CHAPTER 5

Decomposition of Matrix Transformations: Eigenstructures and Quadratic Forms

5.1 INTRODUCTION

In the preceding chapter we discussed various special cases of matrix transformations, such as rotations, reflections, and stretches, and portrayed their effects geometrically. We also pointed out the geometric effect of various *composite* transformations, such as a rotation followed by a stretch.

The motivation for this chapter is, however, just the opposite of that in Chapter 4. Here we start out with a more or less arbitrary matrix transformation and consider ways of *decomposing* it into the product of matrices that are simpler from a geometric standpoint. As such, our objective is to provide, in part, a set of complementary approaches to those illustrated in Chapter 4.

Adopting this reverse viewpoint enables us to introduce a number of important concepts in multivariate analysis–matrix eigenvalues and eigenvectors, the eigenstructure properties of symmetric and nonsymmetric matrices, the singular value decomposition of a matrix and quadratic forms. This new material, along with that of the preceding three chapters, should provide most of the background for understanding vector and matrix operations in multivariate analysis. Moreover, we shall examine concepts covered earlier, such as matrix rank, matrix inverse, and matrix singularity, from another perspective–one drawn from the context of eigenstructures.

Finding the eigenstructure of a square matrix, like finding its inverse, is almost a routine matter in the current age of computers. Nevertheless, it seems useful to discuss the kinds of computations involved even though we limit ourselves to small matrices of order 2×2 or 3×3. In this way we can illustrate many of these concepts geometrically as well as numerically.

Since the topic of eigenstructures can get rather complex, we start off the chapter with an overview discussion of eigenstructures in which the eigenvalues and eigenvectors can be found simply and quickly. Emphasis here is on describing the *geometric* aspects of eigenstructures as related to special kinds of basis vector changes that render the nature of the mapping as simple as possible, for example, as a stretch relative to the appropriate set of basis vectors.

This simple and descriptive treatment also enables us to tie in the present material on eigenstructures with the discussion in Chapter 4 that centered on point and basis vector transformations. In so doing, we return to the numerical example shown in Section 4.3 and obtain the eigenstructure of the transformation matrix described there.

The next main section of the chapter continues the discussion of eigenstructures, but now in the context of multivariate analysis. To introduce this complementary approach–one based on finding a linear composite such that the variance of point projections onto it is maximal–we return to the small numerical problem drawn from the sample data of Table 1.2. We assume that we have a set of mean-corrected scores of twelve employees on X_1 (attitude toward the company) and X_2 (number of years employed by the company). The problem is to find a linear composite of the two separate scores that exhibits maximum variance across individuals. This motivation leads to a discussion of matrix eigenstructures involving *symmetric* matrices and the multivariate technique of principal components analysis.

The next main section of the chapter deals with various properties of matrix eigenstructures. The more common case of symmetric matrices (with real-valued entries) is discussed in some detail, while the more complex case involving eigenstructures of nonsymmetric matrices is described more briefly. The relationship of eigenstructure to matrix rank is also described here.

The singular value decomposition of a matrix either square or rectangular and its relationship to matrix decompositon is another central concept in multivariate procedures. Accordingly, attention is centered on this topic, and the discussion is also related to material covered in Chapter 4. Here, however, we focus on the *decomposition* of matrices into the product of other matrices that individually exhibit rather simple geometric interpretations.

Quadratic forms are next taken up and related to the preceding material. Moreover, additional discussion about the eigenstructure of square nonsymmetric matrices, as related to such multivariate techniques as multiple discriminant analysis and canonical correlation, is presented in the context of the third sample problem in Chapter 1.

Thus, if matrix inversion and matrix rank are important in linear regression and related procedures for studying single criterion, multiple predictor association, matrix eigenstructures and quadratic forms are the essential concepts in dealing with multiple criterion, multiple predictor relationships.

5.2 AN OVERVIEW OF MATRIX EIGENSTRUCTURES

In Chapter 4 we spent a fair amount of time discussing point and basis vector transformations. In particular, in Section 4.3.5 we discussed the problem of finding the transformation matrix $\mathbf{T}^\circ$, relative to some basis $\mathbf{F}$, if we know the transformation matrix $\mathbf{T}$, denoting the same mapping relative to the $\mathbf{E}$ basis. As shown, to find $\mathbf{T}^\circ$ requires that we know $\mathbf{L}$ the transformation that connects $\mathbf{F}$ with $\mathbf{E}$. We can then find $\mathbf{T}^\circ$ from the equation

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

While the discussion at that point may have seemed rather complex, it was pointed out that this procedure for changing basis vectors has practical utility in cases where we are able to find some *special* basis $\mathbf{F}$ in which the matrix (analogous to $\mathbf{T}^\circ$ above) of the linear transformation takes on some particularly simple form, such as a stretch or a stretch followed by a reflection.

The development of a special basis, in which a linear transformation assumes a simple (i.e., diagonal) form, is the motivation for this section of the chapter. As it turns out, if such a basis exists, it will be found from the *eigenstructure* of a matrix that is analogous to $\mathbf{T}$ above.[1] Moreover, the (diagonal) matrix that represents the same transformation relative to the new basis will also be found at the same time. In all cases we assume that the original matrix of the transformation is square (with real-valued entries, of course).

By way of introduction to matrix eigenstructures, let us first take up an even simpler situation than that covered in Section 4.3.5. Assume that we have a 2 x 2 transformation matrix:

$$\mathbf{A} = \begin{bmatrix} -3 & 5 \\ 4 & -2 \end{bmatrix}$$

Next, suppose we wished to find an image vector

$$\mathbf{x}^* = \begin{bmatrix} x_1{}^* \\ x_2{}^* \end{bmatrix}$$

that has the same (or, possibly, precisely the opposite) direction as the preimage vector

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

If we are concerned only with maintaining direction, then $\mathbf{x}^*$ the image vector can be represented by

$$\mathbf{x}^* = \begin{bmatrix} \lambda x_1 \\ \lambda x_2 \end{bmatrix} = \lambda \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

where λ denotes a scalar. That is, we can stretch or compress $\mathbf{x}$, the preimage, in any way we wish as long as $\mathbf{x}^*$ is in the same (or precisely the opposite) direction as $\mathbf{x}$.

If $\mathbf{x}$ is transformed by $\mathbf{A}$ into $\mathbf{x}^* = \lambda\mathbf{x}$, we state the following:

Vectors, which under a given transformation map into themselves or multiples of themselves, are called invariant vectors under that transformation.

It follows, then, that such vectors obey the relation

$$\boxed{\mathbf{A}\mathbf{x} = \lambda\mathbf{x}}$$

where, as noted, λ is a scalar.

[1] What we shall call eigenvalues (and eigenvectors) some authors call characteristic roots (vectors) or latent roots (vectors).

To illustrate, suppose we try out the vector $\mathbf{x}_1 = \left[\begin{smallmatrix}2\\3\end{smallmatrix}\right]$ to see whether this is invariant under $\mathbf{A}$:

$$\begin{array}{ccc} \mathbf{A} & \mathbf{x}_1 & \mathbf{x}_1^* \\ \begin{bmatrix} -3 & 5 \\ 4 & -2 \end{bmatrix} & \begin{bmatrix} 2 \\ 3 \end{bmatrix} = & \begin{bmatrix} 9 \\ 2 \end{bmatrix} \end{array}$$

Such is not the case. We see that the relationship for an invariant vector does not hold, since the components of the vector $\mathbf{x}_1^* = \left[\begin{smallmatrix}9\\2\end{smallmatrix}\right]$ are not constant multiples of the vector $\mathbf{x}_1 = \left[\begin{smallmatrix}2\\3\end{smallmatrix}\right]$. However, let us next try the vector $\mathbf{x}_2 = \left[\begin{smallmatrix}3\\3\end{smallmatrix}\right]$.

$$\begin{array}{ccc} \mathbf{A} & \mathbf{x}_2 & \mathbf{x}_2^* \\ \begin{bmatrix} -3 & 5 \\ 4 & -2 \end{bmatrix} & \begin{bmatrix} 3 \\ 3 \end{bmatrix} = & \begin{bmatrix} 6 \\ 6 \end{bmatrix} = 2 \begin{bmatrix} 3 \\ 3 \end{bmatrix} \end{array}$$

In the case of $\mathbf{x}_2 = \left[\begin{smallmatrix}3\\3\end{smallmatrix}\right]$, we *do* have an invariant vector. Moreover, if we try *any* vector in which the components are in the ratio 1 : 1, we would find that the relation is also satisfied. For example,

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

where $\lambda = 2$ is the constant of proportionality.

Is it the case that only preimage vectors of the form $\mathbf{x}_i = \left[\begin{smallmatrix}k\\k\end{smallmatrix}\right]$ satisfy the relation? Let us try another vector, namely, $\mathbf{x}_3 = \left[\begin{smallmatrix}5\\-4\end{smallmatrix}\right]$:

$$\begin{array}{ccc} \mathbf{A} & \mathbf{x}_3 & \mathbf{x}_3^* \\ \begin{bmatrix} -3 & 5 \\ 4 & -2 \end{bmatrix} & \begin{bmatrix} 5 \\ -4 \end{bmatrix} = & \begin{bmatrix} -35 \\ 28 \end{bmatrix} = -7 \begin{bmatrix} 5 \\ -4 \end{bmatrix} \end{array}$$

We see that this form $\mathbf{x}_j = \left[\begin{smallmatrix}5k\\-4k\end{smallmatrix}\right]$, works also. But are there others? As we shall see, there are no others that are not of the form of either

$$\begin{bmatrix} k \\ k \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} 5k \\ -4k \end{bmatrix}$$

To delve somewhat more deeply into the problem, let us return to the matrix equation

$$\boxed{\mathbf{A}\mathbf{x} = \lambda\mathbf{x}}$$

which can be rearranged (by subtracting $\lambda\mathbf{x}$ from both sides) as follows:

$$\mathbf{A}\mathbf{x} - \lambda\mathbf{x} = \mathbf{0}$$

Or equivalently,

$$\mathbf{A}\mathbf{x} - \lambda\mathbf{I}\mathbf{x} = \mathbf{0}$$

where $\mathbf{I}$ is an identity matrix. Next, we can factor out $\mathbf{x}$ to get

$$(\mathbf{A}-\lambda\mathbf{I})\mathbf{x} = \mathbf{0}$$

As can now be seen, the problem of finding an invariant vector $\mathbf{x}$ is reduced to the problem of solving the equation

$$\boxed{(\mathbf{A}-\lambda\mathbf{I})\mathbf{x} = \mathbf{0}}$$

One trivial solution is, of course, to let $\mathbf{x} = \left[\begin{smallmatrix}0\\0\end{smallmatrix}\right]$. Generally, however, we would be interested in nontrivial solutions; that is, solutions in which $\mathbf{x} \neq \left[\begin{smallmatrix}0\\0\end{smallmatrix}\right]$.

For the moment, let us set $\mathbf{A} - \lambda\mathbf{I}$ equal to $\mathbf{B}$ and examine what is implied about $\mathbf{B}$ if $\mathbf{x}$ is to be nontrivial (i.e., contain nonzero elements). The above expression can then be written as

$$\mathbf{Bx} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

which, in turn, can be written as the set of simultaneous linear equations:

$$ax_1 + bx_2 = 0$$

$$cx_1 + dx_2 = 0$$

After multiplying the first equation by d, the second by $-b$, and adding the two, we have

$$(ad-bc)x_1 = 0$$

We then repeat the process by multiplying the first equation by $-c$, the second by a, and adding the two, to get

$$(ad-bc)x_2 = 0$$

So, if *either* $x_1 \neq 0$ or $x_2 \neq 0$, we must have the situation in which

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = |\mathbf{B}| = |(\mathbf{A}-\lambda\mathbf{I})| = 0$$

What this all says is that the determinant of $\mathbf{A} - \lambda\mathbf{I}$ *must be zero* if we wish to allow x_1 and x_2 to be nonzero.

5.2.1 The Characteristic Equation

Returning to the original expression of $\mathbf{A} - \lambda\mathbf{I}$, the above reasoning says that we want the determinant of this matrix to be zero. We can write out the above matrix explicitly as

$$\mathbf{A}-\lambda\mathbf{I} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

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

Then, we can find the determinant and set it equal to zero:[2]

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

This last expression is called the *characteristic equation* of the transformation matrix **A**. The roots of this equation, which shall be denoted by λ_i, are called *eigenvalues*, and their associated vectors $\mathbf{x}_i$ are obtained by substituting the roots in

$$(\mathbf{A} - \lambda_i \mathbf{I})\mathbf{x}_i = \mathbf{0}$$

and solving for $\mathbf{x}_i$. These vectors $\mathbf{x}_i$ are called *eigenvectors*. They are the vectors that are invariant under transformation by the matrix **A**. That is, by setting up the format of the characteristic equation and then solving for its roots and associated vectors, we have an operational procedure for finding the invariant vectors of interest. We obtain two central results from the process:

1. the eigenvalues λ_i that indicate the magnitude of the stretch (or stretch followed by reflection), and
2. the eigenvectors $\mathbf{x}_i$ that indicate the new directions (basis vectors) along which the stretching or compressing takes place.

In the case of a 2 x 2 matrix, not more than two values of λ_i are possible. We can see this from the fact that the characteristic equation is quadratic, and a quadratic equation has two solutions, or roots. In general, if **A** is $n \times n$, n roots are possible, since a polynomial of degree n is involved.

As indicated above, the characteristic equation of **A** is defined as

$$\boxed{|\mathbf{A} - \lambda_i \mathbf{I}| = 0}$$

The determinant itself is defined as

$$\boxed{|\mathbf{A} - \lambda_i \mathbf{I}|}$$

and is called the *characteristic function* of **A**.

It should be clear, then, that only square matrices have eigenstructures, since we know already that only square matrices have determinants. Moreover, since $\mathbf{Ax} = \lambda\mathbf{x}$, **A** must be square.

5.2.2 A Numerical and Geometric Illustration

Now that we have concerned ourselves with the rationale for finding the eigenstructure (i.e., the eigenvalues and eigenvectors) of a square matrix, let us apply the procedure to the illustrative matrix shown earlier:

$$\mathbf{A} = \begin{bmatrix} -3 & 5 \\ 4 & -2 \end{bmatrix}$$

[2] The reader should note that the characteristic equation is a polynomial of degree n (given that **A** is $n \times n$). It will have, in general, n roots, not all of which may be either real or, even if real, distinct. We consider these possibilities in due course.

First we write

$$(\mathbf{A}-\lambda_i\mathbf{I})\mathbf{x}_i = \mathbf{0}$$

$$\left\{\begin{bmatrix} -3 & 5 \\ 4 & -2 \end{bmatrix} - \lambda_i \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right\} \begin{bmatrix} x_{1i} \\ x_{2i} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Next, we set up the characteristic equation

$$\begin{vmatrix} -3-\lambda_i & 5 \\ 4 & -2-\lambda_i \end{vmatrix} = 0$$

and expand the determinant to get

$$\lambda_i^2 + 5\lambda_i - 14 = 0$$

We then find the roots of this quadratic equation by simple factoring:

$$(\lambda_i + 7)(\lambda_i - 2) = 0$$

$$\lambda_1 = -7$$

$$\lambda_2 = 2$$

Next, let us substitute $\lambda_1 = -7$ in the equation $(\mathbf{A}-\lambda_i\mathbf{I})\mathbf{x}_i = \mathbf{0}$:

$$\left\{\begin{bmatrix} -3 & 5 \\ 4 & -2 \end{bmatrix} - \begin{bmatrix} -7 & 0 \\ 0 & -7 \end{bmatrix}\right\} \begin{bmatrix} x_{11} \\ x_{21} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 4 & 5 \\ 4 & 5 \end{bmatrix} \begin{bmatrix} x_{11} \\ x_{21} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

The obvious solution to the two equations, each of which is

$$4x_{11} + 5x_{21} = 0$$

is the vector

$$\mathbf{x}_1 = \begin{bmatrix} 5 \\ -4 \end{bmatrix}$$

or, as illustrated earlier, more generally,

$$\mathbf{x}_1 = \begin{bmatrix} 5k \\ -4k \end{bmatrix}$$

Next, let substitute $\lambda_2 = 2$ in the same way, so as to find

$$\left\{\begin{bmatrix} -3 & 5 \\ 4 & -2 \end{bmatrix} - \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}\right\} \begin{bmatrix} x_{12} \\ x_{22} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} -5 & 5 \\ 4 & -4 \end{bmatrix} \begin{bmatrix} x_{12} \\ x_{22} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

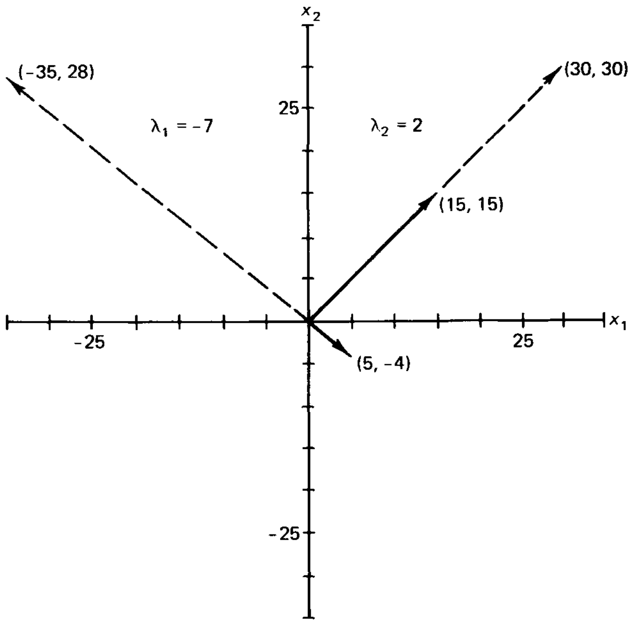

Fig. 5.1 Vectors that are invariant under the transformation matrix **A**.

A solution to these two equations:

$$-5x_{12} + 5x_{22} = 0$$

$$4x_{12} - 4x_{22} = 0$$

is evidently the vector

$$\mathbf{x}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

or, again more generally,

$$\mathbf{x}_2 = \begin{bmatrix} 1k \\ 1k \end{bmatrix}$$

Hence, insofar as $\mathbf{x}_1$ and $\mathbf{x}_2$ are concerned, any vector whose components are in the ratio of either

$$5 : -4 \qquad \text{or} \qquad 1 : 1$$

represents an eigenvector of the transformation given by the matrix **A**.

Figure 5.1 shows the geometric aspects of the preceding computations.[3] If we consider

[3] For ease of presentation, in Fig. 5.1 we let

$$\mathbf{x}_2 = \begin{bmatrix} k \times 1 \\ k \times 1 \end{bmatrix} = \begin{bmatrix} 15 \times 1 \\ 15 \times 1 \end{bmatrix}$$

so that the stretch $\lambda \mathbf{x}_2 = 2\left[\begin{smallmatrix} 15 \\ 15 \end{smallmatrix}\right] = \left[\begin{smallmatrix} 30 \\ 30 \end{smallmatrix}\right]$ is prominent on the figure.

the eigenvector $\mathbf{x}_1 = \left[\begin{smallmatrix}5\\-4\end{smallmatrix}\right]$, we see that this is mapped onto

$$\lambda_1 \begin{bmatrix} x_{11} \\ x_{21} \end{bmatrix} = -7 \begin{bmatrix} 5 \\ -4 \end{bmatrix} = \begin{bmatrix} -35 \\ 28 \end{bmatrix}$$

while the second eigenvector $\mathbf{x}_2 = \left[\begin{smallmatrix}k\times 1\\k\times 1\end{smallmatrix}\right] = \left[\begin{smallmatrix}15\times 1\\15\times 1\end{smallmatrix}\right]$ is mapped onto

$$\lambda_2 \begin{bmatrix} x_{12} \\ x_{22} \end{bmatrix} = 2 \begin{bmatrix} 15 \\ 15 \end{bmatrix} = \begin{bmatrix} 30 \\ 30 \end{bmatrix}$$

Furthermore, the eigenvalues $\lambda_1 = -7$ and $\lambda_2 = 2$ represent stretch (or stretch followed by reflection) constants.

