second matrix **U** since $\mathbf{U}^{-1}$ does not exist. *Hence, whenever a transformation matrix* U *is singular, that mapping is not invertible.*

Panel III shows another case, this one involving a 3 x 3 transformation matrix:

$$\mathbf{W} = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 2 \\ 2 & 4 & 4 \end{bmatrix}$$

as applied to the vertices of the unit cube. Here, the entries in the third row are twice those of the second, and all points of the unit cube lie on a plane through the origin that has the equation $z = 2y$.

Hence, the parallelepiped that would have been obtained had **W** been nonsingular has collapsed into a parallelogram of only two dimensions. Again we have the case in which the original mapping obtained from **W** is not invertible, and the rows (columns) of **W** are not linearly independent.[10]

4.6.6 Finding the Rank of a Matrix via Determinants

A number of methods are available for finding the rank of an arbitrary matrix. Perhaps the most straightforward, if tedious, precedure is by means of determinants. If **A** is square and of order $n \times n$, we first see if $|\mathbf{A}| \neq 0$. If so, then $r(\mathbf{A}) = n$.

If $|\mathbf{A}|$ is zero, we then examine square submatrices of order $(n - 1) \times (n - 1)$. If one of these has a nonzero determinant, then we stop and state that $r(\mathbf{A}) = n - 1$. If all determinants are zero, we continue with square submatrices of order $(n - 2) \times (n - 2)$, and so on.

If **A** is rectangular of order $m \times n$, we know at the outset that $r(\mathbf{A}) \leqslant \min(m, n)$. Having established which order is smaller–suppose it is m–we examine square

[10] Note, however, that the x, y dimensions *are* retained; it is only the third dimension that collapses. The rank of the transformation indicates how many dimensions will be retained (two out of three in this case).

submatrices of order $m \times m$ to see if one can be found whose determinant is nonzero. If so, then $r(\mathbf{A}) = m$. If not, we examine square submatrices of order $(m-1) \times (m-1)$, and so on, as indicated above.

To illustrate how one might go about finding the rank of a matrix via these procedures, we can examine a few examples. First, consider the 2×2 matrix

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

We compute the determinant of **A** and find that $|\mathbf{A}| = -5$. Since $|\mathbf{A}| \neq 0$, we conclude that **A** has rank 2.

On the other hand, consider the following case:

$$\mathbf{B} = \begin{bmatrix} 2 & 3 \\ 1 & 1.5 \end{bmatrix}$$

where we first find that $|\mathbf{B}| = 0$. Since the determinant of the original matrix has vanished, we must examine single-element minors. Since there are first-order determinants, such as $|2|$ or $|3|$, that are not equal to zero, **B** has rank 1.

Now let us examine the 3×4 matrix:

$$\mathbf{C} = \begin{bmatrix} 1 & 1 & 2 & 5 \\ 2 & 2 & 4 & 10 \\ 1 & 0 & 3 & 1 \end{bmatrix}$$

At the outset we know that the rank of **C** cannot exceed 3, the smaller order. However, we see immediately that the second row is twice the first row; hence **C** cannot be of rank 3 since the row vectors are not linearly independent. But, since the first and third rows are linearly independent (i.e., neither is a multiple of the other), **C** has rank 2. This could be checked by observing that any of the six square 2×2 submatrices:

$$\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}; \quad \begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix}; \quad \begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix}; \quad \begin{bmatrix} 1 & 5 \\ 1 & 1 \end{bmatrix}; \quad \begin{bmatrix} 1 & 5 \\ 0 & 1 \end{bmatrix}; \quad \begin{bmatrix} 2 & 5 \\ 3 & 1 \end{bmatrix}$$

made up of elements from rows 1 and 3 each has a nonzero determinant. [Of course, only one such submatrix would be needed to establish that $r(\mathbf{C}) = 2$.]

If the order of the matrix is large, the procedure outlined above becomes rather tedious. Fortunately, other ways of finding the rank of a matrix exist. Some of these are described in Chapter 5, and one of these—based on the *echelon matrix* procedure—is discussed in the last main section of this chapter.

4.6.7 The Uses of Rank in Matrix Algebra

The rank of a matrix plays several important roles in matrix algebra. For example, in solving a set of simultaneous linear equations, it is the case that when (and only when) the rank of the matrix of coefficients equals the rank of the augmented matrix, the set of equations has at least one solution. By "augmented matrix" is meant a matrix consisting of the coefficients to which has been appended an additional column made up of the constants. For example, in the equations described earlier:

$$4x_1 - 10x_2 = -2$$

$$3x_1 + 7x_2 = 13$$

the augmented matrix is

$$\begin{bmatrix} 4 & -10 & -2 \\ 3 & 7 & 13 \end{bmatrix}$$

If the rank of this matrix equals the rank of the matrix of coefficients:

$$\begin{bmatrix} 4 & -10 \\ 3 & 7 \end{bmatrix}$$

the system is consistent and has at least one solution.

Second, if the set of equations is consistent, their solution is unique if (and only if) the rank of the coefficients matrix and, hence, that of the augmented matrix equals the number of unknowns. In this case the sytem can be solved by inversion of the coefficients matrix. If the rank is less than the number of unknowns, then an infinite number of solutions exist. While we do not delve into detailed discussion here of these two major uses of matrix rank (see Appendix B for this), enough has been said to show the importance of the concept in the solution of simultaneous equations.

A second important role that is played by matrix rank concerns those aspects of a configuration of points that are preserved under matrix transformations. As we know, two linearly independent vectors are needed to span a plane, three to span a three-dimensional space, and so on. And, as illustrated in Figs. 4.14 and 4.15, the rank of a matrix determines what aspects of the configuration will be retained after transformation.

Furthermore, as illustrated in Fig. 4.16, if the transformation is not of full rank, the image of a two-dimensional unit square could collapse to a line or to the origin; the image of a three-dimensional unit cube could collapse to an area or to a line, or to the origin. *In general, then, the rank of a transformation matrix determines the dimensionality of the image space.* Perhaps more than anything else, this is the essential aspect of matrix rank in multivariate analysis.

At this point we have discussed matrix inversion as a way to solve simultaneous equations in which the number of equations equals the number of unknowns and, hence, where it is possible that an inverse of the matrix of coefficients exists. We have also examined what happens when a square matrix is singular and the effect that this has on the transformation.

However, what has not been discussed in detail as yet are three related topics:

1. What other procedures are available for finding the rank of a matrix?
2. What procedures, other than computation of the adjoint matrix, are available for finding the inverse of a matrix?
3. How can these numerical methods for finding inverses be applied to problems in multivariate analysis?

The last main section of the chapter takes up these questions.

4.7 METHODS FOR RANK DETERMINATION AND MATRIX INVERSION

In multivariate analysis numerous occasions arise in which we wish to solve a set of simultaneous equations. Often the system will be consistent, and the matrix of coefficients will be square and of rank equal to its order; if so, a solution based on matrix inversion can be found.

Sometimes, however, the matrix of coefficients will either be rectangular or, even if square, singular. In the latter case its rank will not equal its order, and we may wish to find out—by means other than the tedious examination of determinants of square submatrices—what the rank of the transformation matrix is. A highly general approach to determining the rank of a matrix makes use of what are called elementary operations and the associated construction of echelon matrices. We first discuss this alternative approach to the determination of rank and its relationship to the solution of simultaneous linear equations.

We then return to the pivotal method, first used in Chapter 2, to compute determinants. As mentioned there, the pivotal method is also applicable to solving simultaneous equations and computing inverses. We discuss these extensions and illustrate their application to the 4 x 4 matrix that was described in Table 2.2 and to statistical data drawn from the sample problem of Table 1.2.

4.7.1 Elementary Operations

Elementary operations play an essential role in the solution of sets of simultaneous equations. Illustratively taking the case of the rows of a transformation matrix, there are three basic operations—called elementary row operations—that can be used to transform one matrix into another. We may

1. interchange any two rows;
2. multiply any row by a nonzero scalar;
3. add to any given row a scalar multiple of some other row.

If we change some matrix **A** *into another matrix* **B** *by the use of elementary row operations, we say that* **B** *is row equivalent to* **A**.

Elementary row operations involve the multiplication of **A** by special kinds of matrices that effect the above transformations. However, we could just as easily talk about elementary column operations—the same kinds as these shown above—that are applied to the columns of **A**. A matrix so transformed would be called column equivalent to the original matrix. To simplify our discussion, we illustrate the ideas via elementary row operations. The reader should bear in mind, however, that the same approach is applicable to the columns of **A**.

To illustrate how elementary row operations can be applied to the general problem of solving a set of simultaneous equations, let us again consider the two equations described earlier:

$$\mathrm{I}\begin{cases} 4x_1 - 10x_2 = -2 \\ 3x_1 + 7x_2 = 13 \end{cases}$$

Suppose we first multiply all members of the first equation by $-\frac{3}{4}$ and then add the result to those of the second equation:

$$\text{II}\begin{cases} 4x_1 - 10x_2 = -2 \\ \qquad \frac{29}{2}x_2 = \frac{29}{2} \end{cases}$$

Next, let us multiply the second equation by $\frac{2}{29}$. If so, we obtain

$$\text{III}\begin{cases} 4x_1 - 10x_2 = -2 \\ \qquad x_2 = 1 \end{cases}$$

All these sets of equations are row equivalent in the sense that all three sets have the same solution:

$$x_1 = 2$$
$$x_2 = 1$$

and each can be transformed to either of the others via elementary row operations. However, it is clear that the solution to the third set of equations is most apparent and easily found.

While elementary row operations are useful in the general task of solving sets of simultaneous equations, this is not their only desirable feature. A second, and major, attraction is the fact that elementary operations (row or column), as applied to a matrix A, *do not change its rank.* Moreover, as will be shown, elementary operations transform the given matrix in such a way as to make its rank easy to determine by inspection. As it turns out, all three sets of equations above have the same rank (rank 2) since they are all equivalent in terms of elementary row operations.

Elementary row operations are performed by a special set of square, nonsingular matrices called elementary matrices. *An elementary matrix is a nonsingular matrix that can be obtained from the identity matrix by an elementary row operation.* For example, if we wanted to interchange two rows of a matrix, we could do so by means of the permutation matrix

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

For example, if we have the point $\mathbf{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ in two dimensions, the premultiplication of $\mathbf{x}$ by the permutation matrix above would yield:[11]

$$\mathbf{x}^* = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

We note that the coordinates of x have, indeed, been permuted.

[11] In this section of the chapter we use the general format $\mathbf{x}^* = \mathbf{A}\mathbf{x}$, in which the transformation matrix premultiplies the vector (or matrix) of interest. The reader should become comfortable in using either (pre- or postmultiplication) mode.

As mentioned above, elementary matrices are nonsingular. In the 2×2 matrix case, the set of elementary matrices consists of the following:

I. *Permutation*

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

II. *Stretches*

$\begin{bmatrix} k & 0 \\ 0 & 1 \end{bmatrix}$ A stretch or compression of the plane that is parallel to the x axis; $\begin{bmatrix} 1 & 0 \\ 0 & k \end{bmatrix}$ A stretch or compression of the plane that is parallel to the y axis

III. *Shears*

$\begin{bmatrix} 1 & c \\ 0 & 1 \end{bmatrix}$ A shear parallel to the x axis; $\begin{bmatrix} 1 & 0 \\ c & 1 \end{bmatrix}$ A shear parallel to the y axis

Continuing with the numerical example involving $\mathbf{x} = \left[\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right]$, we have

Stretches

$$\begin{bmatrix} k & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} k \\ 2 \end{bmatrix}; \quad \begin{bmatrix} 1 & 0 \\ 0 & k \end{bmatrix}\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2k \end{bmatrix}$$

Shears

$$\begin{bmatrix} 1 & c \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1+2c \\ 2 \end{bmatrix}; \quad \begin{bmatrix} 1 & 0 \\ c & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ c+2 \end{bmatrix}$$

But, there is nothing stopping us from applying, in some prescribed order, a *series* of premultiplications by elementary matrices. Furthermore, the product of a set of nonsingular matrices will itself be nonsingular.