5.2.3 Diagonalizing the Transformation Matrix

Let us return to the two eigenvectors found above and next place them in a matrix, denoted by $\mathbf{U}$:

$$\mathbf{U} = \begin{bmatrix} 5 & 1 \\ -4 & 1 \end{bmatrix}$$

As noted, the two *column* vectors above are the invariant vectors of $\mathbf{A}$. We now ask the question: *How would* $\mathbf{A}$ *behave if one chose as a basis for the space the two eigenvectors, now denoted by* $\mathbf{u}_1$ *and* $\mathbf{u}_2$*, the columns of* $\mathbf{U}$?

As we shall show numerically, if $\mathbf{U}$ is chosen as a new basis of the transformation, originally represented by the matrix $\mathbf{A}$ relative to the standard $\mathbf{E}$ basis, then the new transformation matrix is a stretch. This is represented by the diagonal matrix $\mathbf{D}$, given by the expression

$$\boxed{\mathbf{D} = \mathbf{U}^{-1}\mathbf{A}\mathbf{U}}$$

If $\mathbf{U}$ and $\mathbf{D}$ can be found, we say that $\mathbf{A}$ is *diagonalizable* via $\mathbf{U}$. The matrix $\mathbf{U}$ consists of the eigenvectors of $\mathbf{A}$, and the matrix $\mathbf{D}$ is a diagonal matrix whose entries are the eigenvalues of $\mathbf{A}$. Note, then, that $\mathbf{U}$ must be nonsingular, and we must find its inverse $\mathbf{U}^{-1}$.

Recalling material from Chapter 4, we know that we can find the inverse of $\mathbf{U}$ in the 2 x 2 case simply from the determinant and the adjoint of $\mathbf{U}$:

$$\mathbf{U}^{-1} = \frac{1}{|\mathbf{U}|}\ \mathrm{adj}(\mathbf{U}) = \frac{1}{9}\begin{bmatrix} 1 & -1 \\ 4 & 5 \end{bmatrix} = \begin{bmatrix} 1/9 & -1/9 \\ 4/9 & 5/9 \end{bmatrix}$$

Next, we form the triple product

$$\mathbf{D} = \mathbf{U}^{-1}\mathbf{A}\mathbf{U} = \begin{bmatrix} 1/9 & -1/9 \\ 4/9 & 5/9 \end{bmatrix} \begin{bmatrix} -3 & 5 \\ 4 & -2 \end{bmatrix} \begin{bmatrix} 5 & 1 \\ -4 & 1 \end{bmatrix} = \begin{bmatrix} -7 & 0 \\ 0 & 2 \end{bmatrix}$$

We see that the transformation matrix $\mathbf{A}$, when premultiplied by $\mathbf{U}^{-1}$ and postmultiplied by $\mathbf{U}$, the matrix whose columns represent its eigenvectors, *has* been transformed into a diagonal matrix **D** with entries given by the eigenvalues of **A**. *That is, if a set of basis vectors given by* **U** *is employed, the transformation, represented by* **A**, *behaves as a stretch, or possibly as a stretch followed by a reflection, relative to this special basis of eigenvectors.*

We shall be coming back to this central result several times in the course of elaborating upon matrix eigenstructures. The point to remember here is that we have found an instance where, by appropriate choice of basis vectors, a given linear transformation takes on a particularly simple form. This search for a basis, in which the nature of the transformation is particularly simple, represents the primary motivation for presenting the material of this section.

Next, let us look at the expression

$$\boxed{\mathbf{D} = \mathbf{U}^{-1}\mathbf{A}\mathbf{U}}$$

somewhat more closely. First of all, we are struck by the resemblance of this triple product to the triple product

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

described in Section 4.3.5. There we found that $\mathbf{T}^{\circ}$ denoted the point transformation of a vector $\mathbf{x}^{\circ}$, referred to a basis $\mathbf{f}_i$, onto a vector $\mathbf{x}^{*\circ}$, also referred to **F**. $\mathbf{T}^{\circ}$ can be found if we know **T**, the matrix of the same linear mapping with respect to the original basis $\mathbf{e}_i$, and **L**, the matrix of the transformation linking the $\mathbf{f}_i$ basis to the $\mathbf{e}_i$ basis.

Note in the present case that **D** plays the role of $\mathbf{T}^{\circ}$, **A** plays the role of **T**, and **U** plays the role of $\mathbf{L}'$. As such, the analogy is complete. Since **U** is the transpose of the matrix used to find the two linear combinations with respect to the standard basis $\mathbf{e}_i$:

$$\mathbf{f}_1 = 5\mathbf{e}_1 - 4\mathbf{e}_2; \qquad \mathbf{f}_1 = 5\begin{bmatrix}1\\0\end{bmatrix} - 4\begin{bmatrix}0\\1\end{bmatrix}$$

$$\mathbf{f}_2 = 1\mathbf{e}_1 + 1\mathbf{e}_2; \qquad \mathbf{f}_2 = 1\begin{bmatrix}1\\0\end{bmatrix} + 1\begin{bmatrix}0\\1\end{bmatrix}$$

we see that the analogy does, indeed, hold. The current material thus provides some motivation for recapitulating, and extending, the discussion of point and basis vector transformations in Chapter 4.

5.2.4 Point and Basis Vector Transformations Revisited

Suppose we now tie in directly the current material on the special basis vectors (eigenvectors) obtained by finding the eigenstructure of a matrix to the material covered in Section 4.3.5. There we set up the transformation matrix

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

and the vector $\mathbf{x} = [\frac{1}{2}]$ with respect to the original $\mathbf{e}_i$ basis.

We also considered the transformation matrix

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

which denoted point transformations with respect to the $\mathbf{e}_i$ basis according to the transformation $\mathbf{L}$. As recalled, when we wish to find some new basis $\mathbf{f}_i$ with respect to $\mathbf{e}_i$, we use the transpose of $\mathbf{L}$:

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

Note, then, that we must keep in mind the distinction between a linear transformation τ and its matrix representation with respect to a *particular* basis. By way of review, Fig. 5.2 shows the geometric aspects of the mapping:

$$\mathbf{x}^{*\circ} = \mathbf{T}^\circ \mathbf{x}^\circ = (\mathbf{L}')^{-1}\,\mathbf{T}\mathbf{L}'\mathbf{x}^\circ = \begin{bmatrix} 1.11 & 0.60 \\ 0.34 & 0.59 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 2.31 \\ 1.52 \end{bmatrix}$$

where

$$\mathbf{L}' = \begin{bmatrix} 0.83 & 0.20 \\ 0.55 & 0.98 \end{bmatrix}; \qquad (\mathbf{L}')^{-1} = \begin{bmatrix} 1.39 & -0.28 \\ -0.78 & 1.18 \end{bmatrix}$$

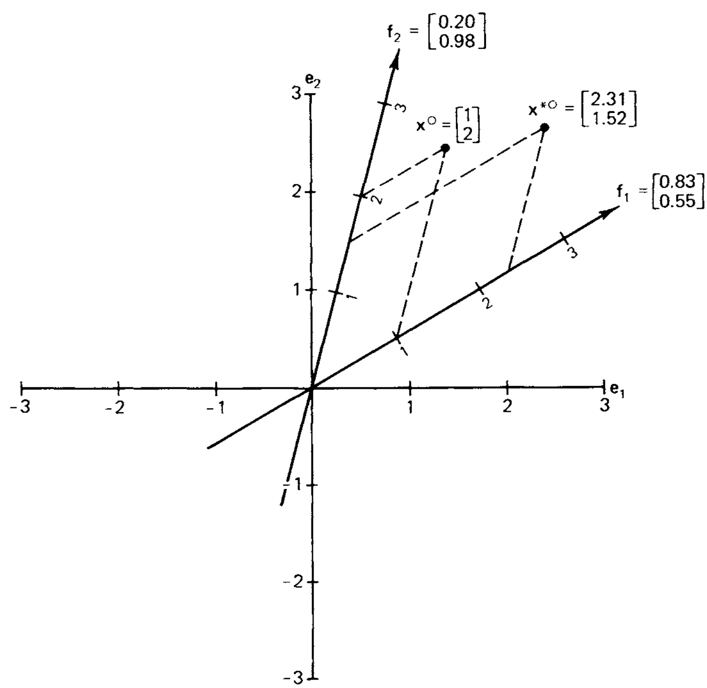

Fig. 5.2 Geometric aspects of the transformation $\mathbf{x}^{*\circ} = \mathbf{T}^\circ \mathbf{x}^\circ$.

in which the vector $\mathbf{x}^{\circ}$ is mapped onto $\mathbf{x}^{*\circ}$ by the point transformation $\mathbf{T}^{\circ}$. As recalled $\mathbf{T}^{\circ}$ is the matrix of the transformation with respect to the $\mathbf{f}_i$ basis (given knowledge of $\mathbf{T}$, the transformation matrix with respect to the $\mathbf{e}_i$ basis), and $\mathbf{L}'$ is the matrix connecting basis vectors in $\mathbf{F}$ with those in $\mathbf{E}$.

In the present context, $\mathbf{U}$ plays the role of $\mathbf{L}'$. Hence, to bring in the new material, we shall want to find the eigenstructure of $\mathbf{T}$. Without delving into computational details, we simply state that the eigenstructure of $\mathbf{T}$ is found in just the same way as already illustrated for the matrix $\mathbf{A}$. In the case of $\mathbf{T}$, the decomposition, as derived from its eigenstructure, is

$$\mathbf{D} = \mathbf{U}^{-1}\mathbf{T}\mathbf{U}$$

$$\begin{array}{ccccc} \mathbf{D} & & \mathbf{U}^{-1} & \mathbf{T} & \mathbf{U} \\ \begin{bmatrix} 1.37 & 0 \\ 0 & 0.66 \end{bmatrix} & = & \begin{bmatrix} -0.80 & -0.62 \\ 0.74 & -0.70 \end{bmatrix} & \begin{bmatrix} 0.9 & 0.44 \\ 0.6 & 0.8 \end{bmatrix} & \begin{bmatrix} -0.69 & 0.61 \\ -0.73 & -0.79 \end{bmatrix} \end{array}$$

Next, in line with the recapitulation in Fig. 5.2, we find the transformation

$$\mathbf{x}^{*\circ} = \mathbf{U}^{-1}\ \mathbf{T}\mathbf{x}^{\circ} = \mathbf{D}\mathbf{x}^{\circ} = \begin{bmatrix} 1.37 & 0 \\ 0 & 0.33 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1.37 \\ 0.66 \end{bmatrix}$$

Figure 5.3 shows the pertinent results from a geometric standpoint. First, we note that the columns of $\mathbf{U}$ appear as the new basis vectors denoted $\mathbf{f}_1$ and $\mathbf{f}_2$, respectively, so as to maintain the analogy with the column vectors of $\mathbf{L}'$ in Fig. 5.2.

First, $\mathbf{U}$ takes $\mathbf{x}^{\circ}$ onto $\mathbf{x}$ with respect to $\mathbf{E}$. Then, the transformation $\mathbf{T}$ takes $\mathbf{x}$ onto $\mathbf{x}^*$ with respect to $\mathbf{E}$. Finally, $\mathbf{U}^{-1}$ takes $\mathbf{x}^*$ onto $\mathbf{x}^{*\circ}$ with respect to $\mathbf{F}$, the matrix of the new basis. The interesting aspect of the exercise, however, is that the $\mathbf{f}_i$ basis (given by $\mathbf{U}$ in the present context) is not just any old basis; rather, it is one in which the mapping of $\mathbf{x}^{\circ}$ onto $\mathbf{x}^{*\circ}$ involves a stretch as given by the transformation

$$\begin{array}{ccccc} & \mathbf{D} & \mathbf{x}^{\circ} & & \mathbf{x}^{*\circ} \\ \mathbf{x}^{*\circ} = \mathbf{D}\mathbf{x}^{\circ} = & \begin{bmatrix} 1.37 & 0 \\ 0 & 0.33 \end{bmatrix} & \begin{bmatrix} 1 \\ 2 \end{bmatrix} & = & \begin{bmatrix} 1.37 \\ 0.66 \end{bmatrix} \end{array}$$

Figure 5.3 shows the point transformation from $\mathbf{x}^{\circ}$ to $\mathbf{x}^{*\circ}$ with respect to the $\mathbf{f}_i$ basis.[4] In one sense, then, the eigenstructure problem is precisely analogous to finding the nature of a transformation relative to two different sets of basis vectors. And this is one reason why the latter topic was discussed in Section 4.3.5. *However, the distinguishing feature of an eigenvector basis is that the nature of the transformation assumes a particularly simple geometric form, such as the stretch noted above.*

[4] As shown in Fig. 5.3, $\mathbf{x}^{\circ} = \left[\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right]$ is stretched along the $\mathbf{f}_1$ axis in the ratio 1.37 : 1 but compressed along the $\mathbf{f}_2$ axis in the ratio 0.67 : 2 (or 0.33 : 1).

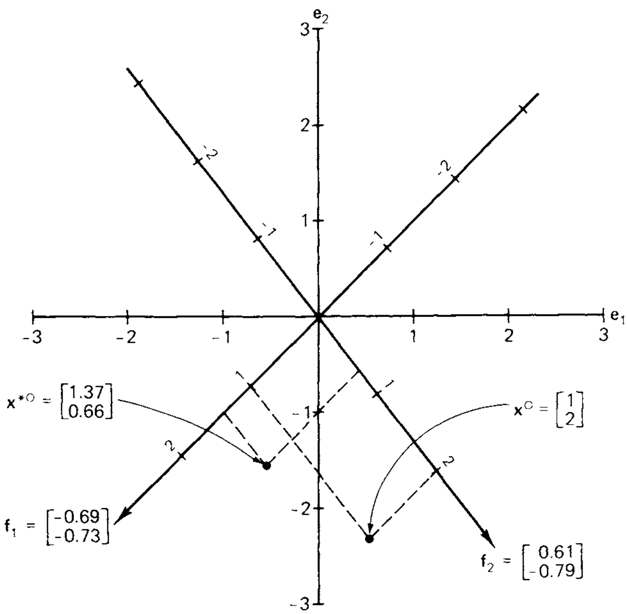

Fig. 5.3 Basis vector transformation involving eigenvectors of **T**.

5.2.5 Recapitulation

In summary, obtaining the eigenstructure of a (square) matrix entails solving the characteristic equation

$$|\mathbf{A}-\lambda_i\mathbf{I}| = 0$$

If **A** is of order $n \times n$, then we shall obtain n roots of the equation; these roots are called eigenvalues. Each eigenvalue can then be substituted in

$$(\mathbf{A}-\lambda_i\mathbf{I})\mathbf{x}_i = \mathbf{0}$$

to obtain its associated eigenvector.

But what about the eigenvalues (and eigenvectors) of some arbitrary matrix **A**? All we have said up to this point is that if **A** is $n \times n$, then n eigenvalues and eigenvectors are obtained. However, we shall find it is possible that

1. some, or all, of the eigenvalues are complex, rather than real valued (even though **A** is real valued);
2. some, or all, of the eigenvectors have complex elements;
3. even if all eigenvalues (and their eigenvectors) are real, some eigenvalues may be zero;
4. even if all eigenvalues are real and nonzero, some may be repeated.

Moreover, so far we have not said very much about the new basis of column eigenvectors in **U**, other than to indicate that it must be nonsingular in order for the relationship

$$\mathbf{D} = \mathbf{U}^{-1}\mathbf{A}\mathbf{U}$$

to hold. Furthermore, by the following algebraic operations:

$$\mathbf{U}\mathbf{D} = \mathbf{U}\mathbf{U}^{-1}\mathbf{A}\mathbf{U} = \mathbf{A}\mathbf{U}$$

and

$$\mathbf{U}\mathbf{D}\mathbf{U}^{-1} = \mathbf{A}\mathbf{U}\mathbf{U}^{-1}$$

we can express **A** as

$$\boxed{\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{U}^{-1}}$$

Other than the conditions that **U** is nonsingular and **D** is diagonal, however, no further requirements have been imposed. We might well wonder if situations exist in which **U** turns out to be orthogonal as well as nonsingular. Also, we recall that the illustrative matrices, whose eigenstructures were found above, are not symmetric. Do special properties exist in the case of symmetric transformation matrices?

We shall want to discuss these questions, and related ones as well, as we proceed to a consideration of eigenstructures in the context of multivariate analysis.

Here a complementary approach to the study of eigenstructures is adopted. Emphasis is now placed on *symmetric* matrices and the role that eigenstructures play in reorienting an original data space with correlated dimensions to uncorrelated axes, along which the objects are maximally separated, that is, display the highest variance. While it may seem that we are starting on a brand-new topic, it turns out that we are still interested in basis vector changes in order to achieve certain simplifications in the transformation. Hence we shall return to the present topic in due course, but now in the context of *symmetric* transformation matrices. As it turns out, the eigenstructure of a symmetric matrix displays properties that can be described more simply than those associated with the nonsymmetric case. Accordingly, we cover this simpler case first and then proceed to the situation involving nonsymmetric matrices.

5.3 TRANSFORMATIONS OF COVARIANCE MATRICES

At the end of Chapter 1, a small sample problem with hypothetical data was introduced in order to illustrate some of the more commonly used multivariate tools. The basic data, shown in Table 1.2, have already been employed as a running example to show

1. how matrix notation can be used to summarize various statistical operations in a concise manner (Chapter 2);
2. the geometric aspects of such statistical measures as standard deviations, correlations, and the generalized dispersion of a covariance matrix (Chapter 3);
3. how the pivotal method can be used to find matrix inverses and solutions to sets of simultaneous equations (Chapter 4).

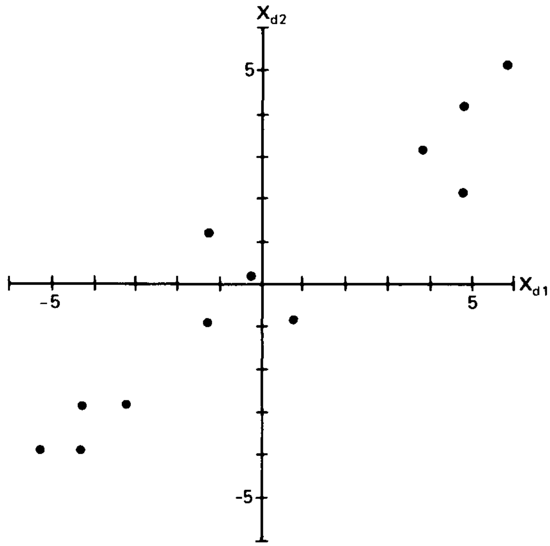

Fig. 5.4 Scatter plot of mean-corrected predictors from sample problem of Chapter 1.

In this chapter we continue to use this small data bank for expository purposes.

Suppose we were to start with a plot of the mean-corrected scores of the two predictors X_{d2} (years employed by the company) versus X_{d1} (attitude toward the company). Figure 5.4 shows the scatter plot obtained from the mean-corrected data of Table 1.2.

We note that X_{d2} and X_{d1} are positively associated (their correlation is 0.95). In line with the discussion of Chapter 1, suppose we wished to replace X_{d1} and X_{d2} by a single linear composite

$$\boxed{z_i = t_1 X_{di1} + t_2 X_{di2}}$$

for $i = 1, 2, \ldots, 12$, the total number of employees, so that the variance of this linear composite

$$\boxed{\mathrm{Var}(z_i) = \sum_{i=1}^{12} (z_i - \bar{z})^2/12}$$

subject to $\mathbf{t}'\mathbf{t} = 1$, is maximized. By this is meant that the twelve employees are maximally separated along the linear composite.