The geometric character of permutations, stretches, and shears has already been illustrated in Section 4.5. Here we are interested in two major applications of elementary row operations and the matrices that represent them:

1. determining the rank of a matrix, and
2. finding the inverse of a matrix, when such inverse exists. Each application is described in turn.

4.7.2 Elementary Operations and Matrix Rank

Three properties of matrix rank are of general interest to matrix algebra:

1. The rank of an $n \times n$ identity matrix $\mathbf{I}_{n \times n}$, is equal to n.
2. The rank of a matrix is not changed by its premultiplication (or postmultiplication) by a nonsingular matrix. In particular, elementary row operations involve nonsingular matrices and, hence, do not change the rank of the matrix being transformed.

3. If two matrices **A** and **B** have ranks that are denoted by $r(\mathbf{A})$ and $r(\mathbf{B})$, and their product **AB** is possible, then the rank of their product, denoted $r(\mathbf{AB})$, is less than or equal to the smaller rank of the two matrices:

$$\boxed{r(\mathbf{AB}) \leqslant \min[r(\mathbf{A}), r(\mathbf{B})]}$$

Suppose we first look at the implications of the second and third of the above properties. The second property indicates that if some matrix **A** has rank k, then multiplication by an elementary matrix will *not* change its rank. This is a special case of the third property, in which both matrices in the matrix product are square and of rank k.

The third property is important to our earlier discussion of invertible transformations. If the transformation matrix has a rank that is less than the matrix being transformed, all information in the preimages will not be preserved in the image space (as was illustrated in Fig. 4.15).

Suppose, now that we wish to find the rank of some arbitrary matrix **A**. Assume that we operate on **A** via a series of elementary row operations. As noted, application of a sequence of elementary matrices, each of which is nonsingular, will not change the rank of **A** but could transform **A** to a structure in which its rank can be determined by inspection. This particular type of matrix—one that is obtained by a series of elementary row operations—is called an *echelon* matrix. To illustrate, consider the following rectangular matrix:

$$\mathbf{H} = \begin{bmatrix} 0 & 1 & 3 & 0 & 4 \\ 0 & 0 & 1 & -2 & 6 \\ 0 & 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

This matrix is an example of an echelon matrix. *An echelon matrix is any matrix, square or rectangular, that exhibits the following structure:*

1. Each of the first k rows ($k \geqslant 0$) of **H** has one or more nonzero elements.
2. For each such row, the first nonzero element, as one reads from left to right, is unity.
3. The arrangement of the first k rows is such that the first nonzero element in a given row is always to the right of the first nonzero element of any row that precedes (or lies above) the given row.
4. After the first k rows, the elements of all remaining rows, if any, are all zero.

The importance of echelon matrices relates to the facts that

1. any matrix **A** can be transformed by a sequence of elementary row operations into echelon form;
2. the rank of the matrix is not altered in the process of changing it to echelon form;
3. the number of nonzero rows in **H**, the echelon form of **A**, equals the rank of **A**.

A couple of caveats are in order, however. First, it should be pointed out that **H**, the echelon form of **A**, does not "equal" **A** but, rather, can be *derived* from **A** by elementary operations. Second, **H** is, in general, not a unique representation of **A**; that is, there is no unique echelon form for a given matrix, **A**. However, neither of these caveats weakens the general usefulness of echelon matrices in rank determination and solving sets of simultaneous equations.

Transforming a given matrix to echelon form represents a relatively straightforward procedure. To illustrate, let us take the matrix

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

and apply elementary row operations to it.

1. Subtract row 1 from row 2; subtract twice row 1 from row 3.

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & -1 & -2 \\ 0 & 0 & -5 & -8 \end{bmatrix}$$

2. Multiply row 3 by $-\frac{1}{5}$.

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & -1 & -2 \\ 0 & 0 & 1 & 8/5 \end{bmatrix}$$

We note that there are three nonzero rows remaining and, hence, the rank of the echelon form of **A**—and the rank of **A** as well—is 3.

Finding the echelon matrix, as indicated by the preceding operations, involves concentrating on one row of the matrix at a time in order to obtain (a) a leading entry of unity in that row and (b) zeros in all lower rows of the column containing the leading entry of unity.

While elementary row operations are useful in finding the echelon form of a matrix and, hence, determining its rank, they are also of value in matrix inversion and the solution of simultaneous linear equations.

4.7.3 Elementary Operations and Simultaneous Equations

Let us now examine how the preceding approach to obtaining echelon matrices can be adapted to solving simultaneous equations. Since we know that the rank of the matrix illustrated above is 3, let us make up a new problem by setting down only the first three columns of the matrix used in the preceding example. Next, assume that the following simultaneous equations represent a system that we would like to solve:

$$\begin{aligned} x_1 + 2x_2 + 3x_3 &= 14 \\ x_1 + 3x_2 + 2x_3 &= 13 \\ 2x_1 + 4x_2 + x_3 &= 13 \end{aligned}$$

As illustrated earlier, this set of equations can be written as

$$\begin{bmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \\ 2 & 4 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 14 \\ 13 \\ 13 \end{bmatrix}$$

We see that the first three columns are the same as those appearing in the preceding example. Now let us go through the echelon procedure once more, but this time apply each elementary row operation to the full, *augmented* matrix consisting of coefficients *and* the vector of constants:

$$\left[\begin{array}{ccc:c} 1 & 2 & 3 & 14 \\ 1 & 3 & 2 & 13 \\ 2 & 4 & 1 & 13 \end{array}\right]$$

1. Subtract row 1 from row 2; subtract twice row 1 from row 3.

$$\left[\begin{array}{ccc:c} 1 & 2 & 3 & 14 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & -5 & -15 \end{array}\right]$$

2. Multiply row 3 by $-\frac{1}{5}$.

$$\left[\begin{array}{ccc:c} 1 & 2 & 3 & 14 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & 1 & 3 \end{array}\right]$$

If we examine the 3 x 3 matrix to the left of the dotted line, we see that the leading entry of each row is unity. Moreover, in echelon form the rank of the matrix is seen, by inspection, to be 3.

If we refer to the original set of equations, it is apparent that we can now obtain this solution quite readily. From the third row of the echelon form, immediately above, we have

$$x_3 = 3$$

If $x_3 = 3$ is then substituted in the second row, we have

$$x_2 - 3 = -1$$

$$x_2 = 2$$

If $x_2 = 2$ and $x_3 = 3$ are substituted in the first row, we have

$$x_1 + 4 + 9 = 14$$

$$x_1 = 1$$

The process of finding the value of x_3 in the third equation first, and then substituting it in the second equation to find x_2, and so on, is called back substitution. (The whole process is just an application of the pivotal method described, in the context of determinants, in Chapter 2, as will be shown later.)

4.7.4 Finding the Inverse

Now that we have seen how a set of simultaneous equations can be solved by means of elementary row operations, let us carry the same general procedure one step further to find the inverse of the matrix of coefficients. For ease of illustration we continue with the same example.

First of all, it should be clear that the set of simultaneous equations shown above can be written in the following form:

$$\begin{bmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \\ 2 & 4 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 14 \\ 13 \\ 13 \end{bmatrix}$$

which, in turn, can be expressed as

$$\mathbf{Ax} = \mathbf{Ib}$$

We could then perform a set of elementary row operations on *both* **A** and **I**. In particular, we shall try to reduce **A** to an identity matrix.

1. Subtract row 1 from row 2; subtract twice row 1 from row 3.

$$\left[\begin{array}{ccc:ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & -1 & -1 & 1 & 0 \\ 0 & 0 & -5 & -2 & 0 & 1 \end{array}\right]$$

2. Subtract twice row 2 from row 1. We include this elementary row operation in order to make sure that column 2 has only a *single* nonzero entry (in row 2).

$$\left[\begin{array}{ccc:ccc} 1 & 0 & 5 & 3 & -2 & 0 \\ 0 & 1 & -1 & -1 & 1 & 0 \\ 0 & 0 & -5 & -2 & 0 & 1 \end{array}\right]$$

3. Multiply row 3 by $-\frac{1}{5}$.

$$\left[\begin{array}{ccc:ccc} 1 & 0 & 5 & 3 & -2 & 0 \\ 0 & 1 & -1 & -1 & 1 & 0 \\ 0 & 0 & 1 & 2/5 & 0 & -1/5 \end{array}\right]$$

The 3 x 3 submatrix on the left is still not an identity matrix. Accordingly, we can apply the following additional row operations:

4. Subtract 5 times row 3 from row 1; add row 3 to row 2.

$$\left[\begin{array}{ccc:ccc} 1 & 0 & 0 & 1 & -2 & 1 \\ 0 & 1 & 0 & -3/5 & 1 & -1/5 \\ 0 & 0 & 1 & 2/5 & 0 & -1/5 \end{array}\right]$$

When we do this, we observe that the left side of the matrix is, indeed, an identity matrix.

However, while we have been transforming **A** to **I** via elementary row operations, we have, at the same time, been transforming **I** on the right side of the dotted line. Recalling that

$$\mathbf{Ax} = \mathbf{b} \qquad \mathbf{A}^{-1}\mathbf{Ax} = \mathbf{A}^{-1}\mathbf{b} \qquad \mathbf{Ix} = \mathbf{A}^{-1}\mathbf{b}$$

it seems reasonable to suppose that we have been obtaining the inverse of **A**. That is

$$\mathbf{A}^{-1} = \begin{bmatrix} 1 & -2 & 1 \\ -3/5 & 1 & -1/5 \\ 2/5 & 0 & -1/5 \end{bmatrix}$$

First, we observe that

$$\underset{\mathbf{A}^{-1}}{\begin{bmatrix} 1 & -2 & 1 \\ -3/5 & 1 & -1/5 \\ 2/5 & 0 & -1/5 \end{bmatrix}} \underset{\mathbf{b}}{\begin{bmatrix} 14 \\ 13 \\ 13 \end{bmatrix}} = \underset{\mathbf{x}}{\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}}$$

as we know it should. Second, we check to see that

$$\underset{\mathbf{I}}{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}} = \underset{\mathbf{A}}{\begin{bmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \\ 2 & 4 & 1 \end{bmatrix}} \underset{\mathbf{A}^{-1}}{\begin{bmatrix} 1 & -2 & 1 \\ -3/5 & 1 & -1/5 \\ 2/5 & 0 & -1/5 \end{bmatrix}}$$

and the solution is complete.

In summary, application of elementary row operations has served three main purposes:

1. By reducing the matrix to echelon form we were able to discern its rank by simply counting the number of rows containing a nonzero entry.

2. If an explicit set of simultaneous equations is involved—here we assume that a unique solution exists—we can apply elementary row operations to the augmented matrix and obtain, via back substitution, the desired values of the unknowns.

3. If the inverse of the given transformation matrix is also desired, we can apply elementary row operations to both **A** and **I**, transforming the former into **I** and the latter into $\mathbf{A}^{-1}$.

In actuality, elementary row operations (or elementary column operations) have applicability to solving sets of simultaneous equations in more general settings where an inverse of the coefficients matrix may not exist. These matters are taken up in Appendix B.

In summary, from a somewhat more theoretical viewpoint, we can use the echelon approach to reduce any matrix to the following form :

$$\left[\begin{array}{c:c} \mathbf{I}_{k\times k} & \phi \\ \hdashline \phi & \phi \end{array}\right]$$

where the first k rows and k columns represent a $k \times k$ identity matrix (of rank k, of course), and the remaining entries (if any) all consist of zeros.