The motivation for doing this sort of thing is often based on parsimony, that is, the desire to replace two or more correlated variables with a single linear composite that, in a sense, attempts to account for as much as possible of the variation shared by the contributory variables. The vector of weights **t** is constrained to be of unit length so that $\mathrm{Var}(z_i)$ cannot be made indefinitely large by making the entries of **t** arbitrarily large.

If our desire is to maximize the variance of the linear composite, how should the weights t_1 and t_2 be chosen so as to bring this about?

To answer this question we need, of course, some kind of criterion. For example, one approach might be to bring the external variable Y into the problem and choose t_1 and t_2 so as to result in a linear composite whose values maximally correlate with Y. As was shown in our discussion of multiple regression in Chapter 4, this involves finding a set of predicted values $\hat{Y}_i$ whose sum of squared deviations from Y_i is minimized.

However, in the present case, let us assume that we choose some "internal" criterion that ignores the external variable Y. The approach suggested earlier is to find a vector that maximizes the variance (or a quantity proportional to this, such as the sum of squares) of the twelve points if they are projected onto this vector. This is also equivalent to minimizing the sum of the squared distances between all pairs of points in which one member of each pair is the to-be-found projection and the other member is the original point.

It is relevant to point out that we are really concerned with two types of vectors. The vector $\mathbf{t} = \left[\begin{smallmatrix} t_1 \\ t_2 \end{smallmatrix}\right]$ is a vector of direction cosines or direction numbers in terms of the original basis. The vector **z** comprises the particular point projections whose variance we are trying to maximize through the particular choice of **t**.

We have, of course, several possible candidates for measuring the original association between X_{d1} and X_{d2}, such as **S**, the SSCP matrix, **C**, the covariance matrix (which is proportional to **S**), and **R**, the correlation matrix.

Illustratively, let us develop the argument in terms of the covariance matrix which, in the sample problem, is

$$\mathbf{C} = \begin{array}{c} \\ X_1 \\ X_2 \end{array}\!\!\begin{array}{c} \begin{array}{cc} X_1 & \quad X_2 \end{array} \\ \begin{bmatrix} 14.19 & 10.69 \\ 10.69 & 8.91 \end{bmatrix} \end{array}$$

Since, as it turns out, we shall be finding two linear composites, we shall refer to these new variables as $\mathbf{z}_1$ and $\mathbf{z}_2$, respectively.

Finding the first of these linear composites $\mathbf{z}_1$ represents the motivation for introducing a new use for computing the eigenstructure of a matrix. That is, we shall wish to find a vector in the space shown in Fig. 5.4 with the property of maximizing the variance of the twelve points if they are projected onto this vector. We might then wish to find a second vector in the same space that obeys certain other properties. If so, what we shall be doing is changing the original basis vectors to a *new* set of basis vectors. (These new basis vectors are often called principal axes.) And, in the course of doing this, it will turn out that we are also decomposing the covariance matrix into the product of simpler matrices from a geometric standpoint in just the same spirit as described in Section 5.2.

5.4 EIGENSTRUCTURE OF A SYMMETRIC MATRIX

Let us now focus on the covariance matrix of the two predictors

$$\begin{array}{cc} & \begin{array}{cc} X_1 & X_2 \end{array} \\ \mathbf{C} = \begin{array}{c} X_1 \\ X_2 \end{array} & \begin{bmatrix} 14.19 & 10.69 \\ 10.69 & 8.91 \end{bmatrix} \end{array}$$

The first thing to be noticed about $\mathbf{C}$ is that it is not only square but also symmetric. Many derived matrices in multivariate analysis exhibit these characteristics. We now wish to consider various linear composites of X_{d1} and X_{d2} that have some chance of maximally separating individuals.

Suppose, arbitrarily, we consider the following, overly simple, linear combinations of $\mathbf{x}_{d1}$ and $\mathbf{x}_{d2}$:

$$\mathbf{z}_1 = 0.707\mathbf{x}_{d1} + 0.707\mathbf{x}_{d2} = 0.707(\mathbf{x}_{d1} + \mathbf{x}_{d2})$$

$$\mathbf{z}_2 = 0.707\mathbf{x}_{d2} - 0.707\mathbf{x}_{d1} = 0.707(\mathbf{x}_{d2} - \mathbf{x}_{d1})$$

In the case of $\mathbf{z}_1$ we are giving equal weight to $\mathbf{x}_{d1}$ and $\mathbf{x}_{d2}$, while in the case of $\mathbf{z}_2$ we are concerned with their difference, that is, the "increment" (component by component) of $\mathbf{x}_{d2}$ over $\mathbf{x}_{d1}$. Notice, further, that (a) the transformation matrix representing these linear combinations, which we shall call $\mathbf{T}$, is orthogonal and (b) we shall be postmultiplying each point represented as a row vector in $\mathbf{X}_d$ by the matrix

$$\mathbf{T} = \begin{bmatrix} 0.707 & -0.707 \\ 0.707 & 0.707 \end{bmatrix}$$

As surmised, 0.707 is chosen so that $(0.707)^2 + (0.707)^2 = 1$, and we have a set of direction cosines. Since $\mathbf{T}$ is orthogonal, the following relationships hold:

$$\mathbf{T}'\mathbf{T} = \mathbf{T}\mathbf{T}' = \mathbf{I}$$

Finally, we also note that the determinant $|\mathbf{T}| = 1$. Hence, a proper rotation, as a matter of fact, a 45° rotation, is entailed since $\cos 45° = 0.707$. We first ask: Suppose one were to consider the mean-corrected matrix $\mathbf{X}_d$ of the two predictors. What is the relationship of the derived matrix $\mathbf{Z}$, found by the rotation $\mathbf{X}_d\mathbf{T}$, to the matrix $\mathbf{X}_d$?

Panel I of Table 5.1 shows the linear composites for $\mathbf{z}_1$ and $\mathbf{z}_2$, respectively, as obtained from the 45° rotation matrix. Since we shall be considering a second transformation subsequently, we use the notation $\mathbf{Z}_a$ and $\mathbf{T}_a$ to denote the particular rotation (involving a 45° angle) that is now being applied.

The solid-line vectors, $\mathbf{z}_{1a}$ and $\mathbf{z}_{2a}$, of Fig. 5.5 show the results of rotating the original basis vectors 45° counterclockwise. If we project the twelve points onto the new basis vectors $\mathbf{z}_{1a}$ and $\mathbf{z}_{2a}$, we find the projections shown in the matrix $\mathbf{Z}_a$ in Panel I of Table 5.1. Note further that, within rounding error, the mean of each column of $\mathbf{Z}_a$ is zero. (In general, if a set of vectors is transformed by a linear function, we will find that the means of the transformed vectors are given by that linear function applied to the means of the original variables.) A more interesting aspect of the transformation is: What is the relationship of the *new* covariance matrix, derived from the transformed matrix $\mathbf{Z}_a$, to that derived from the original matrix $\mathbf{X}_d$?

TABLE 5.1

Linear Composites Based on 45° and 38° Rotations of Original Basis Vectors

I 45° rotation						II 38° rotation					
$\mathbf{Z}_a$		$\mathbf{X}_d$		$\mathbf{T}_a$		$\mathbf{Z}_b$		$\mathbf{X}_d$		$\mathbf{T}_b$	
−6.48	0.94	−5.25	−3.92	0.707	−0.707	−6.55	0.15	−5.25	−3.92	0.787	−0.617
−5.78	0.23	−4.25	−3.92	0.707	0.707	5.76	−0.46	−4.25	−3.92	0.617	0.787
−5.07	0.94	−4.25	−2.92			−5.15	0.32	−4.25	−2.92		
−4.36	0.23	−3.25	−2.92			−4.36	−0.29	−3.25	−2.92		
−1.53	0.23	−1.25	−0.92			−1.55	0.05	−1.25	−0.92		
0.12	1.65	−1.25	1.08			−0.32	1.62	−1.25	1.08		
−0.12	0.23	−0.25	0.08			−0.15	0.09	−0.25	0.08		
−0.12	−1.19	0.75	−0.92			0.02	−1.19	0.75	−0.92		
4.83	0.47	3.75	3.08			4.85	0.11	3.75	3.08		
4.83	−1.89	4.75	2.08			5.02	−1.29	4.75	2.08		
6.24	0.47	4.75	4.08			6.26	0.28	4.75	4.08		
7.66	−0.47	5.75	5.08			7.66	0.45	5.75	5.08		
Variances						Variances					
22.23	0.87					22.56	0.54				

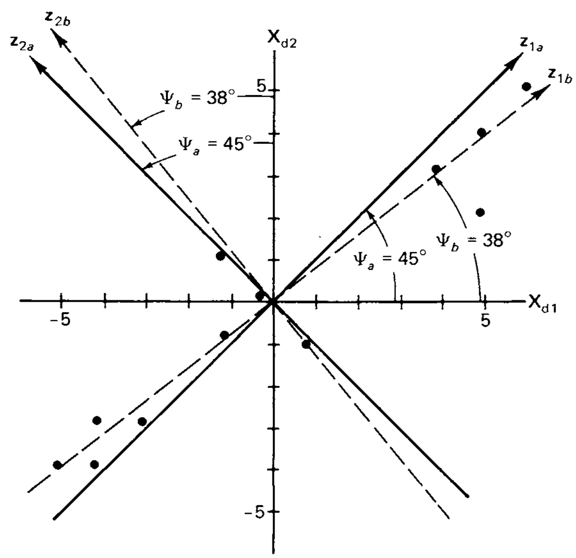

Fig. 5.5 Applying 45° and 38° counterclockwise rotations to axes of Fig. 5.4.

We recall from Chapter 2 that the covariance matrix can be computed from the raw-data matrix $\mathbf{X}$ by the expression

$$\mathbf{C} = 1/m[\mathbf{X}'\mathbf{X} - (1/m)(\mathbf{X}'\mathbf{1})(\mathbf{1}'\mathbf{X})]$$

where $\mathbf{X}'\mathbf{X}$ is called the minor product moment of the raw-data matrix $\mathbf{X}$, and the second term in the brackets is the correction for means.

In the present case by using $\mathbf{X}_d$ the mean of each column is already zero and similarly so for the columns of the transformed matrix $\mathbf{Z}_a$. Hence, the second term on the right of the above equation consists of a null matrix. The original covariance matrix can then be restated as

$$\mathbf{C}(X) = \begin{bmatrix} 14.19 & 10.69 \\ 10.69 & 8.91 \end{bmatrix}$$

5.4.1 The Behavior of the Covariance Matrix under Linear Transformation

If $\mathbf{C}(X)$ is computed as shown above, we could find $\mathbf{C}(Z_a)$ by the same procedure, namely,

$$\mathbf{C}(Z_a) = 1/m\, \mathbf{Z}_a'\mathbf{Z}_a$$

since, as before, the mean of each column of $\mathbf{Z}_a$ (within rounding error) is also zero and, hence $1/m(\mathbf{Z}_a'\mathbf{1})(\mathbf{1}'\mathbf{Z}_a) = \phi$.

Actually, however, a much more direct way to find $\mathbf{C}(Z_a)$ is available. Since $\mathbf{Z}_a = \mathbf{X}_d\mathbf{T}_a$ and $\mathbf{C}(Z_a) = 1/m\, \mathbf{Z}_a'\mathbf{Z}_a$, we have

$$\mathbf{C}(Z_a) = [(\mathbf{X}_d\mathbf{T}_a)'(\mathbf{X}_d\mathbf{T}_a)]/m = [\mathbf{T}_a'\mathbf{X}_d'\mathbf{X}_d\mathbf{T}_a]/m = \mathbf{T}_a'\,[\mathbf{C}(X)]\,\mathbf{T}_a$$

That is, we can find $\mathbf{C}(Z_a)$ *through knowledge of the transformation matrix* $\mathbf{T}_a$ *and* $\mathbf{C}(X)$, *the covariance matrix prior to transformation.*

To find $\mathbf{C}(Z_a)$ we compute

$$\mathbf{C}(Z_a) = \overset{\mathbf{T}_a'}{\begin{bmatrix} 0.707 & 0.707 \\ -0.707 & 0.707 \end{bmatrix}} \overset{\mathbf{C}(X)}{\begin{bmatrix} 14.19 & 10.69 \\ 10.69 & 8.91 \end{bmatrix}} \overset{\mathbf{T}_a}{\begin{bmatrix} 0.707 & -0.707 \\ 0.707 & 0.707 \end{bmatrix}}$$

$$= \begin{bmatrix} 22.23 & -2.64 \\ -2.64 & 0.87 \end{bmatrix}$$

The first thing to be noticed about $\mathbf{C}(Z_a)$ is that the sum of the main diagonal entries (22.23 + 0.87) is, within rounding, equal to the sum of the main diagonal entries of $\mathbf{C}(X)$. The second thing of interest is that $\mathbf{C}(Z_a)$ continues to remain symmetric, but now *the off-diagonal entries are much smaller in absolute value* (2.64) *than their counterpart entries* (10.69) *in* $\mathbf{C}(X)$.

What has happened here is that the rotation of axes, effected by $\mathbf{T}$, has resulted in a new set of basis vectors in which projections on the first of the new axes display a *considerably larger variance* than either contributing dimension x_{d1} or x_{d2}.

We might next inquire if we can do still better in variance maximization. Does some *other* rotation result in a still higher sum of squares for the *first* transformed dimension? What we could do, of course, is to proceed by brute force. Based on what we have seen so far, we could try other orthogonal transformations in the vicinity of a 45° angle and duly note how the parceling out of variance between z_1 and z_2 behaves under each transformation. Fortunately, however, an analytical approach is available—one that again utilizes the concept of the eigenstructure of a matrix.

5.4.2 The Characteristic Equation

In preceding chapters we have talked about changing the basis of a vector space. We have also discussed transformations that involve a rotation (proper and improper) and a stretch. Finally, we know that a linear combination such as $z_1 = 0.707X_{d1} + 0.707X_{d2}$ can be so expressed that the sum of the squared weights equals unity. That is, we can–and have done so in the preceding illustration–normalize the weights of the linear combination so that they appear as direction cosines.

Suppose we take just one column vector of some new rotation matrix $\mathbf{T}$ (for the moment we drop the subscript for ease of exposition) and wish to maximize the expression

$$\mathbf{t}_1'[\mathbf{C}(X)]\mathbf{t}_1$$

subject to the normalization constraint that $\mathbf{t}_1'\mathbf{t}_1 = 1$. This restriction on the length of $\mathbf{t}_1$ will ensure that our transformation meets the unit length condition for a rotation. And, incidentally, this restriction will also ensure that the resultant scalar $\mathbf{t}_1'[\mathbf{C}(X)]\mathbf{t}_1$ cannot be made arbitrarily large by finding entries of $\mathbf{t}_1$ with arbitrarily large values.

The above problem is a more or less standard one in the calculus, namely, optimizing

$$F = \mathbf{t}_1'[\mathbf{C}(X)]\mathbf{t}_1 - \lambda(\mathbf{t}_1'\mathbf{t}_1 - 1)$$

where λ is an additional unknown in the problem, called a *Lagrange multiplier.*

While we shall not go into details (see Appendix A for these), we can briefly sketch out their nature by differentiating F with respect to the elements of $\mathbf{t}_1$ and setting this partial derivative equal to the zero vector. The appropriate (symbolic) derivative is

$$\frac{\partial F}{\partial \mathbf{t}_1} = 2[\mathbf{C}(X)\mathbf{t}_1 - \lambda\mathbf{t}_1]$$

Setting this expression equal to the zero vector, dividing both sides by 2, and factoring out $\mathbf{t}_1$ leads to

$$[\mathbf{C}(X) - \lambda\mathbf{I}]\mathbf{t}_1 = \mathbf{0}$$

where $\mathbf{I}$ is the identity matrix. In terms of our specific problem, we have

$$\begin{array}{ccccc} \mathbf{C}(X)-\lambda\mathbf{I} & & \mathbf{t}_1 & & \mathbf{0} \\ \begin{bmatrix} 14.19-\lambda & 10.69 \\ 10.69 & 8.91-\lambda \end{bmatrix} & & \begin{bmatrix} t_{11} \\ t_{21} \end{bmatrix} & = & \begin{bmatrix} 0 \\ 0 \end{bmatrix} \end{array}$$

We may also recall from the calculus that satisfaction of the above equation is a necessary condition for a maximum (or minimum).[5]

Immediately we are struck by the resemblance of the above expression to the matrix equation of Section 5.2:

$$\boxed{(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}}$$

[5] Although not shown here, sufficiency conditions are also met.

that was used in finding the eigenstructure of **A**. Indeed, the only basic differences here are that $\mathbf{C}(X)$ is symmetric, and the eigenvectors are to be normalized to unit length.

If the matrix $\mathbf{C}(X) - \lambda\mathbf{I}$, for a fixed value of λ, were nonsingular (i.e., possessed an inverse), it would always be the case that the only possible solution to the equation involves setting $\mathbf{t}_1$ equal to the zero vector. Hence, in line with the discussion of Section 5.2, we want to accomplish *just the opposite*, namely, to find a λ that will make $\mathbf{C}(X) - \lambda\mathbf{I}$ *singular.* But we recall that singular matrices have determinants of zero. Hence we want to find a value for λ that satisfies the characteristic equation:

$$\boxed{|\mathbf{C}(X) - \lambda\mathbf{I}| = 0}$$

Another way of putting things is that we wish to find $\mathbf{t}_1$ such that

$$\boxed{\mathbf{C}(X)\mathbf{t}_1 = \lambda\mathbf{t}_1}$$

where $\mathbf{t}_1$ is a vector, which if premultiplied by $\mathbf{C}(X)$, results in a vector $\lambda\mathbf{t}_1$ whose components are proportional to those of $\mathbf{t}_1$.

This, of course, is the same line of reasoning that we discussed earlier in the chapter in the context of finding a new basis in which the matrix of the transformation in terms of that new basis could be denoted by a stretch or, possibly, by a stretch followed by a reflection.

As recalled, however, for an $n \times n$ matrix one obtains n roots in solving the characteristic equation. Since we wish to maximize F, we shall be on the lookout for the largest λ_i obtained from solving the characteristic equation. *That is, we shall order the roots from large to small and choose that eigenvector* $\mathbf{t}_i$ *corresponding to the largest* λ_i.

Either the approach described in Section 5.2 or the current approach leads to the same type of result so long as we remember to *order* the roots of $|\mathbf{C}(X) - \lambda| = 0$ from large to small.[6] As observed, $|\mathbf{C}(X) - \lambda| = 0$ is simply the characteristic equation of the covariance matrix $\mathbf{C}(X)$.

Now, while we initially framed the problem in terms of a single eigenvector $\mathbf{t}_1$ and a single eigenvalue λ_1, the characteristic equation, as formulated above, will enable us to solve for two eigenvalues λ_1 and λ_2, and two eigenvectors $\mathbf{t}_1$ and $\mathbf{t}_2$. As already noted in Section 5.2, if $\mathbf{C}(X)$ is of order $n \times n$, we shall obtain n eigenvalues and n associated eigenvectors.

5.4.3 Finding the Eigenvalues and Eigenvectors of C(X)

It is rather interesting that following either (a) the (present) variance-maximizing path or (b) the basis vector transformation path that seeks a new basis in which vectors are mapped onto scalar multiples of themselves leads to the same result–the characteristic equation. However, let us now concentrate our attention on the variance-maximizing path as the one that appears more appropriate from an intuitive standpoint in the context of the current problem.

[6] It should be mentioned, however, that $\mathbf{C}(X)$, the covariance matrix, exhibits special properties in that it is symmetric and represents the minor product moment of another matrix (in this case $\mathbf{X}_d/\sqrt{m}$). As such, all of its eigenvalues will be real and nonnegative and **T** will be orthogonal. We discuss these special properties later on.