In the preceding example involving the nonsingular matrix **A**, no zeros appeared since **A** itself was of full rank. In other instances zeros will be found. In general, the above form is found in two steps. Given an arbitrary matrix **B**, we first apply a set of elementary row operations to get the echelon form:

$$\mathbf{H} = \mathbf{FB}$$

where **F** denotes the (nonsingular) matrix product of *all* of the separate elementary matrices used to carry out the reduction of **B** to echelon form. This step has been illustrated above.

Next, we could apply elementary column operations in order to get the matrix product:

$$\mathbf{HG} = \left[\begin{array}{c:c} \mathbf{I}_{k\times k} & \phi \\ \hdashline \phi & \phi \end{array}\right] = \mathbf{FBG}$$

where $r(\mathbf{B}) = k$. In the preceding numerical example it was not necessary to apply elementary column operations after the elementary row operations were performed. In other cases, it may be more efficient to reduce the matrix to echelon form via row operations and then to obtain the form above via elementary column operations that entail postmultiplication of the echelon matrix **H**. At any rate we include the general approach that involves both row and column operations.

The complete transformation that entails row and column operations is also called an equivalence transformation. *More formally, if* **F** *and* **G** *are nonsingular, an equivalence transformation of an arbitrary matrix* **B** *is defined as*

$$\boxed{\mathbf{C} = \mathbf{FBG}}$$

and **C** is defined to be equivalent, via elementary row and column operations, to the given matrix **B**.

4.7.5 The Pivotal Method

Having seen how elementary operations are used to (a) determine rank, (b) solve a set of simultaneous equations, and (c) find the inverse of the coefficients matrix, our discussion of the pivotal method, first described in Chapter 2, can now be completed. For ease of reference, Table 2.2 is reproduced here as Table 4.15.

TABLE 4.15

Evaluating a Determinant by the Pivotal Method

Row no.	Original matrix				Identity matrix				Check sum column
0	1	2	3	4	5	6	7	8	9
01	$\boxed{2}$	3	1	2	1	0	0	0	9
02	4	2	3	4	0	1	0	0	14
03	1	4	2	2	0	0	1	0	10
04	3	1	0	1	0	0	0	1	6
10	1	1.5	0.5	1	0.5	0	0	0	4.5
11		$\boxed{-4}$	1	0	−2	1	0	0	−4
12		2.5	1.5	1	−0.5	0	1	0	5.5
13		−3.5	−1.5	−2	−1.5	0	0	1	7.5
20		1	−0.25	0	0.5	−0.25	0	0	1
21			$\boxed{2.125}$	1	−1.75	0.625	1	0	3.0
22			−2.375	−2	0.25	−0.875	0	1	−4.0
30			1	0.471	−0.824	0.294	0.471	0	1.412
31				$\boxed{-0.881}$	−1.707	−0.177	1.119	1	−0.646
40				1	1.938	0.201	−1.270	−1.135	0.733
30*			1		−1.737	0.199	1.069	0.534	1.067
20*		1			0.066	−0.200	0.267	0.134	1.267
10*	1				−0.668	−0.001	0.334	0.667	1.332

$|\mathbf{A}| = (2)(-4)(2.125)(-0.881) = 15$

In Chapter 2 we illustrated how the pivotal (boxed) entries in Table 4.15 were obtained and how $|\mathbf{A}|$ was found as the product of the four pivotal items. From what we have now learned, we see that the pivotal method is just a systematic way of applying elementary row operations, leading in the example of Table 4.15 to the echelon matrix:

$$\mathbf{H} = \begin{bmatrix} 1 & 1.5 & 0.5 & 1 \\ 0 & 1 & -0.25 & 0 \\ 0 & 0 & 1 & 0.471 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

which, we note, is of full rank where $r(\mathbf{H})$ equals n, its order. Our task now is to explain the back substitution procedure that leads to the inverse of the original 4×4 matrix (call it $\mathbf{A}$) in the top left-hand corner of Table 4.15.

First, if we treat column 5 as a vector of constants, it should be apparent from earlier discussion how we can obtain the following values of x_4, x_3, x_2, x_1 via back substitution:

$$x_4 = 1.938 \qquad \text{(from row 40 and column 5)}$$

$$x_3 + 0.471(1.938) = -0.824 \qquad \text{(from row 30 and column 5)}$$

$$x_3 = -1.737$$

$$x_2 - 0.25(-1.737) + 0(1.938) = 0.5 \qquad \text{(from row 20 and column 5)}$$

$$x_2 = 0.066$$

$$x_1 + 1.5(0.066) + 0.5(-1.737) + 1(1.938) = 0.5 \qquad \text{(from row 10 and column 5)}$$

$$x_1 = -0.668$$

These entries are shown as a column vector in rows 40, 30*, 20*, and 10*, and column 5.

What we are doing here is building up the inverse of **A** one column at a time. Then we repeat the whole procedure, using the entries of column 6, followed by column 7, and finally by column 8. The result of all of this is the inverse $\mathbf{A}^{-1}$, shown in rows 40, 30*, 20*, and 10*, and columns 5, 6, 7, and 8.

However, it is important to note that the inverse $\mathbf{A}^{-1}$ is shown in Table 4.15 in *reverse row order.* That is, the *first* row of $\mathbf{A}^{-1}$ is given by

$$(-0.668 \quad -0.001 \quad 0.334 \quad 0.667)$$

as designated by the row 10*. Hence, if one uses the row order 10*, 20*, 30*, and 40, the inverse $\mathbf{A}^{-1}$ will be in correct row order.

The pivotal procedure proceeds somewhat differently from that used in Section 4.7.4, but the results are the same. Rather than applying additional elementary operations to transform the echelon matrix **H** into identity form, in the pivotal method we simply treat each of the columns—columns 5 through 8—as a separate set of constants and solve for $\mathbf{A}^{-1}$ in this manner. Either way we have the relation

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$$

as we should.