The problem now is to solve for the eigenstructure of $\mathbf{C}(X)$. First, we shall want to find the eigenvalues of the characteristic equation

$$|\mathbf{C}(X)-\lambda_i\mathbf{I}| = \begin{vmatrix} 14.19-\lambda_i & 10.69 \\ 10.69 & 8.91-\lambda_i \end{vmatrix} = 0$$

Expansion of the second-order determinant is quite easy. In terms of the above problem we can express the characteristic equation as

$$(14.19-\lambda_i)(8.91-\lambda_i)-(10.69)^2 = 0$$

$$\lambda_i^2-23.1\lambda_i + 126.433-114.276 = 0$$

$$\lambda_i^2-23.1\lambda_i + 12.157 = 0$$

The simplest way of solving the above equation is to use the quadratic formula of the general form $y = ax^2 + bx + c$. First, we note that the coefficients in the preceding expression are

$$a = 1; \qquad b = -23.1; \qquad c = 12.157$$

Next, let us substitute these in the general quadratic formula:

$$\lambda_i = \frac{-b \pm \sqrt{b^2-4ac}}{2a} = \frac{23.1 \pm \sqrt{(-23.1)^2-4(12.157)}}{2} = \frac{23.1 \pm \sqrt{484.982}}{2}$$

$$\lambda_1 = 22.56; \qquad \lambda_2 = 0.54$$

As could be inferred from the sign of the discriminant $(b^2 - 4ac)$ of the general quadratic, we have a case in which the roots λ_1 and λ_2 are both real and are unequal. If we go back to our original problem of trying to optimize $F[\mathbf{C}(X)]$ subject to $\mathbf{t}'\mathbf{t} = 1$, we see that $\lambda_1 = 22.56$ denotes the *maximum variance achievable* along one dimension by a linear composite of the original basis vectors.

In the 2×2 case, solving for the eigenvectors $\mathbf{t}_1$ and $\mathbf{t}_2$ is rather simple. Let us first substitute the value of $\lambda_1 = 22.56$ in the expression $\mathbf{C}(X) - \lambda\mathbf{I}$:

$$\begin{bmatrix} 14.19-22.56 & 10.69 \\ 10.69 & 8.91-22.56 \end{bmatrix} = \overset{\mathbf{C}(\mathbf{X})-\lambda_1\mathbf{I}}{\begin{bmatrix} -8.37 & 10.69 \\ 10.69 & -13.65 \end{bmatrix}}$$

The next step is to set up the simultaneous equations needed to solve for $\mathbf{t}_1$, the first eigenvector:

$$\overset{\mathbf{C}(X)-\lambda_1\mathbf{I}}{\begin{bmatrix} -8.37 & 10.69 \\ 10.69 & -13.65 \end{bmatrix}} \begin{bmatrix} t_{11} \\ t_{21} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

If we set $t_{21} = 1$ in the first of the two equations:

$$-8.37t_{11} + 10.69t_{21} = 0$$

we obtain $t_{11} = 1.277$. (Note that these values also satisfy the second equation.) This gives us the first (nonnormalized) eigenvector:

$$\mathbf{t}_1 = \begin{bmatrix} 1.277k \\ 1k \end{bmatrix}$$

As we know, we can then divide the components of $\mathbf{t}_1$ by $\|\mathbf{t}_1\| = 1.622$, its length, to get the normalized version:

$$(\text{norm})\mathbf{t}_1 = \begin{bmatrix} 0.787 \\ 0.617 \end{bmatrix}$$

In exactly the same way, we substitute $\lambda_2 = 0.54$ and perform the following calculations:

$$\begin{bmatrix} 14.19-0.54 & 10.69 \\ 10.69 & 8.91-0.54 \end{bmatrix} = \overset{\mathbf{C}(X)-\lambda_2\mathbf{I}}{\begin{bmatrix} 13.65 & 10.69 \\ 10.69 & 8.37 \end{bmatrix}}$$

$$\overset{\mathbf{C}(X)-\lambda_2\mathbf{I}}{\begin{bmatrix} 13.65 & 10.69 \\ 10.69 & 8.37 \end{bmatrix}} \begin{bmatrix} t_{12} \\ t_{22} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Next, we set $t_{22} = 1$ in the first of the two equations:

$$13.65t_{12} + 10.69t_{22} = 0$$

to obtain $t_{12} = -0.783$ and note, further, that these values also satisfy the second equation. We then obtain

$$\mathbf{t}_2 = \begin{bmatrix} -0.783k \\ 1k \end{bmatrix}; \qquad (\text{norm})\mathbf{t}_2 = \begin{bmatrix} -0.617 \\ 0.787 \end{bmatrix}$$

Let us now assemble the two normalized eigenvectors into the matrix, which we shall denote by $\mathbf{T}_b$:

$$\mathbf{T}_b = \begin{bmatrix} 0.787 & -0.617 \\ 0.617 & 0.787 \end{bmatrix}$$

Then, as before, we can collect the various terms of the decomposition into the following triple product:

$$\mathbf{Z}_b = \mathbf{D}_b = \begin{bmatrix} 22.56 & 0 \\ 0 & 0.54 \end{bmatrix} = \mathbf{T}_b^{-1}\mathbf{C}(X)\mathbf{T}_b$$

While we have just found $\mathbf{Z}_b$ and $\mathbf{T}_b$, and we know $\mathbf{C}(X)$ to begin with, we must still solve for $\mathbf{T}_b^{-1}$. We obtain $\mathbf{T}_b^{-1}$ from

$$\mathbf{T}_b = \begin{bmatrix} 0.787 & -0.617 \\ 0.617 & 0.787 \end{bmatrix}$$

by the now-familiar adjoint matrix method, first described in Chapter 4.

$$\mathbf{T}_b^{-1} = \frac{1}{|\mathbf{T}_b|} \operatorname{adj}(\mathbf{T}_b) = \frac{1}{1}\begin{bmatrix} 0.787 & 0.617 \\ -0.617 & 0.787 \end{bmatrix} = \begin{bmatrix} 0.787 & 0.617 \\ -0.617 & 0.787 \end{bmatrix}$$

Here we see the somewhat surprising result that $\mathbf{T}_b^{-1} = \mathbf{T}_b{}'$.

However, as recalled from Chapter 4, one of the properties of an orthogonal matrix is that its inverse equals its transpose:

$$\boxed{\mathbf{T}^{-1} = \mathbf{T}'}$$

and we see that such is the case here. Moreover, it is easily seen that $\mathbf{T}_b$ exhibits the orthogonality conditions of $\mathbf{t}'_{1b}\mathbf{t}_{2b} = 0$ and $\mathbf{t}'_{1b}\mathbf{t}_{1b} = \mathbf{t}'_{2b}\mathbf{t}_{2b} = 1$.

Thus, in the case of a *symmetric* matrix, illustrated by $\mathbf{C}(X)$, the general diagonal representation

$$\boxed{\mathbf{D} = \mathbf{U}^{-1}\mathbf{A}\mathbf{U}}$$

of Section 5.2 now can be written as

$$\boxed{\mathbf{D} = \mathbf{T}'\mathbf{C}(X)\mathbf{T}}$$

where $\mathbf{T}$ is orthogonal and $\mathbf{D}$ continues to be diagonal.

By the same token we can write

$$\boxed{\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{U}^{-1}}$$

in the case of a symmetric matrix as

$$\boxed{\mathbf{C}(X) = \mathbf{T}\mathbf{D}\mathbf{T}'}$$

Thus, the rather interesting outcome of all of this is that starting with a *symmetric* matrix $\mathbf{C}(X)$, we obtain an eigenstructure in which the matrix of eigenvectors is orthogonal. That is, not only is the transformation represented by a stretch (as also found in Section 5.2), but the basis vectors themselves, in the symmetric matrix case, are orthonormal. We shall return to this finding in due course. For the moment, however, let us pull together the results so far, particularly as they relate to the problem of computing some "best" linear composite for the sample problem.

5.4.4 Comparing the Results

When we rather arbitrarily tried a 45° rotation of the original axes in order to obtain the linear composites $\mathbf{Z}_a$ shown in Panel I of Table 5.1, we noted that the covariance matrix derived from this transformation was

$$\mathbf{C}(Z_a) = \mathbf{T}_a'\mathbf{C}(X)\mathbf{T}_a = \begin{bmatrix} 22.23 & -2.64 \\ -2.64 & 0.87 \end{bmatrix}$$

Let us now compare this result with the maximum variance attainable from $\mathbf{T}_b$. From the immediately preceding discussion, we know that the comparable results are

$$\mathbf{C}(Z_b) = \underset{\mathbf{T}_b'}{\begin{bmatrix} 0.787 & 0.617 \\ -0.617 & 0.787 \end{bmatrix}} \underset{\mathbf{C}(X)}{\begin{bmatrix} 14.19 & 10.69 \\ 10.69 & 8.91 \end{bmatrix}} \underset{\mathbf{T}_b}{\begin{bmatrix} 0.787 & -0.617 \\ 0.617 & 0.787 \end{bmatrix}}$$

$$= \begin{bmatrix} 22.56 & 0 \\ 0 & 0.54 \end{bmatrix}$$

The matrix $\mathbf{T}_b$ in the present case involves a counterclockwise rotation of 38°.

Panel II of Table 5.1 shows the computed $\mathbf{Z}_b$ values. Also, Fig. 5.5 shows this optimal rotation as a dashed line for comparison with the 45° rotation tried earlier. Clearly, the two rotations are quite close to each other. Moreover, when we compare the first (selected to be the largest) eigenvalues above:

$$\lambda_1 = 22.23 \qquad \text{(for 45° rotation)}$$

$$\lambda_1 = 22.56 \qquad \text{(for 38° optimal rotation)}$$

we again see how close the 45° rotation, which was just guessed at for expository purposes, is to the optimal.

Another point to note is that the off-diagonal elements of $\mathbf{C}(\mathbf{Z}_b)$, after the optimal rotation, are zero, and hence, $\mathbf{z}_{1b}$ and $\mathbf{z}_{2b}$ the two linear composites obtained from the optimal rotation, are uncorrelated. This is a bonus provided by the fact that $\mathbf{C}(X)$ is symmetric and, hence, $\mathbf{T}_b$ is orthogonal.

To recapitulate, we have demonstrated how a set of vectors represented by the matrix $\mathbf{X}_d$ and a transformation of those vectors $\mathbf{C}(X) = (1/m)\mathbf{X}_d'\mathbf{X}_d$ can be rotated such that the projections of the points $\mathbf{X}_d\mathbf{T}_b$ (with $\mathbf{T}_b$ orthogonal) onto the first axis $\mathbf{z}_{1b}$ have the property of maximum variance. Moreover, at the same time it turns out that the new covariance matrix

$$\mathbf{C}(Z_b) = 1/m\,[\mathbf{T}_b'\mathbf{X}_d'\mathbf{X}_d\mathbf{T}_b] = \begin{bmatrix} 22.56 & 0 \\ 0 & 0.54 \end{bmatrix}$$

is *diagonal.* That is, all vectors referenced according to this new basis will be mapped onto scalar multiples of themselves by a stretch. This means that all cross products in the $\mathbf{C}(Z_b)$ matrix vanish, as noted above.

Hence, we have diagonalized the original transformation $\mathbf{C}(X)$ by finding a rotation $\mathbf{T}_b$ of the $\mathbf{X}_d$ space that has the effect of making $\mathbf{C}(Z_b)$ a diagonal matrix. The second axis $\mathbf{z}_{2b}$ will be orthogonal to the first or variance-maximizing axis $\mathbf{z}_{1b}$.

Reflecting a bit on the above example and observing the configuration of points in Fig. 5.5, it would appear that the point pattern roughly resembles an ellipse. Furthermore, the new axes, $\mathbf{z}_{1b}$ and $\mathbf{z}_{2b}$, correspond, respectively, to the major and minor axes of that ellipse, called principal axes in multivariate analysis.

If the distribution of points is multivariate normal, it turns out that the loci of equal probability are represented by a family of concentric ellipses (in two dimensions) or ellipsoids or hyperellipsoids (in higher dimensions). The "ellipse" in Fig. 5.5 could be construed as an estimate of one of these concentration ellipses.

It also turns out that by solving for the eigenstructure of $\mathbf{C}(X)$, we also obtain the axes of the "ellipse." This reorientation of the plane along the axes of the implied ellipse in Fig. 5.5 (via the 38° rotation of basis vectors) will also be relevant to quadratic forms, a topic that is discussed later in the chapter.

5.5 PROPERTIES OF MATRIX EIGENSTRUCTURES

At this point we have discussed eigenstructures from two different, and complementary, points of view:

1. finding a new basis of some linear transformation so that the transformation relative to that new basis assumes a particularly simple form, such as a stretch or a stretch followed by reflection;
2. finding a new basis—by means of a rotation—so that the variance of a set of points is maximized if they are projected onto the first axis of the new basis; the second axis maximizes residual variance for that dimensionality, and so on.

In the first approach no mention was made of any need for the basis transformation to be symmetric. In the second case the presence of a symmetric matrix possessed the advantage of producing an orthonormal basis of eigenvectors (a rotation).

In the recapitulation of Section 5.2.5, we alerted the reader to a number of problems concerning the eigenstructure of nonsymmetric matrices. In general, even though we assume throughout that $\mathbf{A}$ has all real-valued entries, if $\mathbf{A}$ is nonsymmetric,

1. we may not be able to diagonalize it via $\mathbf{U}^{-1}\mathbf{A}\mathbf{U}$;[7]
2. even if it can be diagonalized, the eigenvalues and eigenvectors of $\mathbf{A}$ need not all be real;
3. even if the eigenvalues (and eigenvectors) of $\mathbf{A}$ are all real valued, they need not be all nonzero;[8]
4. even if they are all nonzero, they need not be all distinct.[9]

[7] However, any matrix *can* be made similar to an upper triangular matrix (a square matrix with all zeros below the main diagonal). We do not pursue this more general topic here.

[8] If $\mathbf{A}$ has at least one zero eigenvalue, it is singular, a point that will be discussed in more detail in Section 5.6.

[9] Points 3 and 4 pertain to symmetric matrices as well.

Delving into the properties of eigenstructures involving complex eigenvalues and eigenvectors would take us too far afield in this book.

Fortunately for the reader all nonsymmetric matrices of interest to us in multivariate analysis will have *real* eigenvalues and *real* eigenvectors. However, if $\mathbf{A}$ is nonsymmetric, then $\mathbf{U}$, the new basis of eigenvectors, is not orthogonal. Moreover, the problem of dealing with zero (or nonzero but nondistinct) eigenvalues must be contended with in any case, and will be discussed in the context of symmetric matrices.

5.5.1 Properties of Symmetric Matrices

As could be inferred from earlier discussion, in order to satisfy the expression

$$\boxed{\mathbf{D} = \mathbf{U}^{-1}\mathbf{A}\mathbf{U}}$$

$\mathbf{U}^{-1}$ must, of course, exist. However, it turns out that in order for some (not necessarily symmetric) matrix $\mathbf{A}_{n \times n}$ to be made diagonal, it is *necessary and sufficient that* $\mathbf{U}$ *consist of linearly independent vectors* and, hence, forms a basis (and possesses an inverse). If this condition is not met, then $\mathbf{A}$ is not diagonalizable. However, even if a matrix is diagonalizable, it may not necessarily be orthogonally so. And, even if a matrix can be made diagonal, it need not consist of eigenvalues and eigenvectors that are all real valued.

Symmetric matrices take care of these problems. If $\mathbf{A}$ is symmetric, it is not only *always* diagonalizable but, in addition, it is *orthogonally* diagonalizable where we have the relation

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

This is a very important condition since it states that for any pair of *distinct* eigenvectors $\mathbf{u}_i$, $\mathbf{u}_j$ their scalar product $\mathbf{u}_i'\mathbf{u}_j = 0$.

Notice that this was the situation in Section 5.4.3 in which we had the result

$$\begin{array}{cccc} & \mathbf{T}_b' & \mathrm{C}(X) & \mathbf{T}_b \\ \mathrm{C}(Z_b) = & \begin{bmatrix} 0.787 & 0.617 \\ -0.617 & 0.787 \end{bmatrix} & \begin{bmatrix} 14.19 & 10.69 \\ 10.69 & 8.91 \end{bmatrix} & \begin{bmatrix} 0.787 & -0.617 \\ 0.617 & 0.787 \end{bmatrix} \end{array}$$

$$= \begin{bmatrix} 22.56 & 0 \\ 0 & 0.54 \end{bmatrix}$$

Not only is $\mathrm{C}(Z_b)$ diagonal, but $\mathbf{T}_b$ is orthogonal. Since orthonormal basis vectors are highly convenient to work with in multivariate analysis, orthogonally diagonalizable matrices are useful to have.

A further differentiating property for symmetric matrices versus their nonsymmetric counterparts is also useful in multivariate applications. *If a symmetric matrix* $\mathbf{A}$ *consists of all real-valued entries, then all of its eigenvalues and associated eigenvectors will be real valued.*

In practice, however, even the nonsymmetric matrices that we encounter in multivariate analysis–such as those that arise in multiple discriminant analysis and canonical correlation–will have real eigenvalues. Hence, in the kinds of applications of relevance to multivariate analysis, the researcher does not need to worry very much about cases involving complex eigenvalues and eigenvectors. Still it is nice to know that the problem of complex eigenvalues and eigenvectors does not arise if $\mathbf{A}$ is symmetric.

Now, let us next examine the case of equal eigenvalues in symmetric matrices. Suppose we have tied λ_i's of multiplicities l_k for blocks $k = 1, 2, \ldots, s$ where

$$\sum_{k=1}^{s} l_k = n$$

First, it is comforting to know that the orthogonality property is maintained *across* subsets of eigenvectors associated with tied eigenvalues. That is, if $\mathbf{t}_i$ and $\mathbf{t}_j$ are drawn from *different* sets, then $\mathbf{t}_i'\mathbf{t}_j = 0$. The problem, then, is to obtain a set of orthogonal eigenvectors *within* each tied set of eigenvalues. Since this can usually be done in an infinity of ways, the solution will not be unique.

To illustrate, suppose we have eigenvalues $\lambda_1 = 1$, $\lambda_2 = 2$, $\lambda_3 = 2$ with associated eigenvectors $\mathbf{t}_1$, $\mathbf{t}_2$, and $\mathbf{t}_3$. We note that λ_2 and λ_3 are tied. In the present case there is, in a sense, too much freedom with regard to the eigenvectors associated with λ_2 and λ_3. What can be done, however, is

1. Find the eigenvector $\mathbf{t}_1$ associated with the distinct eigenvalue λ_1; this is done routinely in the course of substituting λ_1 in the equation $(\mathbf{A} - \lambda_1 \mathbf{I})\mathbf{t}_1 = \mathbf{0}$.
2. Choose eigenvectors $\mathbf{t}_2$ and $\mathbf{t}_3$ (e.g., via Gram–Schmidt orthonormalization) so that they form an orthonormal set *within* themselves. Each, of course, will already be orthogonal to $\mathbf{t}_1$.

While the above orientation is arbitrary, in view of the equality of λ_2 and λ_3, it does represent a way to deal with the problem of subsets of eigenvalues that are equal to each other.

If it turned out that *all* eigenvalues were equal, that is, $\lambda_1 = \lambda_2 = \lambda_3$, then we have a case in which the transformation of $\mathbf{A}$ is a scalar transformation with coefficient λ (i.e., just the scalar λ times an identity transformation). As such, all (nonzero) vectors in the original space can serve as eigenvectors. Thus, *any* set of mutually orthonormal vectors–including the original orthonormal basis that could lead to this condition–can serve as a new basis.