Furthermore, had an explicit set of constants been present, in the usual form of a set of simultaneous equations, the same procedure could have been applied to this column as well. In brief, the pivotal method can be used for the same purposes noted earlier:

1. finding the determinant of **A**,
2. finding the rank of **A**,
3. solving an explicit set of simultaneous equations,
4. finding the inverse of **A** (if it exists).

In applying the pivotal method one should be on the lookout for cases in which a zero appears in the leading diagonal position. If such does occur, we cannot, of course, divide that row by the pivot. If the matrix is nonsingular—which will usually be the case in data-based applications—the presence of a leading zero suggests that we want to move to the next row and select it as the pivot (i.e., permute rows) before proceeding. This has no effect on the inverse, but will reverse the sign of the determinant (if an odd number of such transpositions occur).

If the matrix is singular, the situation will be revealed by the presence of *all zeros* in some row to be pivoted. This is the type of situation that we encountered earlier in transforming a matrix to echelon form.

In summary, the pivotal method is now seen as just a specific step-by-step way to go about applying elementary row operations. While other approaches are available, the pivotal method does exhibit the virtues of simplicity and directness. To round out discussion, we consider its use in finding the inverse of a correlation matrix in the context of multiple regression.

4.7.6 Matrix Inversion in Multiple Regression

The sample data introduced in Table 1.2 involved three variables:

Y number of days absent

X_1 attitude toward the company

X_2 number of years employed by the company

As discussed in Section 1.6.2, the least-squares principle entails finding a linear equation that minimizes the sum of the squared deviations

$$\sum_{i=1}^{12} e_i^2 = \sum_{i=1}^{12} (Y_i - \hat{Y}_i)^2$$

between the original criterion Y_i and the predicted criterion $\hat{Y}_i$.

The linear equation is represented by the form

$$\hat{Y}_i = b_0 + b_1 X_1 + b_2 X_2$$

where b_0, b_1, and b_2 are parameters to be estimated (by least squares) from the data.

The derivation of the normal equations underlying the least-squares procedure is described in Appendix A.

If we work with the original data from the sample problem, the normal equations can be represented in matrix form as

$$(\mathbf{X}'\mathbf{X})\mathbf{b} = \mathbf{X}'\mathbf{y}$$

where $\mathbf{X}$ is the matrix of predictor variable scores (to which has been appended a column vector of unities); $\mathbf{b}$ is the vector of parameters:

$$\mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix}$$

to be solved for, and $\mathbf{y}$ is the criterion vector.

The intercept b_0 denotes the value of Y when X_1 and X_2 are each zero; b_1 measures the change in Y per unit change in X_1 and b_2 measures the change in Y per unit change in X_2.

TABLE 4.16

Computing the Matrix Product Required for Solving the Normal Equations in the Sample Regression Problem

$$\mathbf{X'X} = \begin{bmatrix} 1,1,1,\ldots,1 \\ 1,2,2,\ldots,12 \\ 1,1,2,\ldots,10 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 2 & 2 \\ \vdots & \vdots & \vdots \\ 1 & 12 & 10 \end{bmatrix} = \begin{bmatrix} 12 & 75 & 59 \\ 75 & 639 & 497 \\ 59 & 497 & 397 \end{bmatrix}$$

$$\mathbf{X'y} = \begin{bmatrix} 1,1,1,\ldots,1 \\ 1,2,2,\ldots,12 \\ 1,1,2,\ldots,10 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 1 \\ \vdots \\ 16 \end{bmatrix} = \begin{bmatrix} 75 \\ 702 \\ 542 \end{bmatrix}$$

Table 4.16 illustrates the computations required to obtain $\mathbf{X'X}$ and $\mathbf{X'y}$. To solve for $\mathbf{b}$, the vector of regression parameters, we express the normal equations in the form

$$\boxed{\mathbf{b} = (\mathbf{X'X})^{-1}\mathbf{X'y}}$$

and proceed to solve for $\mathbf{b}$.

Table 4.17 shows the pivotal method applied to the sample problem. In this case the original matrix is only 3×3 so the procedure involves fewer steps; otherwise, the approach is the same as that followed in Table 4.15. After $(\mathbf{X'X})^{-1}$ is found by the pivotal procedure, this is postmultiplied by $\mathbf{X'y}$. The solution vector $\mathbf{b}$ is shown in the lower portion of Table 4.17. The vector of parameters $\mathbf{b}$ was actually obtained via computer and so differs somewhat from that found by carrying only three decimal places in Table 4.17.

The desired linear equation is

$$\hat{Y}_i = -2.263 + 1.550X_1 - 0.239X_2$$

In Chapter 6 we discuss this sample problem in considerably more detail. At this point, however, we simply wish to show how the pivotal procedure can be used to find a matrix inverse in the context of multiple regression analysis.

Matrix inversion represents a central concept in multivariate analysis. While illustrated here in the context of multiple regression, matrix inversion goes well beyond this type of application. It is used in a variety of multivariate techniques, including analysis of variance and covariance, discriminant analysis and canonical correlation, to name some of the procedures. Along with matrix eigenstructures, to be discussed in Chapter 5, matrix inversion represents one of the most important and commonly applied operations in all of multivariate analysis.

TABLE 4.17

Applying the Pivotal Method to the Sample Regression Problem

Row no. 0	Original matrix 1	2	3	Identity matrix 4	5	6	Check sum column 7
01	$\boxed{12}$	75	59	1	0	0	147
02	75	639	497	0	1	0	1212
03	59	497	397	0	0	1	954
10	1	6.250	4.917	0.083	0	0	12.250
11		$\boxed{170.250}$	128.225	−6.225	1	0	293.250
12		128.250	106.897	−4.897	0	1	231.250
20		1	0.753	−0.037	0.006	0	1.722
21			$\boxed{10.325}$	−0.152	−0.770	1	10.403
30			1	−0.015	−0.075	0.097	1.007
20*		1		−0.026	0.062	−0.073	0.963
10*	1			0.320	−0.019	−0.021	1.280

$$\mathbf{b} = \overset{(\mathbf{X}'\mathbf{X})^{-1}}{\begin{bmatrix} 0.320 & -0.019 & -0.021 \\ -0.026 & 0.062 & -0.073 \\ -0.015 & -0.075 & 0.097 \end{bmatrix}} \overset{\mathbf{X}'\mathbf{y}}{\begin{bmatrix} 75 \\ 702 \\ 542 \end{bmatrix}}; \qquad \mathbf{b} = \begin{bmatrix} -2.263 \\ 1.550 \\ -0.239 \end{bmatrix}$$

4.8 SUMMARY

This chapter has focused on linear transformations, a key concept in multivariate analysis. As indicated at the outset of the chapter, all matrix transformations are linear transformations. Furthermore, in two or three dimensions various kinds of matrix transformations can be portrayed geometrically.

We first reviewed transformations involving orthogonal matrices, that is, rotations. We described how one can move a point (or points) relative to a fixed basis or, alternatively, leave the point fixed and rotate the basis vectors. This idea was then extended to linear transformations generally. Numerical examples were used to illustrate the concepts.

We next discussed matrix inversion and its role in solving for the original vector, given some transformed vector and the matrix of the transformation. Allied concepts involving the adjoint of a matrix and the use of the matrix inverse in finding transformations relative to a new set of basis vectors were also discussed. The concept of basis vector transformation was then extended to the case of showing the effect of a given transformation on points referred to two different sets of basis vectors if one knows (a) the matrix of the transformation with respect to one basis and (b) the linear transformation that relates the two sets of basis vectors. Some discussion was also presented on general (oblique) coordinate systems.

The next principal section of the chapter attempted to integrate many of our earlier comments, in this and preceding chapters, by showing what happens geometrically as various kinds of matrix transformations are applied. An important extension of this concept involves the idea of composite mappings and, in particular, the observation that any nonsingular matrix transformation with real-valued entries can be decomposed into the unique product of either a rotation followed by a stretch followed by another rotation, or a rotation followed first by a reflection and then by a stretch that is followed by another rotation. Explanation of this assertion represents one of the main topics of the next chapter.

The next major area of interest concerned the rank of a matrix and its relationship to matrix inversion and linear independence. Various properties of matrix rank were listed and illustrated in the context of linear transformations. As was pointed out, the rank of a transformation matrix is important in determining what characteristics of the original space are preserved under a matrix transformation.

A topic that is related to the foregoing involves the application of elementary operations to determine matrix rank, solving sets of simultaneous equations, and matrix inversion. This subject was taken up next, and various examples of applying elementary operations were worked through and the results integrated with earlier material.

The concluding section dealt with numerical procedures for matrix inversion and emphasized the pivotal method, first introduced in Chapter 2. This example was continued here and, in addition, the pivotal method was applied to inverting a cross-products matrix in the context of multiple regression.

REVIEW QUESTIONS

1. Let g_1 denote a mapping of E^3 (Euclidean 3-space) that represents the projection of a point onto the xy plane; let g_2 denote a mapping of E^3 that represents the projection of a point onto the z axis. Describe the following mappings geometrically:

a. $g_1 + g_2$ b. $-g_2$ c. $g_1 - g_2$ d. $3g_1 + g_2$

2. Use the transformation

$$\mathbf{x}^* = \begin{bmatrix} 1 & 0 \\ 1/2 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

to carry out the following transformations:

a. a triangle with vertices (1, 1), (4, 2), (5, 4)
b. a parallelogram with vertices (1, 1), (4, 2), (5, 4), (2, 3)
c. a rectangle with vertices (1, 2), (2, 2), (2, 5), (1, 5)

By means of diagrams show the original and transformed figures. What aspects—lengths, angles, areas—are invariant over the transformation?

3. Using the same figures as those of the preceding question,

a. show the effect of a stretch with $k_{11} = 2$ and $k_{22} = 3$;
b. show the effect of a shear parallel to the x axis with $c = 2$.

4. Use 2 x 2 matrices to represent the following transformations:

a. A counterclockwise rotation of points through 45° followed by the translation $x_1^* = x_1 + 2; x_2^* = x_2 - 1$.

b. A counterclockwise rotation of points through 30° followed by the stretch $x_1^* = 3x_1; x_2^* = x_2/2$.

c. A stretch $x_1^* = 3x_1; x_2^* = x_2/2$, followed by a counterclockwise rotation of points through a 30° angle.

5. Consider the following transformations:

$$\mathbf{A}_1 = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}; \qquad \mathbf{A}_2 = \begin{bmatrix} 1/2 & 0 \\ 0 & 1/2 \end{bmatrix}; \qquad \mathbf{A}_3 = \begin{bmatrix} 0.707 & 0.707 \\ -0.707 & 0.707 \end{bmatrix}$$

Express in matrix form, and show geometrically, the following composite transformations:

a. $\mathbf{A}_2\mathbf{A}_1$ b. $\mathbf{A}_3\mathbf{A}_1\mathbf{A}_2$ c. $\mathbf{A}_2\mathbf{A}_3$ d. $\mathbf{A}_3\mathbf{A}_2\mathbf{A}_1$

6. Find $\mathbf{A}^{-1}$, if it exists, and verify that $\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$ for the following:

a. $\mathbf{A} = \begin{bmatrix} 5 & 2 \\ 7 & 5 \end{bmatrix}$ b. $\mathbf{A} = \begin{bmatrix} 8 & 4 \\ -4 & -2 \end{bmatrix}$

c. $\mathbf{A} = \begin{bmatrix} 1 & 0 & -1 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}$ d. $\mathbf{A} = \begin{bmatrix} a & -b \\ -c & -d \end{bmatrix}$