In general, if we have k eigenvalues, all with the same value λ, then we must first find k linearly independent eigenvectors, all having eigenvalue λ. Then we orthonormalize them via some process like Gram–Schmidt. If we have two or more subsets of tied eigenvalues, the orthonormalizing process is done separately within set. As noted earlier, all eigenvectors in different sets, where the eigenvalues differ, will already be orthogonal.

However, tied eigenvalues arise only rarely in data-based product-moment matrices, such as $\mathbf{C}(X)$ and $\mathbf{R}(X)$. However, if they do, the analyst should be aware that the representation of $\mathbf{A}$ in terms of its eigenstructure is not unique, even though $\mathbf{A}$ may be nonsingular.

In summary, if $\mathbf{A}_{n \times n}$ is symmetric, we can say the following:

1. An orthogonal transformation $\mathbf{T}$ can be found such that $\mathbf{A}$ can be made diagonal by

$$\boxed{\mathbf{D} = \mathbf{T}'\mathbf{A}\mathbf{T}}$$

where $\mathbf{D}$ is diagonal, and the columns of $\mathbf{T}$ are eigenvectors of $\mathbf{A}$.

2. All eigenvalues and eigenvectors are real.
3. If $\lambda_i \neq \lambda_j$, then $\mathbf{t}_i'\mathbf{t}_j = 0$.
4. If tied eigenvalues occur, of multiplicity l_k for some block k, then $\mathbf{A}$ has l_k but not more than l_k mutually orthogonal eigenvectors corresponding to the kth block of tied eigenvalues. In such cases, the eigenstructure of $\mathbf{A}$ will not be unique, but $\mathbf{T}$, the $n \times n$ matrix of eigenvectors, will still be nonsingular.

The topic of *zero eigenvalues*—and their relationship to matrix rank—is so important that a special section in the chapter is reserved for it. For the moment, however, we turn to some additional properties of eigenstructures, appropriate (unless stated otherwise) for both the symmetric and nonsymmetric cases.

5.5.2 Additional Properties of Eigenstructures

Two quite general properties of eigenstructures that apply to either the nonsymmetric or symmetric cases are:

1. The sum of the eigenvalues of the eigenstructure of a matrix equals the sum of the main diagonal elements (called the *trace*) of the matrix. That is, for some matrix $\mathbf{W}_{n \times n}$,

$$\boxed{\sum_{i=1}^{n} \lambda_i = \sum_{i=1}^{n} w_{ii}}$$

2. The product of the eigenvalues of $\mathbf{W}$ equals the determinant of $\mathbf{W}$:

$$\boxed{\prod_{i=1}^{n} \lambda_i = |\mathbf{W}|}$$

Notice here that if $\mathbf{W}$ is singular, at least one of its eigenvalues must be zero in order that $|\mathbf{W}| = 0$, the condition that must be met in order for $\mathbf{W}$ to be singular. However, even though $\mathbf{W}$ may be singular, $\mathbf{T}$ in the expression

$$\mathbf{D} = \mathbf{T}'\mathbf{W}\mathbf{T}$$

is still orthogonal (and, hence, nonsingular).

In addition to the above, a number of other properties related to sums, products, powers, and roots should also be mentioned. (The last two properties that are listed pertain only to symmetric matrices with nonnegative eigenvalues.)

3. If we have the matrix $\mathbf{B} = \mathbf{A} + k\mathbf{I}$, where k is a scalar, then the eigenvectors of $\mathbf{B}$ are the same as those of $\mathbf{A}$, and the ith eigenvalue of $\mathbf{B}$ is

$$\boxed{\lambda_i + k}$$

where λ_i is the ith eigenvalue of $\mathbf{A}$.

4. If we have the matrix $\mathbf{C} = k\mathbf{A}$, where k is a scalar, then $\mathbf{C}$ has the same eigenvectors as $\mathbf{A}$ and

$$\boxed{k\lambda_i}$$

is the ith eigenvalue of $\mathbf{C}$, where λ_i is the ith eigenvalue of $\mathbf{A}$.

5. If we have the matrix $\mathbf{A}^p$ (where p is a positive integer), then $\mathbf{A}^p$ has the same eigenvectors as $\mathbf{A}$ and

$$\boxed{\lambda_i{}^p}$$

is the ith eigenvalue of $\mathbf{A}^p$, where λ_i is the ith eigenvalue of $\mathbf{A}$.

6. If $\mathbf{A}^{-1}$ exists, then $\mathbf{A}^{-p}$ has the same eigenvectors as $\mathbf{A}$ and

$$\boxed{\lambda_i{}^{-p}}$$

is the eigenvalue of $\mathbf{A}^{-p}$ corresponding to the ith eigenvalue of $\mathbf{A}$. In particular, $1/\lambda_i$ is the eigenvalue of $\mathbf{A}^{-1}$ corresponding to λ_i, the ith eigenvalue of $\mathbf{A}$.

7. If a symmetric matrix $\mathbf{A}$ can be written as the product

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

where $\mathbf{D}$ is diagonal with all entries nonnegative and $\mathbf{T}$ is an orthogonal matrix of eigenvectors, then

$$\mathbf{A}^{1/2} = \mathbf{TD}^{1/2}\mathbf{T}'$$

and it is the case that $\mathbf{A}^{1/2}\mathbf{A}^{1/2} = \mathbf{A}$.[10]

8. If a symmetric matrix $\mathbf{A}^{-1}$ can be written as the product

$$\mathbf{A}^{-1} = \mathbf{TD}^{-1}\mathbf{T}'$$

where $\mathbf{D}^{-1}$ is diagonal with all entries nonnegative and $\mathbf{T}$ is an orthogonal matrix of eigenvectors, then

$$\mathbf{A}^{-1/2} = \mathbf{TD}^{-1/2}\mathbf{T}'$$

and it is the case that $\mathbf{A}^{-1/2}\mathbf{A}^{-1/2} = \mathbf{A}^{-1}$.

[10] In Chapter 2 the square root of a diagonal matrix was defined as $\mathbf{D}^{1/2}\mathbf{D}^{1/2} = \mathbf{D}$ where the diagonal elements of $\mathbf{D}^{1/2}$ are $\sqrt{d_{ii}}$ and it was assumed that all d_{ii} in $\mathbf{D}$ are nonnegative to begin with. In the present case $\mathbf{A}^{1/2}\mathbf{A}^{1/2} = \mathbf{A}$, where $\mathbf{A}$ need not be diagonal, but the conditions stated in point 7 must be met. In the present context we see that $(\mathbf{TD}^{1/2}\mathbf{T}')(\mathbf{TD}^{1/2}\mathbf{T}') = \mathbf{TD}^{1/2}\mathbf{D}^{1/2}\mathbf{T}' = \mathbf{TDT}' = \mathbf{A}$.

We can illustrate these properties of eigenstructures by means of a simple example:

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

The eigenvalues of $\mathbf{A}$ are $\lambda_1 = 3$ and $\lambda_2 = 1$. The associated (and normalized) eigenvectors are

$$\mathbf{t}_1 = \begin{bmatrix} 0.707 \\ 0.707 \end{bmatrix}; \qquad \mathbf{t}_2 = \begin{bmatrix} 0.707 \\ -0.707 \end{bmatrix}$$

Since $\mathbf{A}$ is symmetric, we can write the decomposition as

$$\mathbf{D} = \mathbf{T}'\mathbf{A}\mathbf{T} = \begin{bmatrix} 0.707 & 0.707 \\ 0.707 & -0.707 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 0.707 & 0.707 \\ 0.707 & -0.707 \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}$$

The various properties, discussed above, are now illustrated:

Trace of $\mathbf{A}$

$$\operatorname{tr}(\mathbf{A}) = a_{11} + a_{22} = \lambda_1 + \lambda_2 = 3 + 1 = 4$$

Determinant of $\mathbf{A}$

$$|\mathbf{A}| = \lambda_1\lambda_2 = 3(1) = 3$$

Eigenvalues of $\mathbf{A} + 2\mathbf{I}$

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

$$\lambda^2 - 8\lambda + 15 = 0$$

$$\lambda_1 = 5; \quad \lambda_2 = 3$$

Eigenvalues of $2\mathbf{A}$

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

$$\lambda^2 - 8\lambda + 12 = 0$$

$$\lambda_1 = 6; \quad \lambda_2 = 2$$

Eigenvalues of $\mathbf{A}^2$

$$\mathbf{A}^2 = \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix}$$

$$\lambda^2 - 10\lambda + 9 = 0$$

$$\lambda_1 = 9 \qquad \lambda_2 = 1$$

Eigenvalues of $\mathbf{A}^{-2}$

$$\mathbf{A}^{-1} = \begin{bmatrix} 2/3 & -1/3 \\ -1/3 & 2/3 \end{bmatrix}; \quad \mathbf{A}^{-2} = \begin{bmatrix} 5/9 & -4/9 \\ -4/9 & 5/9 \end{bmatrix}$$

$$\lambda^2 - 4/3\lambda + 1/3 = 0 \qquad \lambda^2 - 10/9\lambda + 1/9 = 0$$

$$\lambda_1 = 1/3 \quad \lambda_2 = 1 \qquad\qquad \lambda_1 = 1/9 \quad \lambda_2 = 1$$

The Square Root of $\mathbf{A}$

$$\mathbf{A}^{1/2} = \mathbf{T}\mathbf{D}^{1/2}\mathbf{T}' = \begin{bmatrix} 0.707 & 0.707 \\ 0.707 & -0.707 \end{bmatrix} \begin{bmatrix} \sqrt{3} & 0 \\ 0 & \sqrt{1} \end{bmatrix} \begin{bmatrix} 0.707 & 0.707 \\ 0.707 & -0.707 \end{bmatrix}$$

$$= \begin{bmatrix} 1.366 & 0.366 \\ 0.366 & 1.366 \end{bmatrix}; \qquad \mathbf{A}^{1/2}\mathbf{A}^{1/2} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

The Square Root of $\mathbf{A}^{-1}$

$$\mathbf{A}^{-1/2} = \mathbf{T}\mathbf{D}^{-1/2}\mathbf{T}' = \begin{bmatrix} 0.707 & 0.707 \\ 0.707 & -0.707 \end{bmatrix} \begin{bmatrix} 1/\sqrt{3} & 0 \\ 0 & 1/\sqrt{1} \end{bmatrix} \begin{bmatrix} 0.707 & 0.707 \\ 0.707 & -0.707 \end{bmatrix}$$

$$= \begin{bmatrix} 0.788 & -0.211 \\ -0.211 & 0.788 \end{bmatrix}; \qquad \mathbf{A}^{-1/2}\mathbf{A}^{-1/2} = \begin{bmatrix} 2/3 & -1/3 \\ -1/3 & 2/3 \end{bmatrix}$$

$$\begin{array}{ccc} \mathbf{A}^{1/2} & \mathbf{A}^{-1/2} & \mathbf{I} \\ \begin{bmatrix} 1.366 & 0.366 \\ 0.366 & 1.366 \end{bmatrix} & \begin{bmatrix} 0.788 & -0.211 \\ -0.211 & 0.788 \end{bmatrix} = & \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \end{array}$$

The preceding properties are quite useful in various aspects of multivariate analysis, and we shall return to a discussion of some of them later in the chapter.

5.6 EIGENSTRUCTURES AND MATRIX RANK

In Chapter 4 we described two procedures for finding the rank of a matrix, square or rectangular, as the case may be:

1. The examination of various square submatrices in order to find that one with the largest order for which the determinant is nonzero.
2. The echelon matrix approach followed by a count of the number of rows with at least one nonzero entry.

Eigenstructures are computed only for square matrices. However, by some procedures to be described in this section, we shall see how eigenstructures also provide a way to determine the rank of *any* matrix, even if the matrix is rectangular.

In addition, it is now time to discuss the topic of zero eigenvalues in solving for the eigenstructure of a matrix.[11] As noted in Section 5.5, the presence of one or more zero eigenvalues is sufficient evidence that the matrix **A** is singular.

5.6.1 Square Matrices

First, we recall that if $\mathbf{A}_{n \times n}$ is symmetric, then all of its eigenvalues are real. It is possible, of course, that some may be positive, others negative, and some even zero. Also, from the previous section we know that

$$|\mathbf{A}| = \prod_{i=1}^{n} \lambda_i$$

Hence, if any λ_i is zero, **A** is singular. But what about the rank of **A**? Or, if **A** is rectangular, how can its rank be found by means of eigenstructures?

In the case where $\mathbf{A}_{n \times n}$ is symmetric, finding the rank of **A** is simple. We first find the eigenstructure of **A** and then count the number–including multiple values, if any are present–of *nonzero* (either positive or negative) eigenvalues. The number of nonzero eigenvalues of **A** is equal to its rank. Since we assume here that any square matrix (symmetric or nonsymmetric) of interest to us in multivariate analysis will have real eigenvalues–this, of course, must be the case if **A** is symmetric–we can use this same procedure for finding the rank of any **A** as long as **A** is square.

[11] Our remarks will pertain to nonsymmetric as well as symmetric matrices.

5.6.2 Rectangular Matrices

Finding the rank of $\mathbf{A}_{m \times n}$, where $m \neq n$, by means of eigenstructures rests on an important fact about the minor and major product moments of a matrix:

$\mathbf{A}'\mathbf{A}$; minor product moment of $\mathbf{A}$

$\mathbf{A}\mathbf{A}'$; major product moment of $\mathbf{A}$

and that fact is

$$\boxed{r(\mathbf{A}'\mathbf{A}) = r(\mathbf{A}\mathbf{A}') = r(\mathbf{A})}$$

The rank of either a minor or major product moment is the same as the rank of the matrix itself.

Since $r(\mathbf{A}'\mathbf{A}) = r(\mathbf{A}\mathbf{A}')$, we should, of course, find the eigenstructure of the *smaller* of these two orders, so as to reduce computational time and effort. And, if the researcher finds it easier to work with the eigenstructures of symmetric matrices, if $\mathbf{A}_{n \times n}$ is square but nonsymmetric, one can also compute its product moment, either minor or major, and find the eigenstructure of the symmetrized matrix.

Another virtue attaches to finding the eigenstructure of a product moment matrix, $\mathbf{A}'\mathbf{A}$ or $\mathbf{A}\mathbf{A}'$, and that is that *all eigenvalues will be either positive or zero*. In the process of finding the product moment, any negative eigenvalues of $\mathbf{A}$ will become positive in the case of either $\mathbf{A}'\mathbf{A}$ or $\mathbf{A}\mathbf{A}'$, as we shall note later on.

For the moment, however, let us set down the procedure for rank determination in a step-by-step way. First, if $\mathbf{A}$ is originally nonsymmetric or rectangular, we can always find the minor product moment ($\mathbf{A}'\mathbf{A}$) or the major product moment ($\mathbf{A}\mathbf{A}'$) of $\mathbf{A}$, whichever is of smaller order. The product-moment matrix will be square and symmetric. Furthermore, the eigenstructure of the product-moment matrix will exhibit either positive or zero eigenvalues, and a general procedure for rank determination can be followed. This general procedure, applicable to finding the rank of square but nonsymmetric and rectangular matrices alike, is as follows:

1. Find $\mathbf{A}'\mathbf{A}$ or $\mathbf{A}\mathbf{A}'$, whichever is of smaller order in the case of rectangular matrices. Their product will be symmetric, and *all* eigenvalues will be nonnegative. The number of positive eigenvalues of $\mathbf{A}'\mathbf{A}$ (or $\mathbf{A}\mathbf{A}'$) equals the rank of $\mathbf{A}$.
2. If $r(\mathbf{A}) = n$ and $\mathbf{A}$ is $n \times n$, then the vectors, either row or column vectors in $\mathbf{A}$, are linearly independent.
3. If $r(\mathbf{A}) = n$ and $n < m$ (where $\mathbf{A}$ is of order $m \times n$), then the row vectors are linearly dependent. If $r(\mathbf{A}) = m < n$, then the column vectors are linearly dependent.
4. If $r(\mathbf{A}) < n \leqslant m$, then the set of either row or column vectors are linearly dependent and $r(\mathbf{A}) = k$ is the largest number of linearly independent vectors in $\mathbf{A}$. (The number k is the number of positive eigenvalues in $\mathbf{A}'\mathbf{A}$ or $\mathbf{A}\mathbf{A}'$.)

Next, suppose that the symmetric matrix being examined is still of the form $\mathbf{A}'\mathbf{A}$ or $\mathbf{A}\mathbf{A}'$, where we have adopted this form because $\mathbf{A}$ is either rectangular or nonsymmetric. However, even if $\mathbf{A}'\mathbf{A}$ (or $\mathbf{A}\mathbf{A}'$) is nonsingular, some of the (positive) λ_i may be equal to each other. If so, their multiplicities are still counted up in finding the rank of $\mathbf{A}$. If $\mathbf{A}$ (or $\mathbf{A}'\mathbf{A}$ or $\mathbf{A}\mathbf{A}'$) is singular with a subset of l_k nondistinct eigenvalues, we can still find a mutually orthonormal set of eigenvectors of rank l_k by some process, such as the Gram–Schmidt orthonormalization process, for the tied block k.[12]

In summary, finding the eigenstructure of a matrix—either symmetric to begin with or else a derived product-moment matrix—is a highly useful procedure for determining matrix rank. If the $n \times n$ original symmetric matrix has rank $r(\mathbf{A}) = k$, then $k \leqslant n$ nonzero eigenvalues will be found. If the *derived* matrix

$$\mathbf{A}'\mathbf{A} \qquad \text{or} \qquad \mathbf{A}\mathbf{A}' \qquad \text{(whichever is of smaller order)}$$

has $r(\mathbf{A}'\mathbf{A})$ or $r(\mathbf{A}\mathbf{A}')$ equal to k, then $k \leqslant \min(m, n)$ positive eigenvalues will be found. In short, one can *always* find the rank of a matrix via the eigenstructure approach.

5.6.3 Special Characteristics of Product-Moment Matrices

Product-moment matrices, like the SSCP, covariance, and correlation matrices, play a unique role in multivariate analysis. For example, let us return to the covariance matrix used in the eigenstructure problem of Section 5.4:

$$\mathbf{C}(X) = \begin{array}{c} \\ X_1 \\ X_2 \end{array} \begin{array}{c} \begin{array}{cc} X_1 & \quad X_2 \end{array} \\ \begin{bmatrix} 14.19 & 10.69 \\ 10.69 & 8.91 \end{bmatrix} \end{array}$$

We recall that this represents the minor product moment found from

$$\mathbf{C}(X) = (1/m)\mathbf{X}_\mathrm{d}'\mathbf{X}_\mathrm{d}$$

where $\mathbf{X}_\mathrm{d}$ is the matrix of deviations about column means.

By way of summarizing some aspects of matrix rank and their relationship to eigenstructures, let us set down a few properties of product-moment matrices, such as $\mathbf{C}(X)$. We can illustrate the properties in terms of the covariance matrix:

1. If $\mathbf{C}(X)$, the covariance matrix, has all distinct eigenvalues, it can be written uniquely as the triple product

$$\boxed{\mathbf{C}(X) = \mathbf{TDT}'}$$

where $\mathbf{D}$ is diagonal, and $\mathbf{T}$ is an orthogonal matrix of associated eigenvectors.

2. Since $\mathbf{C}(X)$ is of product-moment form, all of its eigenvalues are nonnegative, and we can always order the eigenvalues of $\mathbf{C}(X)$ from large to small.

3. Whether nonsingular or not, the rank of $\mathbf{C}(X)$ equals the number of positive eigenvalues in its eigenstructure (since all eigenvalues in this case are nonnegative).

4. It is generally the case that if *any* square matrix $\mathbf{A}$ is symmetric with nonnegative eigenvalues, then $\mathbf{A} = \mathbf{B}'\mathbf{B}$. One way of writing the relationship involves defining $\mathbf{B}'$ as

$$\boxed{\mathbf{B}' = \mathbf{TD}^{1/2}}$$

[12] In a sense, then, the problem of dealing with tied (but nonzero) eigenvalues is independent of the problem of determining the rank of a matrix.

where $\mathbf{D}^{1/2}$ is a diagonal matrix of the square roots of the eigenvalues of **A**, and **T** is the orthogonal matrix whose columns are the associated eigenvectors of **A**.

5. Since all of the eigenvalues of $\mathbf{C}(X)$ are either positive or zero, there exists a matrix **B** of order $k \times n$ such that

$$\boxed{\mathbf{C}(X) = \mathbf{B}'\mathbf{B}}$$

6. The preceding definition of $\mathbf{B}' = \mathbf{T}\mathbf{D}^{1/2}$ is, however, not unique. If $\mathbf{B}_1 = \mathbf{V}\mathbf{B}$ is the product of an orthogonal matrix **V** and **B**, then **A** can also be written, equally appropriately, as

$$\mathbf{A} = \mathbf{B}_1{}'\mathbf{B}_1 = (\mathbf{V}\mathbf{B})'(\mathbf{V}\mathbf{B}) = \mathbf{B}'\mathbf{V}'\mathbf{V}\mathbf{B}$$

The last three points can be illustrated by returning to the eigenstructure of $\mathbf{C}(X)$ in the sample problem. First, we can write $\mathbf{C}(X)$ as

$$\mathbf{C}(X) = \overset{\mathbf{T}_b}{\begin{bmatrix} 0.787 & -0.617 \\ 0.617 & 0.787 \end{bmatrix}} \overset{\mathbf{D}_b}{\begin{bmatrix} 22.56 & 0 \\ 0 & 0.54 \end{bmatrix}} \overset{\mathbf{T}_b{}'}{\begin{bmatrix} 0.787 & 0.617 \\ -0.617 & 0.787 \end{bmatrix}}$$

Next, we define $\mathbf{B}' = \mathbf{T}_b\ \mathbf{D}_b^{1/2}$ and $\mathbf{B} = \mathbf{D}_b^{1/2}\ \mathbf{T}_b'$ and, furthermore, restate $\mathbf{C}(X)$:

$$\mathbf{B}' = \begin{bmatrix} 3.74 & -0.45 \\ 2.93 & 0.58 \end{bmatrix}; \qquad \mathbf{B} = \begin{bmatrix} 3.74 & 2.93 \\ -0.45 & 0.58 \end{bmatrix}; \qquad \mathbf{B}'\mathbf{B} = \begin{bmatrix} 14.19 & 10.69 \\ 10.69 & 8.81 \end{bmatrix}$$

We can then check to see that $\mathbf{B}'\mathbf{B} = \mathbf{C}(X)$.

Next, however, let us take some orthogonal matrix **V** and write

$$\mathbf{B}_1 = \overset{\mathbf{V}}{\begin{bmatrix} 0.707 & 0.707 \\ -0.707 & 0.707 \end{bmatrix}} \overset{\mathbf{B}}{\begin{bmatrix} 3.74 & 2.93 \\ -0.45 & 0.58 \end{bmatrix}} = \begin{bmatrix} 2.33 & 2.48 \\ -2.96 & -1.66 \end{bmatrix}$$

Then, we can write

$$\mathbf{C}(X) = \overset{\mathbf{B}_1{}'}{\begin{bmatrix} 2.33 & -2.96 \\ 2.48 & -1.66 \end{bmatrix}} \overset{\mathbf{B}_1}{\begin{bmatrix} 2.33 & 2.48 \\ -2.96 & -1.66 \end{bmatrix}} = \begin{bmatrix} 14.19 & 10.69 \\ 10.69 & 8.91 \end{bmatrix}$$

and see that $\mathbf{C}(X)$ can be reproduced in this way just as well as the original way. *This latter property:*

$$\boxed{\mathbf{C}(X) = \mathbf{B}'\mathbf{B} = \mathbf{B}'\mathbf{V}'\mathbf{V}\mathbf{B}}$$

where **V** *is orthogonal* ($\mathbf{V}'\mathbf{V} = \mathbf{V}\mathbf{V}' = \mathbf{I}$) *is of particular relevance to factor analysis and has to do with the so-called rotation problem.* That is, suppose we were first to find the eigenstructure of $\mathbf{C}(X)$ and then write $\mathbf{C}(X)$ in the context of the sample problem as

$$\mathbf{C}(X) = \mathbf{B}'\mathbf{B} = [\mathbf{T}_b\mathbf{D}_b^{1/2}\mathbf{D}_b^{1/2}\mathbf{T}_b{}']$$

By means of the preceding argument, $\mathbf{B}' = \mathbf{T}_b\mathbf{D}_b^{1/2}$ is *not* unique, and we are free to rotate **B** as we please. This, of course, introduces a certain ambiguity into the question of what factor dimensions are "correct." As recalled, the principal components axes found from the *unique* representation:

$$\mathbf{C}(X) = \mathbf{T}_b\mathbf{D}_b\mathbf{T}_b'$$

are unambiguous in the sense of *maximizing variance* in the derived covariance matrix $\mathbf{C}(Z_b)$.

In summary, data-based product-moment matrices exhibit a number of virtues, such as real-valued eigenstructures and orthogonal transformations for rotating the matrix to diagonal form. As we saw in the sample problem, one can order the eigenvalues from largest to smallest in the process of transforming a data matrix into a set of linear composites that are mutually orthogonal. Finally, the rank of product-moment matrices is easily discerned by simply counting up the number of positive eigenvalues. In most data-based problems the rank of $\mathbf{C}(X)$, and other types of derived product-moment matrices, will equal the order of the (minor) product-moment matrix.

5.6.4 Recapitulation

At this point we have covered quite a bit of ground regarding eigenstructures and matrix rank. In the case of nonsymmetric matrices in general, we noted that even if a (square) matrix **A** could be diagonalized via

$$\boxed{\mathbf{D} = \mathbf{U}^{-1}\mathbf{A}\mathbf{U}}$$

the eigenvalues and eigenvectors need not be all real valued. (Fortunately, in the types of matrices encountered in multivariate analysis, we shall always be dealing with real-valued eigenvalues and eigenvectors.)

In the case of symmetric matrices, any such matrix **A** is diagonalizable, and orthogonally so, via

$$\boxed{\mathbf{D} = \mathbf{T}'\mathbf{A}\mathbf{T}}$$

where $\mathbf{T}' = \mathbf{T}^{-1}$ since **T** is orthogonal.[13] Moreover, all eigenvalues and eigenvectors are necessarily real. If the eigenvalues are not all distinct, an orthogonal basis—albeit not a unique one—can still be constructed. Furthermore, eigenvectors associated with distinct eigenvalues are already orthogonal to begin with.

The rank of any matrix **A**, square or rectangular, can be found from its eigenstructure or that of its product moment matrices. If $\mathbf{A}_{n\times n}$ is symmetric, we merely count up the number of nonzero eigenvalues k and note that $r(\mathbf{A}) = k \leqslant n$. If **A** is nonsymmetric or rectangular, we can find its minor (or major) product moment and then compute the eigenstructure. In this case, all eigenvalues are real and nonnegative. To find the rank of **A**, we simply count up the number of positive eigenvalues k and observe that $r(\mathbf{A}) = k \leqslant \min(m, n)$ if **A** is rectangular or $r(\mathbf{A}) = k \leqslant n$ if **A** is square.

[13] Note also that $\mathbf{A} = \mathbf{T}\mathbf{D}^{1/2}\mathbf{D}^{1/2}\mathbf{T}' = \mathbf{T}\mathbf{D}\mathbf{T}'$, as desired.

Finally, if **A** is of product-moment form to begin with, or if **A** is symmetric with nonnegative eigenvalues, then it can be written—although not uniquely so—as $\mathbf{A} = \mathbf{B}'\mathbf{B}$. The lack of uniqueness of **B** was illustrated in the context of axis rotation in factor analysis.

5.7 THE SINGULAR VALUE DECOMPOSITION (SVD) OF A MATRIX

With some oversimplification, we can summarize the intent of the chapter so far by saying that, given a transformation matrix **A**, we would like to find a representation of it in some way that simplifies its geometric nature.

In the case of a square, but nonsymmetric, matrix **A**, we found that under fairly general circumstances (in which the vectors of **U** are linearly independent), **A** could be written as

$$\boxed{\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{U}^{-1}}$$

where **U** is nonsingular and **D** is diagonal. The eigenvalues and eigenvectors of **A** need not all be real valued, however.

In the case of a symmetric matrix **A**, *orthogonal* diagonalization can be achieved. In this case the above equation is satisfied and, in addition, we have

$$\boxed{\mathbf{A} = \mathbf{T}\mathbf{D}\mathbf{T}'}$$

since, given an orthogonal matrix **T**, we know that $\mathbf{T}' = \mathbf{T}^{-1}$. Moreover, all real-valued symmetric matrices are orthogonally diagonalizable with real-valued eigenvalues and eigenvectors.

Not all matrix transformations are symmetric, however, and not all matrices are square. Thus, with the exception of the preceding discussion of matrix rank, we have said relatively little on the topics of (a) the eigenstructure of square, nonsymmetric matrices and (b) the role of eigenstructure analysis in dealing with rectangular matrices which, by definition, do not have eigenstructures. It is now time to bring up these aspects, particularly the latter one.

This section of the chapter deals with *singular value decomposition*, the most general decomposition of a transformation matrix that is discussed in the book. As we shall see, *any* matrix can be decomposed into its basic structure (although not necessarily uniquely so). As such, basic structure represents an extremely powerful concept in multivariate analysis and unifies much of our earlier discussion of matrix decomposition.[14] Again, we shall tie in the current material with some comments made on related matters in Chapter 4.

In Section 4.5.5 we demonstrated that an arbitrary nonsingular matrix

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

[14] While eigenstructure analysis plays a central role in finding the basic structure of a matrix, it should be stressed that the concepts are distinct.

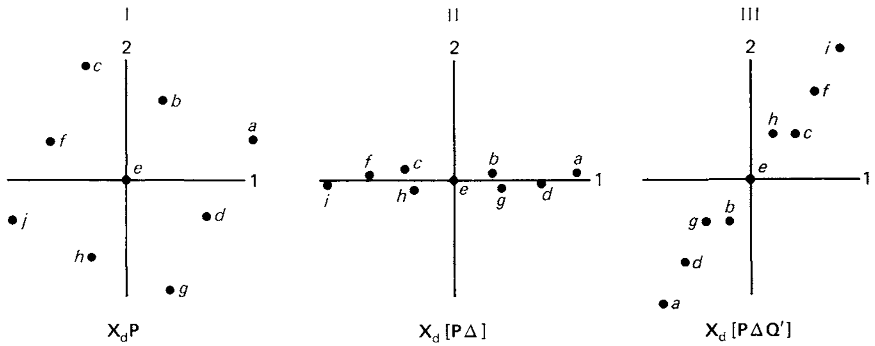

Fig. 5.6 Decomposition of general linear transformation $\mathbf{V} = \mathbf{P\Delta Q'}$ (reproduced from Fig. 4.13).

could be uniquely decomposed into the triple product

$$\mathbf{V} = \mathbf{P\Delta Q'} = \overset{\mathbf{P}}{\begin{bmatrix} -0.41 & -0.91 \\ -0.91 & 0.41 \end{bmatrix}} \overset{\mathbf{\Delta}}{\begin{bmatrix} 5.47 & 0 \\ 0 & 0.37 \end{bmatrix}} \overset{\mathbf{Q'}}{\begin{bmatrix} -0.58 & -0.82 \\ 0.82 & -0.58 \end{bmatrix}}$$

where $\mathbf{P}$ is orthogonal (specifically, an improper rotation), $\mathbf{\Delta}$ is diagonal (a stretch), and $\mathbf{Q'}$ is orthogonal (a proper rotation).

For ease of reference, Fig. 5.6 reproduces Fig. 4.13 in which a square lattice of points (shown in Panel II of Fig. 4.10) was transformed, in three stages, by the matrices making up the specific decomposition of $\mathbf{V}$. Each stage is shown in Fig. 5.6. Moreover, at that point we indicated that *any* nonsingular matrix could be uniquely decomposed into the product of (a) a rotation–stretch–rotation or (b) a rotation–reflection–stretch–rotation. It turns out, however, that this type of decomposition is a very general type of decomposition. *It is so general, in fact, that any matrix, square or rectangular, nonsingular or singular, symmetric or nonsymmetric, can be so decomposed, albeit not necessarily uniquely so.*

In the case of symmetric matrices, we know, of course, that the geometric relationship does hold:

$$\boxed{\mathbf{A} = \mathbf{TDT'}}$$

since, in this case, $\mathbf{T}$ and $\mathbf{T'}$ are orthogonal (rotations), and $\mathbf{D}$ is diagonal (a stretch).[15] Thus, if $\mathbf{A}$ is symmetric, the above decomposition holds as a special case.

However, as already observed in describing the eigenstructure of nonsymmetric matrices, there is no requirement that $\mathbf{T}$ be orthogonal. Furthermore, no such decomposition–orthogonal or otherwise–has been discussed for rectangular matrices.

We now consider the cases in which $\mathbf{A}$ is either square but nonsymmetric or $\mathbf{A}$ is rectangular. As it turns out, both cases can be handled by the same procedure, and we shall illustrate the approach by assuming that $\mathbf{A}$ is rectangular, of order $m \times n$. For convenience, assume that $\mathbf{A}$ is "vertical" with $m > n$, although this is not essential. We shall also assume for the moment that rank of $\mathbf{A}$ is k ($k \leqslant n < m$).

[15] Or, possibly, a stretch preceded by a reflection.

We first examine some of the algebra of this type of decomposition. Then we apply the decomposition to an illustrative problem. Let us set down our objective at the outset. And that is: Given an arbitrary rectangular transformation matrix **A**, we wish to find a way to express **A** in terms of the product of three, relatively simple, matrices:

1. an $m \times k$ matrix **P** which is orthonormal by columns and, hence, satisfies the relation $\mathbf{P}'\mathbf{P} = \mathbf{I}$;
2. a $k \times k$ matrix Δ which is diagonal and consists of k positive diagonal entries ordered from large to small (with ties allowed);
3. an $n \times k$ matrix **Q** which is orthonormal by columns and, hence, satisfies the relation $\mathbf{Q}'\mathbf{Q} = \mathbf{I}$.

The representation of the matrix **A** *as the triple product* $\mathbf{P}\Delta\mathbf{Q}'$ *is called its singular value decomposition* (abbreviated SVD for reasons of brevity from now on). It is occasionally referred to as basic structure (Horst, 1963).

5.7.1 The Algebra of Singular Value Decomposition (SVD)

The mathematical aspects of SVD become rather complex, and so we shall settle for a less technical discussion. Given an arbitrary rectangular matrix $\mathbf{A}_{m \times n}$, where $m > n$ and $r(\mathbf{A}) = k \leqslant n < m$, its SVD involves the triple product

$$\mathbf{A}_{m \times n} = \mathbf{P}_{m \times k}\Delta_{k \times k}\mathbf{Q}'_{k \times n}$$

where **P** and **Q** are each orthonormal by columns ($\mathbf{P}'\mathbf{P} = \mathbf{I}_{k \times k}$; $\mathbf{Q}'\mathbf{Q} = \mathbf{I}_{k \times k}$), and $\Delta_{k \times k}$ is diagonal with ordered positive entries.

Note, in particular, that **Q** cannot be orthogonal (where $\mathbf{Q}\mathbf{Q}' = \mathbf{I}_{n \times n}$) unless $k = n$. Moreover, **P** cannot be orthogonal (where $\mathbf{P}\mathbf{P}' = \mathbf{I}_{m \times m}$) unless $k = n = m$. As such, $\mathbf{P}_{m \times k}$ and $\mathbf{Q}_{n \times k}$ might be called *orthonormal sections* in which all columns are of unit length and mutually orthogonal.

Next, let us comment on the diagonal matrix $\Delta_{k \times k}$, which we will call the *singular value* (*or SV*) *matrix*. First, as we shall see, all elements of $\Delta_{k \times k}$ can be

1. taken to be positive;
2. ordered from large to small (with ties allowed).

Moreover, it will turn out that there is one and only one SV matrix for any given matrix; that is, the singular values, which are the diagonal entries of the SV matrix of the decomposition, are always unique, and this will be true regardless of whether **A** is of full rank, square, or rectangular. It may happen, however, that some entries in $\Delta_{k \times k}$ are tied. If such is the case, only those portions of **P** and **Q** that relate to distinct entries in $\Delta_{k \times k}$ will be unique, a point to which we return later. *Finally, the rank of* **A** *is given quite simply by the number of positive entries in* Δ, *the SV matrix.*

For the moment, let us examine the relationship of $\mathbf{A} = \mathbf{P}\Delta\mathbf{Q}'$ to its major and minor product moments, $\mathbf{A}\mathbf{A}'$ and $\mathbf{A}'\mathbf{A}$, respectively:

$$\mathbf{A}\mathbf{A}' = (\mathbf{P}\Delta\mathbf{Q}')(\mathbf{P}\Delta\mathbf{Q}')' = \mathbf{P}\Delta\mathbf{Q}'\mathbf{Q}\Delta\mathbf{P}'$$

but since $\mathbf{Q}'\mathbf{Q} = \mathbf{I}$ and letting $\Delta^2 = \mathbf{D}$, we see that **D** is still diagonal. Thus, we get

$$\mathbf{A}\mathbf{A}' = \mathbf{P}\mathbf{D}\mathbf{P}'$$

Furthermore,

$$\mathbf{A}'\mathbf{A} = (\mathbf{P}\boldsymbol{\Delta}\mathbf{Q}')'(\mathbf{P}\boldsymbol{\Delta}\mathbf{Q}') = \mathbf{Q}\boldsymbol{\Delta}\mathbf{P}'\mathbf{P}\boldsymbol{\Delta}\mathbf{Q}'$$

but since $\mathbf{P}'\mathbf{P} = \mathbf{I}$ and $\boldsymbol{\Delta}^2 = \mathbf{D}$, we get

$$\mathbf{A}'\mathbf{A} = \mathbf{Q}\mathbf{D}\mathbf{Q}'$$

Note that in both cases we have the eigenstructure formulation shown earlier for the case of symmetric matrices, namely, the triple product of an orthogonal, diagonal, and transposed orthogonal matrix. This is not surprising since both $\mathbf{A}\mathbf{A}'$ and $\mathbf{A}'\mathbf{A}$ are symmetric. However, the diagonal matrix $\mathbf{D}$ of each triple product *is the same* for both product moments $\mathbf{A}\mathbf{A}'$ and $\mathbf{A}'\mathbf{A}$. Furthermore,

1. all entries of $\mathbf{D}$ are real;
2. all entries of $\mathbf{D}$ are nonnegative;
3. all positive entries of $\mathbf{D}$ can be ordered from large to small (with possible ties, of course).

We take advantage of these facts in describing $\boldsymbol{\Delta}$, the $k \times k$ portion of $\mathbf{D}$ that has positive (as opposed to zero) entries, in terms of the following definition:

$$\boxed{\boldsymbol{\Delta}_{k \times k} = \mathbf{D}^{1/2}_{k \times k}}$$

At this point, then, we are starting to get some hints about how to find $\mathbf{P}_{m \times k}$, $\boldsymbol{\Delta}_{k \times k}$, and $\mathbf{Q}'_{k \times n}$ As observed above, $\mathbf{P}_{m \times k}$ represents the first k columns of the orthogonal matrix $\mathbf{P}_{m \times m}$ obtained from the eigenstructure of $\mathbf{A}\mathbf{A}'$, while $\mathbf{Q}_{n \times k}$ represents the first k columns of the orthogonal matrix $\mathbf{Q}_{n \times n}$ obtained from the eigenstructure of $\mathbf{A}'\mathbf{A}$. Their common diagonal matrix $\mathbf{D}$ has all nonnegative entries. Furthermore, we can order (with ties allowed) the positive entries of $\mathbf{D}$ from large to small, until we get k of them. The remaining entries on the main diagonal will all be zero. The columns of $\mathbf{P}$ and $\mathbf{Q}$ can, of course, be made to correspond to the ordered diagonal elements in D and, hence, to their square roots in $\boldsymbol{\Delta}$.