7. What is the rank of the following matrices:

a. $\mathbf{A} = \begin{bmatrix} 2 & -1 \\ 3 & 4 \end{bmatrix}$ b. $\mathbf{A} = \begin{bmatrix} 3 & 2 & 4 \\ 2 & 0 & 2 \\ 4 & 2 & 3 \end{bmatrix}$

c. $\mathbf{A} = \begin{bmatrix} 4 & 3 & 1 & 18 \\ 2 & 1 & 3 & 10 \\ 5 & 7 & -2 & 29 \end{bmatrix}$

8. Use elementary row operations to find the echelon form of

a. $\mathbf{A} = \begin{bmatrix} 0 & 5 & 6 & 2 & 0 \\ 1 & 2 & 3 & 9 & 2 \\ 0 & 1 & 2 & 1 & 4 \\ 2 & 0 & 4 & 0 & 1 \end{bmatrix}$ b. $\mathbf{A} = \begin{bmatrix} 1 & 4 & 3 \\ 2 & 6 & 8 \\ 3 & 2 & 1 \end{bmatrix}$

c. $\mathbf{A} = \begin{bmatrix} 2 & 3 & 8 \\ 1 & 1 & 1 \\ 6 & 2 & 3 \\ 4 & 7 & 0 \\ 3 & 2 & 2 \end{bmatrix}$

9. For each of the following sets of equations, determine the rank of the coefficients matrix and the rank of the augmented matrix. Calculate the inverse (if it exists) by elementary row operations.

a.
$$\begin{aligned} 2x_1 + x_2 + 4x_3 &= 13 \\ x_1 - x_2 + x_3 &= 2 \\ x_1 + 2x_2 + 3x_3 &= 10 \end{aligned}$$

b.
$$\begin{aligned} x_1 + 2x_2 + 6x_3 &= 16 \\ x_1 + 3x_2 - x_3 &= 12 \\ x_1 + 2x_2 + 0x_3 &= 10 \end{aligned}$$

c.
$$\begin{aligned} x_1 - 2x_2 + x_3 &= 7 \\ 3x_1 - 2x_2 + 5x_3 &= 43 \\ x_1 + 2x_2 + 3x_3 &= 29 \end{aligned}$$

10. Use the pivotal method to solve the following set of equations; invert the matrix of coefficients and find its determinant:

$$\begin{aligned} 2x_1 + 5x_2 + 3x_3 &= 1 \\ 3x_1 + x_2 + 2x_3 &= 1 \\ x_1 + 2x_2 + x_3 &= 0 \end{aligned}$$

11. Use the adjoint method to invert the 2 x 2 correlation matrix of predictors in the sample problem of Table 1.2 and then perform the indicated multiplication:

$$\mathbf{R} = \begin{array}{c} \\ X_1 \\ X_2 \end{array} \begin{array}{c} \begin{array}{cc} X_1 & X_2 \end{array} \\ \begin{bmatrix} 1.00 & 0.95 \\ 0.95 & 1.00 \end{bmatrix} \end{array}; \qquad \mathbf{b}^* = \mathbf{R}^{-1}\mathbf{r}(y)$$

where $\mathbf{r}(y)$ is the vector of simple correlations between the criterion Y and the predictors X_1 and, X_2, respectively:

$$\mathbf{r}(y) = \begin{bmatrix} 0.95 \\ 0.89 \end{bmatrix}$$

The vector $\mathbf{b}^*$ denotes the standardized regression coefficients $b_1{}^*$ and $b_2{}^*$ (often called beta weights) that measure the change in Y per unit change in X_1 and X_2, respectively, when all variables are expressed in terms of mean zero and unit standard deviation.

12. Let

$$\mathbf{T} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$$

denote the matrix of a point transformation in the basis $\mathbf{E}$. Let the basis vector transformation relating $\mathbf{F}$ to $\mathbf{E}$ be

$$\begin{aligned} \mathbf{f}_1 &= 11\mathbf{e}_1 + 7\mathbf{e}_2 \\ \mathbf{f}_2 &= 3\mathbf{e}_1 + 2\mathbf{e}_2 \end{aligned}$$

Find the point transformation $\mathbf{T}^\circ$ corresponding to $\mathbf{T}$ relative to the basis $\mathbf{F}$.

a. If $\mathbf{x} = [\begin{smallmatrix} 2 \\ 3 \end{smallmatrix}]$ in $\mathbf{E}$, what is it after transformation in $\mathbf{F}$?
b. If $\mathbf{x}^* = [\begin{smallmatrix} 3 \\ 1 \end{smallmatrix}]$ in $\mathbf{E}$, what is its preimage in $\mathbf{E}$?
What is the determinant of $\mathbf{T}$? of $\mathbf{T}^\circ$?

13. A linear transformation relative to the **E** basis has been defined as

$$\begin{aligned} x_1^* &= x_1 + x_2 \\ x_2^* &= x_1 \qquad + x_3 \\ x_3^* &= \qquad x_2 + x_3 \end{aligned}$$

The transformation linking **F**, some new basis, with the original **E** basis is

$$\mathbf{L} = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 1 & 1 \\ 1 & 3 & 1 \end{bmatrix}$$

Find the point transformation defined above in terms of the new basis **F**.

14. Relative to the **E** basis we have the linear transformation:

$$\begin{aligned} x_1^* &= 7x_1 - 2x_2 \\ x_2^* &= -2x_1 + 6x_2 - 2x_3 \\ x_3^* &= \qquad - 2x_2 + 5x_3 \end{aligned}$$

A transformation of **E** to **F** has been found and is represented by

$$\begin{aligned} \mathrm{f}_1 &= \mathrm{e}_1/3 + 2\mathrm{e}_2/3 + 2\mathrm{e}_3/3 \\ \mathrm{f}_2 &= 2\mathrm{e}_1/3 + \mathrm{e}_2/3 - 2\mathrm{e}_3/3 \\ \mathrm{f}_3 &= -2\mathrm{e}_1/3 + 2\mathrm{e}_2/3 - \mathrm{e}_3/3 \end{aligned}$$

What (particularly simple) form does the mapping of x onto x* take with respect to the new basis **F**?