Next, let us take the argument one step further. If we let $\boldsymbol{\Delta}$ be the first k rows and k columns embedded in a (larger) $m \times n$ *rectangular* matrix, with $m - k$ rows and $n - k$ columns of zeros elsewhere, both $\mathbf{P}$ and $\mathbf{Q}'$ could be made fully orthogonal and, in this sense, properly constitute rotation matrices of order $m \times m$ and $n \times n$, respectively. This generalization can be called the *full* SVD of a matrix.

The preceding generalization is a significant one. It tells us that *any* matrix—not just square, nonsingular ones—with real-valued entries can be represented as the product of

1. a rotation (possibly followed by a reflection), followed by
2. a stretch, followed by
3. a rotation.

Note further that if, indeed, $\mathbf{A}$ is *symmetric* to begin with, we have the special case

$$\mathbf{A} = \mathbf{T}\mathbf{D}\mathbf{T}'$$

since $\mathbf{A}\mathbf{A}' = \mathbf{A}'\mathbf{A}$, and therefore $\mathbf{P}' = \mathbf{Q}' = \mathbf{T}'$. Hence, this same approach to matrix decomposition can be applied even to the more familiar case of a symmetric matrix.

However, the diagonal **D** is to be interpreted as Δ in the current context since we refer to **A** rather than to its product moments **AA**′ or **A**′**A**.

In summary, *any* matrix of real entries has a SVD and can be written as

$$\mathbf{A} = \mathbf{P}\Delta\mathbf{Q}'$$

where Δ is diagonal, and **P** and **Q** are orthonormal by columns. The concept of *full* SVD embeds the $k \times k$ diagonal matrix of positive entries in an $m \times n$ matrix in which $m - k$ rows and $n - k$ columns are zeros. Alternatively, we can require Δ to be $k \times k$ with all positive entries and, hence, **P** and **Q** will not, in general, be rotation matrices, although they will still be orthonormal by columns and constitute orthonormal sections.

Finally, a special case of the above involves the case in which **A** is symmetric to begin with. If so, it can be written as

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

where **T**, of course, is an orthogonal matrix and $\mathbf{D} = \Delta$ in the present discussion.

Figure 5.7 shows schematically the two cases that we have been considering. Panel I shows the case in which **P** is orthonormal by columns ($\mathbf{P}'\mathbf{P} = \mathbf{I}$) but is not orthogonal. Similarly, **Q** is orthonormal by columns ($\mathbf{Q}'\mathbf{Q} = \mathbf{I}$) but is not orthogonal. The diagonal matrix Δ has k positive eigenvalues, where $k < n < m$.

Panel II shows the full SVD in the sense that **P** and **Q** can be made orthogonal by embedding Δ in a larger matrix, consisting of $m - k$ rows and $n - k$ columns of zeros (in addition, of course, to the zeros in the off-diagonal entries of the $k \times k$ portion). Not surprisingly, the $m - k$ columns of **P** and the $n - k$ rows of **Q**—while they could be made orthogonal—are mainly of theoretical interest since those dimensions would be annihilated by the $m - k$ rows of zeros and the $n - k$ columns of zeros in the $m \times n$ matrix in which Δ is embedded.

5.7.2 Conditions for Full Rank Matrices

The foregoing discussion applies to any matrix of interest since any matrix possesses a SVD, written as the triple product $\mathbf{P}\Delta\mathbf{Q}'$. While any matrix can be decomposed into its SVD, not all matrices are full rank. This distinction needs to be made in discussing the rank of a matrix.

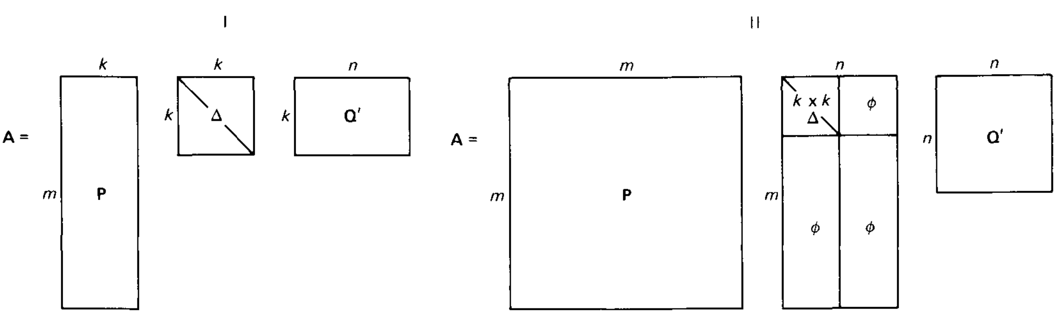

Fig. 5.7 Alternative formulations of the basic structure of $\mathbf{A}_{m \times n}$

A full rank matrix **A** is one whose rank equals its smaller order. If **A** is $m \times n$ (and $m > n$), and if **A** is full rank then $r(\mathbf{A}) = k = n$.

If **A** is square and full rank, then $r(\mathbf{A})$ equals its (common) order, and we have called this kind of matrix *nonsingular*. If **A** is rectangular and full rank, then the rank of the major product moment $\mathbf{AA}'$ or the minor product moment $\mathbf{A}'\mathbf{A}$ is equal to the *smaller order* of **A**. If **A** is not full rank, then its rank is $k(k < n \leqslant m)$.

This can all be summarized by saying that any matrix **A** can be decomposed into the SVD

$$\boxed{\mathbf{A}_{m \times n} = \mathbf{P}_{m \times k}\Delta_{k \times k}\mathbf{Q}'_{k \times n}}$$

where $k \leqslant \min(m, n)$. *Then* **A** *is full rank if and only if* $k = \min(m,n)$.

5.7.3 Finding the Singular Value Decomposition (SVD)

It is one thing to define the SVD of a matrix **A** and quite another to solve for its representation as $\mathbf{A} = \mathbf{P}\Delta\mathbf{Q}'$. Finding the SVD of a matrix **A** makes use of concepts already discussed under the topic of eigenstructure. First, as previously discussed, if $\mathbf{A} = \mathbf{P}\Delta\mathbf{Q}'$, then

$$\mathbf{A}'\mathbf{A} = \mathbf{Q}\Delta\mathbf{P}'\mathbf{P}\Delta\mathbf{Q}'$$

and, since **P** is orthonormal by columns, we have

$$\mathbf{A}'\mathbf{A} = \mathbf{Q}\Delta^2\mathbf{Q}'$$

Then, after we solve for Δ^2 and **Q** by finding the eigenstructure of the symmetric matrix $\mathbf{A}'\mathbf{A}$, we can find Δ and then find **P** from

$$\boxed{\mathbf{P} = \mathbf{A}\mathbf{Q}\Delta^{-1}}$$

Since Δ is diagonal, Δ^{-1} consists simply of the reciprocals of the diagonal entries of Δ. At this point, then, we have a procedure for solving for the triple product

$$\mathbf{A} = \mathbf{P}\Delta\mathbf{Q}'$$

Note that the above implies the singular value matrix for the SVD of **A** is the square root of the diagonal matrix of eigenvalues of the minor product moment matrix of **A**, $\mathbf{A}'\mathbf{A}$. The diagonal entries of Δ are called "singular values" (thus the term "singular value matrix" for Δ), so that the singular values of **A** are the square roots of the eigenvalues of $\mathbf{A}'\mathbf{A}$. The matrix **Q** is the matrix of principal *eigenvectors* (those associated with positive eigenvalues) of $\mathbf{A}'\mathbf{A}$, but it is also called the matrix of principal *right singular vectors* of **A**. It has already been seen that the k nonzero eigenvalues of the minor product moment of **A** ($\mathbf{A}'\mathbf{A}$) are identical to the (k) nonzero eigenvalues of its *major* product moment matrix ($\mathbf{AA}'$). It turns out that **P** is the matrix whose columns are the principal eigenvectors of $\mathbf{AA}'$ (those associated with its k nonzero eigenvalues and ordered in the same descending order in which those eigenvalues themselves are conventionally ordered in the diagonal eigenvalue matrix $\mathbf{D} = \Delta^2$). The columns of **P** are also called principal *left singular vectors* of **A**.

Thus the singular value decomposition (SVD) decomposes $\mathbf{A}$ into a triple product of three matrices, the first that of its *left singular vectors* ($\mathbf{P}$), the second the diagonal *singular value* matrix(Δ), and the third being the *transpose* of the matrix of *right singular vectors* ($\mathbf{Q}'$). These matrices, as we have seen, all are very closely related to, and can be defined via, the eigenvectors of the minor and major product moment matrices of $\mathbf{A}$. (We could, for example, estimate $\mathbf{P}$ directly by computing the eigenvectors of $\mathbf{AA}'$, instead of using the equation $\mathbf{P} = \mathbf{AQ}\Delta^{-1}$ above, but this approach would be computationally more expensive, in general, and could also lead to certain indeterminacy problems in case of tied eigenvalues—so the approach discussed above is definitely preferable to this alternative approach.)

The procedure involves the following steps, assuming first that $\mathbf{A}$ is of order $m \times n$ with $n \leq m$:

1. Compute the minor product moment $\mathbf{A}'\mathbf{A}$ which results in a square symmetric matrix of order $n \times n$, where $n \leq m$.
2. Find the eigenstructure of $\mathbf{A}'\mathbf{A}$, thus yielding the matrix of eigenvalues Δ^2 of rank k ($k \leq n$) and the matrix $\mathbf{Q}$ where $\mathbf{Q}$ is the matrix of associated eigenvectors that are orthonormal by columns.
3. Find the square roots of the diagonal entries of Δ^2.
   These are called "singular values of $\mathbf{P}$".
4. Find Δ^{-1}, the reciprocals of the diagonal entries of Δ.
5. Find $\mathbf{P} = \mathbf{AQ}\Delta^{-1}$.

It turns out that if $k = n$, then $\mathbf{A}$ is full rank. If $k < n$, then $\mathbf{A}$ is not full rank but of rank k.

On the other hand, if $n > m$ we apply the same type of procedure to $\mathbf{A}'$ and transpose the result. That is, let

$$\mathbf{A}' = \mathbf{P}_1\Delta_1\mathbf{Q}_1'$$

Then its transpose is

$$\mathbf{A} = (\mathbf{P}_1\Delta_1\mathbf{Q}_1')' = \mathbf{Q}_1\Delta_1\mathbf{P}_1'$$

Then, if we define

$$\mathbf{P} \equiv \mathbf{Q}_1; \qquad \Delta \equiv \Delta_1; \qquad \mathbf{Q} \equiv \mathbf{P}_1$$

it turns out that we have the desired decomposition of $\mathbf{A}$ into the SVD

$$\mathbf{A} = \mathbf{P}\Delta\mathbf{Q}'$$

5.7.4 Illustrating the Singular Value Decomposition (SVD) Procedure

To illustrate the procedure described above, let us take a particularly simple case involving a 3 x 2 matrix:

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

Since $m > n$, we first find the (smaller) minor product-moment matrix:

$$\mathbf{A}'\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 1 & 9 \end{bmatrix} = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 0 & 2 \\ -1 & 1 \end{bmatrix}$$

Table 5.2 shows a summary of the computations involved in finding the eigenstructure of $\mathbf{A}'\mathbf{A}$. Note that this part of the problem is a standard one in finding the eigenstructure of a symmetric matrix.

After finding Δ^2 and $\mathbf{Q}$, by means of solving the characteristic equation, we find Δ and then Δ^{-1}. The last step is to solve for $\mathbf{P}$ in the equation

$$\boxed{\mathbf{P} = \mathbf{A}\mathbf{Q}\Delta^{-1}}$$

These results also appear in Table 5.2. Finally, we assemble the triple product

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 0 & 2 \\ -1 & 1 \end{bmatrix} = \underset{\mathbf{P}}{\begin{bmatrix} -0.70 & 0.52 \\ -0.65 & -0.20 \\ -0.28 & -0.83 \end{bmatrix}} \underset{\mathbf{\Delta}}{\begin{bmatrix} 3.02 & 0 \\ 0 & 1.36 \end{bmatrix}} \underset{\mathbf{Q}'}{\begin{bmatrix} -0.14 & -0.99 \\ 0.99 & -0.14 \end{bmatrix}}$$

As can be noted, after taking the transpose of $\mathbf{Q}$, the matrix $\mathbf{A}$ has been decomposed into the product of an orthonormal (by columns) matrix times a diagonal times a (square) orthogonal matrix. In general, however, $\mathbf{Q}'$ will be orthonormal by rows if $\mathbf{A}$ is not full rank.

TABLE 5.2

Finding the Eigenstructure of $\mathbf{A}'\mathbf{A}$

Minor product-moment matrix

$$\mathbf{A}'\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 1 & 9 \end{bmatrix}$$

Characteristic equation

$$|(\mathbf{A}'\mathbf{A}) - \lambda \mathbf{I}| = \begin{vmatrix} 2-\lambda & 1 \\ 1 & 9-\lambda \end{vmatrix} = 0$$

$$\lambda^2 - 11\lambda + 17 = 0$$

Quadratic formula

$$y = ax^2 + bx + c$$

Substitution in general quadratic

$$\frac{-b \pm \sqrt{b^2 - 4ac}}{2a} = \frac{11 \pm \sqrt{(-11)^2 - 4(17)}}{2}$$

Eigenvalues of $\mathbf{A}'\mathbf{A}$

$$\lambda_1 = 9.14$$

$$\lambda_2 = 1.86$$

Matrix of eigenvectors of $\mathbf{A}'\mathbf{A}$

$$\mathbf{Q} = \begin{bmatrix} -0.14 & 0.99 \\ -0.99 & -0.14 \end{bmatrix}$$

Diagonal singular value (SV) matrix

$$\Delta = \begin{bmatrix} (9.14)^{1/2} & 0 \\ 0 & (1.86)^{1/2} \end{bmatrix} = \begin{bmatrix} 3.02 & 0 \\ 0 & 1.36 \end{bmatrix}$$

Solving for the matrix $\mathbf{P}$

$$\mathbf{P} = \underset{\mathbf{A}}{\begin{bmatrix} 1 & 2 \\ 0 & 2 \\ -1 & 1 \end{bmatrix}} \underset{\mathbf{Q}}{\begin{bmatrix} -0.14 & 0.99 \\ -0.99 & -0.14 \end{bmatrix}} \underset{\Delta^{-1}}{\begin{bmatrix} 1/3.02 & 0 \\ 0 & 1/1.36 \end{bmatrix}} = \begin{bmatrix} -0.70 & 0.52 \\ -0.65 & -0.20 \\ -0.28 & -0.83 \end{bmatrix}$$

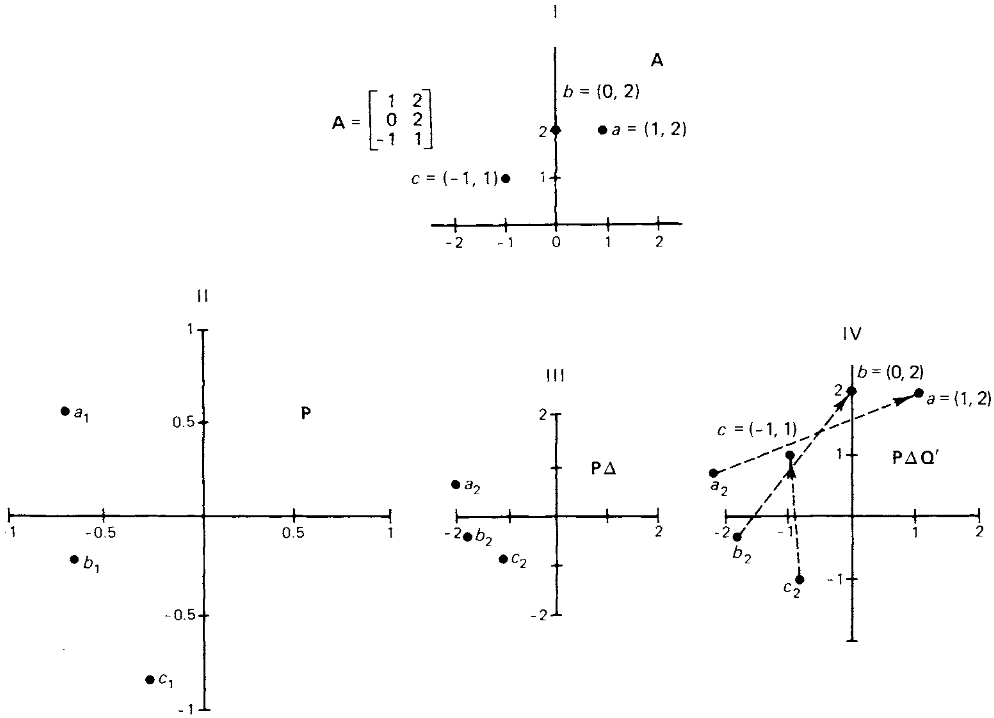

Fig. 5.8 Decomposition of **A** to SVD **A** = **PΔQ′**. Key: I, "target" configuration as defined by **A** = **PΔQ′**; II, orthonormal-by-columns matrix **P**; III, application of stretch (defined by Δ) to **P**; IV, rotation of **PΔ** via matrix **Q′**.

Figure 5.8 shows the separate aspects of the decomposition. We first start with the matrix **P**. **P** is then differentially stretched in accordance with **Δ**. Finally, the points of **PΔ** are rotated in accordance with **Q′**, leading to a reproduction of what we started out with, namely, the matrix **A** shown at the top of the figure.

5.7.5 The SVD of the Sample Problem Matrix of Predictors

In Section 5.4 we found the eigenstructure of the covariance matrix of the two predictor variables in the sample problem. As recalled,

$$\mathbf{C}(X) = \mathbf{X}_d'\mathbf{X}_d/m = \mathbf{TDT}'$$

$$= \overset{\mathbf{T}}{\begin{bmatrix} 0.787 & 0.617 \\ 0.617 & -0.787 \end{bmatrix}} \overset{\mathbf{D}}{\begin{bmatrix} 22.56 & 0 \\ 0 & 0.54 \end{bmatrix}} \overset{\mathbf{T}'}{\begin{bmatrix} 0.787 & 0.617 \\ 0.617 & -0.787 \end{bmatrix}} = \begin{bmatrix} 14.19 & 10.69 \\ 10.69 & 8.91 \end{bmatrix}$$

where **T** and **T′** are orthogonal and **D** is diagonal.

Suppose, now, that we wished to find the SVD of $\mathbf{X}_d/\sqrt{m}$, the mean-corrected matrix of predictors, scaled by the square root of the sample size.

Based on what we have just covered, we know how to proceed. First, we find

$$\Delta^{-1} = D^{-1/2} = \begin{bmatrix} 1/\sqrt{22.56} & 0 \\ 0 & 1/\sqrt{0.54} \end{bmatrix} = \begin{bmatrix} 0.211 & 0 \\ 0 & 1.361 \end{bmatrix}$$

Next, analogous to solving for $\mathbf{P}$ in the expression $\mathbf{P} = \mathbf{AQ}\Delta^{-1}$, we now solve for $\mathbf{U}$ in the expression

$$\mathbf{U} = \mathbf{X}_\mathrm{d}/\sqrt{m}\,\mathbf{T}\Delta^{-1}$$

which leads to the SVD of $\mathbf{X}_\mathrm{d}/\sqrt{m}$ in the present notation as

$$\boxed{\mathbf{X}_\mathrm{d}/\sqrt{m} = \mathbf{U}\Delta\mathbf{T}'}$$

Since no new principles are involved, we do not go through the extensive computations to find $\mathbf{U}$, which is of order 12×2. What can be said, however, is that any data matrix, $\mathbf{X}$, $\mathbf{X}_\mathrm{d}$, $\mathbf{X}_\mathrm{s}$, $\mathbf{X}/\sqrt{m}$, $\mathbf{X}_\mathrm{s}/\sqrt{m}$ can be expressed in terms of SVD in just the way described above.

5.7.6 Recapitulation

The concept of SVD represents the most general of decompositions that are considered in this chapter. We have only provided introductory material on the topic, but, having done so, it seems useful to recapitulate the main results and add a few more comments as well:

1. Any matrix $\mathbf{A}$ can be decomposed into the triple product:

$$\mathbf{A} = \mathbf{P}\Delta\mathbf{Q}'$$

where $\mathbf{P}$ and $\mathbf{Q}$ are each orthonormal by columns, and Δ is diagonal with ordered positive elements.

2. The number of positive elements in Δ, the singular value (SV) matrix, is equal to the rank of $\mathbf{A}$. Moreover, the SV matrix is unique.

3. If all entries of Δ are distinct, then $\mathbf{P}$ and $\mathbf{Q}'$ are also unique (up to a possible reflection).

4. If some entries of Δ are tied, then those portions of $\mathbf{P}$ and $\mathbf{Q}'$ corresponding to tied blocks of entries are not unique. The portions of $\mathbf{P}$ and $\mathbf{Q}$ corresponding to the subset of distinct entries of Δ are unique (up to a reflection), however.

5. Mutually orthogonal vectors for tied blocks of entries in Δ can also be found by the Gram–Schmidt process, after first finding a set of r linearly independent vectors in the tied block. (These are unique up to orthogonal rotation within the r-dimensional subspace corresponding to the r tied eigenvalues.)

6. A full rank matrix is one whose rank equals its smaller dimension.

7. A square full rank matrix is one whose rank equals its smaller order. More commonly, a square full rank matrix is called nonsingular.

8. If $\mathbf{A}$ is nonsingular, then $\mathbf{P}'\mathbf{P} = \mathbf{P}\mathbf{P}' = \mathbf{I}$; $\mathbf{Q}'\mathbf{Q} = \mathbf{Q}\mathbf{Q}' = \mathbf{I}$, and we have the case of rotation–stretch–rotation or rotation–reflection–stretch–rotation.

9. If $\mathbf{A}$ is rectangular or square but singular, the concept of full SVD, in which $\Delta_{k\times k}$ is embedded in a larger $m \times n$ matrix (see Fig. 5.7), still involves the sequence of transformations shown immediately above. Some dimensions of $\mathbf{P}$ and/or $\mathbf{Q}$ are annihilated, however.

10. The orthogonal diagonalization of a symmetric matrix

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

was shown to be a special case of SVD.

11. The concepts of matrix nonsingularity and decomposition uniqueness should be kept separate. A square matrix $\mathbf{A}_{n \times n}$ can be nonsingular but still nonunique in terms of its SVD if it contains tied (positive) entries in $\Delta_{n \times n}$.

12. A square matrix $\mathbf{A}_{n \times n}$ can be singular but still unique in terms of its SVD if it contains all distinct entries in $\Delta_{n \times n}$, thus implying that only one entry is zero.

13. A square matrix $\mathbf{A}_{n \times n}$ can, of course, be both nonsingular and unique in terms of SVD if all entries in $\Delta_{n \times n}$ are positive and distinct.

As we know at this point, if a matrix is of rank k, then the SVD procedure will reproduce it in terms of the triple product $\mathbf{P}\Delta\mathbf{Q}'$, where the diagonal SV matrix Δ is $k \times k$.

What has not been covered is the case in which we would like to approximate $\mathbf{A}$ with a triple product whose diagonal is of order *less* than $k \times k$. This type of problem crops up in principal components analysis, among other things, when we wish to reduce the original space to fewer dimensions with the least loss of information.

Fortunately, SVD provides a reduced-rank approximation to $\mathbf{A}$ whose sum of squared deviations between $\mathbf{A}$ and $\mathbf{P}\Delta\mathbf{Q}'$ is minimal for the order of the diagonal being retained. While it would take us too far afield to explore the topic of matrix approximation via SVD, this turns out to be another valuable aspect of the technique. Not surprisingly, the fact that the entries of Δ are ordered from large to small figures prominently in this type of approximation.

5.8 QUADRATIC FORMS

In multivariate analysis one often encounters situations in which the mapping of some vector entails a quadratic, rather than linear, function.[16] At first blush it may seem surprising that matrix algebra is relevant for this situation. After all, thus far we have emphasized the applicability of matrices to linear transformations. It is now time to expand the topic and consider quadratic functions and, in particular, quadratic forms.

5.8.1 Linear Forms

We have already encountered linear forms in our discussion of simultaneous equations in Chapter 4. If we have a set of variables x_i and a set of coefficients a_i, a linear form can be written in scalar notation as

$$g(x) = a_1x_1 + a_2x_2 + \cdots + a_nx_n = \sum_{i=1}^{n} a_ix_i$$

in which all x_i, as noted, are of the first degree. In vector notation we have

[16] Clearly, the idea of the *variance* of some variable entails a quadratic function, and variances represent a central concept in statistical analysis.

$$g(\mathbf{x}) = \mathbf{a}'\mathbf{x} = (a_1, a_2, \ldots, a_n)\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

which, of course, equals some scalar, once we assign numerical values to **a** and **x**.

Next, suppose we consider a set of several linear forms, with the matrix of coefficients given by **A** and the vector of constants given by **c**. Then we have

$$\mathbf{Ax} = \mathbf{c} = \overset{\mathbf{A}}{\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}} \overset{\mathbf{x}}{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}} = \overset{\mathbf{c}}{\begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}}$$

This, of course, represents a set of simultaneous linear equations. Hence, a linear form is simply a linear function in a set of variables x_i.

5.8.2 Bilinear Forms

Bilinear forms involve only a slight extension of the above. Here we have two sets of variables x_i and y_j, each of the first degree, as illustrated specifically by

$$f(x, y) = x_1y_1 + 6x_2y_1 - 4x_3y_1 + 2x_1y_2 + 3x_2y_2 + 2x_3y_2$$

in which exactly one x_i and one y_j (each of the first degree) appears in each term. More generally, expressions of this type can be written in scalar notation as

$$f(x, y) = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}x_iy_j$$

and are called bilinear forms in x_i and y_j. If we write the vectors $\mathbf{x}' = (x_1, x_2, \ldots, x_m)$ and $\mathbf{y}' = (y_1, y_2, \ldots, y_n)$, a bilinear form involves terms in which every possible combination of vector components is formed. In matrix notation we can write a bilinear form as

$$f(\mathbf{x}, \mathbf{y}) = \mathbf{x}'\mathbf{A}\mathbf{y}$$

In the numerical example above, we have

$$a_{11} = 1; \qquad a_{12} = 2; \qquad a_{21} = 6; \qquad a_{22} = 3; \qquad a_{31} = -4; \qquad a_{32} = 2$$

and the function can be expressed as

$$\begin{array}{cccc} & \mathbf{x}' & \mathbf{A} & \mathbf{y} \\ f(\mathbf{x},\mathbf{y}) = & (x_1, x_2, x_3) & \begin{bmatrix} 1 & 2 \\ 6 & 3 \\ -4 & 2 \end{bmatrix} & \begin{bmatrix} y_1 \\ \\ y_2 \end{bmatrix} \end{array}$$

$$= (x_1 + 6x_2 - 4x_3, 2x_1 + 3x_2 + 2x_3)\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

$$= x_1 y_1 + 6x_2 y_1 - 4x_3 y_1 + 2x_1 y_2 + 3x_2 y_2 + 2x_3 y_2$$

The matrix **A** is called the matrix of the bilinear form, and it determines the form completely. Note that, in general, **A** need not be square.

By assigning different values to **x** and **y** one obtains different values of the bilinear form, each of which is a scalar. The set of all such scalars, for a given domain of **x** and **y**, is the range of the bilinear form.

5.8.3 Quadratic Forms

Next, let us specialize the bilinear form to the specific case in which $\mathbf{x} = \mathbf{y}$. In this case we assume that **y** can be replaced by **x** and, given their same dimensionality, the matrix of coefficients **A** will be square rather than rectangular. For example,

$$f(x_1, x_2) = 2x_1^2 + 5x_1x_2 + 3x_1x_2 + 6x_2^2$$

can now be written in matrix form as

$$f(\mathbf{x}) = (x_1, x_2)\begin{bmatrix} 2 & 3 \\ 5 & 6 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = (2x_1 + 5x_2, 3x_1 + 6x_2)\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$
$$= 2x_1^2 + 5x_1x_2 + 3x_1x_2 + 6x_2^2$$

and the result, again, is a scalar, once numerical values are assigned to x_1 and x_2. Also, by assigning different values to **x** over its domain, we can obtain the range of $f(\mathbf{x})$, the quadratic form.

By way of formal definition, *a quadratic form is a polynomial function of* x_1, $x_2, \ldots, x_n$ that is homogeneous and of second degree. For example, in the case of two variables, we have

$$f(x_1, x_2) = x_1^2 + 6x_1x_2 + 9x_2^2$$

However, we can also write this as

$$f(\mathbf{x}) = x_1^2 + 6x_1x_2 + 9x_2^2$$

in which the vector $\mathbf{x}' = (x_1, x_2)$ is mapped from a two-dimensional space into a one-dimensional space. Similarly, an example of a quadratic form in three variables is

$$f(\mathbf{x}) = x_1^2 + x_2^2 + x_3^2$$

where the vector $\mathbf{x}' = (x_1, x_2, x_3)$ is mapped from three dimensions to one dimension.

In general, a quadratic form in n dimensions can be written in scalar notation as

$$q(\mathbf{u}) = \sum_{i,j}^{n} a_{ij}u_iu_j$$

where $\mathbf{u}' = (u_1, u_2, \ldots, u_n)$, the a_{ij} are real-valued coefficients and the $u_i u_j$ are the preimages of the mapping. If $i = j$, we obtain the squared term $a_iu_i^2$, and if $i \neq j$, we obtain the cross-product term $a_{ij}u_iu_j$.

By "homogeneous" we mean that all terms are of the above form and, in particular, there are no linear terms in the u_i's nor is there a constant term. While the function

$$v = x_1^2 + 2x_2^2 + x_1x_2 + x_1 + 3x_2$$

is a second-degree polynomial, it is not a quadratic form since the last two terms are not of the general form $a_{ij}u_iu_j$.

Quadratic forms are of particular interest to multivariate data analysis inasmuch as we are often concerned with what happens to variances and covariances under various linear functions of a set of multivariate data.

While we did not bring up the topic of quadratic forms at that time, our diagonalization of the sample problem covariance matrix in Sections 5.3 and 5.4 involved a quadratic form, with matrix $\mathbf{C}(X)$. Indeed, *all of the cross-product matrices employed in multivariate analysis,* such as the raw cross product, SSCP, covariance, and correlation matrices, are illustrations of quadratic forms. In these cases the diagonal entries are some measure of single-variable dispersion, and the off-diagonal entries are some measure of covariation between a pair of variables.

In working with quadratic forms, our motivation is similar to diagonalizing transformation matrices. That is, we shall wish to find a linear function of the original data that has the effect of leading to a cross-products matrix in which two things are desired: (a) an arrangement of the linear composites so that the main diagonal entries in the cross-product matrix decrease in size and (b) off-diagonal entries of the cross-products matrix being zero, indicating uncorrelatedness of all pairs of composites. This, of course, is the same motivation underlying principal components analysis, as illustrated in Section 5.4.

5.8.4 An Illustrative Problem

Suppose we have the quadratic form $q(\mathbf{x}) = 66x_1^2 + 24x_1x_2 + 59x_2^2$. This can be expressed in matrix product form as

$$\mathbf{q}'\mathbf{A}\mathbf{q} = (x_1, x_2)\begin{bmatrix} 66 & 12 \\ 12 & 59 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$= (66x_1 + 12x_2, 12x_1 + 59x_2)\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$= 66x_1^2 + 12x_1x_2 + 12x_1x_2 + 59x_2^2$$

$$\mathbf{q}'\mathbf{A}\mathbf{q} = 66x_1^2 + 24x_1x_2 + 59x_2^2$$

Notice that we set the off-diagonal entries of $\mathbf{A}$ to half the coefficient of x_1x_2 which is $24/2 = 12$.

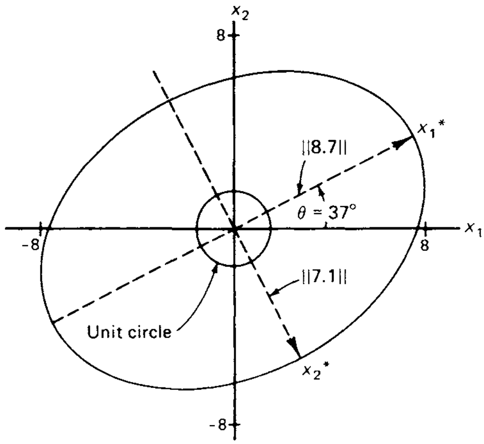

Fig. 5.9 Change of basis vectors of matrix representing quadratic form.

Notice further that a quadratic form involves a transformation into one dimension of an n-component vector in which the transformation is characterized by an $n \times n$ symmetric matrix.[17] On the other hand, a linear mapping of a vector in n dimensions into one dimension entails a single vector (either a $1 \times n$ or an $n \times 1$ matrix), whose entries are usually expressed as direction cosines.

Now let us see what happens when we take various values of x_1 and x_2 and substitute them in $\mathbf{q}'\mathbf{A}\mathbf{q}$. The way this can be done graphically, as shown in Fig. 5.9, is to take various pairs of x_1, x_2 values on the unit circle in which we have the condition

$$(x_1, x_2)\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 1$$

For example let

$x_1 = 1;$	$x_2 = 0$	$\mathbf{q}'\mathbf{A}\mathbf{q} = 66$	$(\mathbf{q}'\mathbf{A}\mathbf{q})^{1/2} = 8.1$
$x_1 = 0;$	$x_2 = 1$	$= 59$	$= 7.7$
$x_1 = -0.707;$	$x_2 = -0.707$	$= 74.5$	$= 8.7$
$x_1 = 0.707;$	$x_2 = 0.707$	$= 74.5$	$= 8.7$
$x_1 = -0.707;$	$x_2 = 0.707$	$= 50.5$	$= 7.1$
$x_1 = 0.707;$	$x_2 = -0.707$	$= 50.5$	$= 7.1$

We can select still other vectors of points on the unit circle and multiply the length or distance from the origin of each by $(\mathbf{q}'\mathbf{A}\mathbf{q})^{1/2}$. This "stretching" of $\mathbf{q}$ on the unit circle

[17] The matrix of a quadratic form does not have to be represented by a symmetric matrix. However, the original matrix can always be symmetrized by setting each off-diagonal entry equal to half the sum of the original off-diagonal entries; that is, $(a_{ij} + a_{ji})/2$.

into the point $[(\mathbf{q}'\mathbf{A}\mathbf{q})^{1/2}]\mathbf{q}$ results in the ellipse shown in Fig. 5.9. This ellipse may be viewed as a geometric representation of the quadratic form.

Now, however, suppose we consider another quadratic form:

$$\mathbf{u}'\mathbf{B}\mathbf{u} = 75x_1^{*2} + 50x_2^{*2}$$

that can, in turn, be represented as

$$\mathbf{u}'\mathbf{B}\mathbf{u} = (x_1^*, x_2^*)\begin{bmatrix} 75 & 0 \\ 0 & 50 \end{bmatrix}\begin{bmatrix} x_1^* \\ x_2^* \end{bmatrix}$$

We see from Fig. 5.9 that if we rotate the coordinate system 37° to the axes x_1^* and x_2^*, then *this function also lies on the previously obtained ellipse.* Vector lengths of the major and minor semiaxes of the ellipse are $\sqrt{75} = 8.7$ and $\sqrt{50} = 7.1$, respectively. That is, by a change in orientation of the axes, we obtain a *new* representation of the quadratic form in which the x_1x_2 cross product vanishes. Moreover, one axis of this form coincides with the major axis of the ellipse, while the other corresponds to the minor axis of the ellipse. These axes are usually referred to as principal axes. By eliminating the cross-product term the second matrix is seen to be a simpler representation of the quadratic form than the first. Moreover, the entries of the diagonal matrix **B** are in decreasing order.

5.8.5 Finding the New Basis Vectors

As the reader has probably surmised already, the new representation of the quadratic form $\mathbf{u}'\mathbf{B}\mathbf{u}$ is obtained by solving for the eigenstructure of **A**. Primarily in the nature of review we set up the characteristic equation

$$|\mathbf{A} - \lambda\mathbf{I}| = \begin{vmatrix} 66-\lambda & 12 \\ 12 & 59-\lambda \end{vmatrix} = 0$$

and solve for its eigenvalues by finding the second-order determinant and setting it equal to zero:

$$\lambda^2 - 125\lambda + 3750 = 0$$

$$(\lambda - 75)(\lambda - 50) = 0$$

$$\lambda_1 = 75; \qquad \lambda_2 = 50$$

After substitution of λ_1 and λ_2, we find the normalized eigenvectors $\begin{bmatrix} -0.8 \\ -0.6 \end{bmatrix}$ and $\begin{bmatrix} 0.6 \\ -0.8 \end{bmatrix}$, which can be arranged in the matrix **Q**.

$$\mathbf{Q} = \begin{bmatrix} -0.8 & 0.6 \\ -0.6 & -0.8 \end{bmatrix}$$

Notice that $|\mathbf{Q}| = 1$ and $\mathbf{Q}\mathbf{Q}' = \mathbf{Q}'\mathbf{Q} = \mathbf{I}$. That is, **Q** is orthogonal and represents a proper rotation.

We then have the relationship

$$\mathbf{Q}'\mathbf{A}\mathbf{Q} = \mathbf{D}$$

$$\begin{bmatrix} -0.8 & -0.6 \\ 0.6 & -0.8 \end{bmatrix} \begin{bmatrix} 66 & 12 \\ 12 & 59 \end{bmatrix} \begin{bmatrix} -0.8 & 0.6 \\ -0.6 & -0.8 \end{bmatrix} = \begin{bmatrix} 75 & 0 \\ 0 & 50 \end{bmatrix}$$

Finally, we see that the last matrix on the right is equal to **B**, the diagonal matrix of the new quadratic form that reorients the axes. Furthermore, if we reflect the first column of **Q**, we note that $\cos 37° = 0.8$, indicating that $\mathbf{x}_1{}^*$ makes an angle of 37° with the horizontal axis, while $\mathbf{x}_2{}^*$ makes an angle of −53° with the horizontal axis.

In brief, no new principles are involved in the present diagonalization process. As noted, **A** is symmetric to begin with, so all of our previous discussion about diagonalizing symmetric matrices is relevant here. We note that the present formula

$$\mathbf{D} = \mathbf{Q}'\mathbf{A}\mathbf{Q}$$

is the same as that found in Section 5.2:

$$\mathbf{D} = \mathbf{T}'\mathbf{A}\mathbf{T}$$

where **D** is diagonal. The matrix **Q**, an orthogonal matrix, is the same as **T** in the context of Section 5.2.

5.8.6 Types of Quadratic Forms

Quadratic forms can be classified according to the nature of the eigenvalues of the matrix of the quadratic form:

1. If all λ_i are positive, the form is said to be *positive definite.*
2. If all λ_i are negative, the form is said to be *negative definite.*
3. If all λ_i are nonnegative (positive or zero), the form is said to be *positive semidefinite.*
4. If all λ_i are nonpositive (zero or negative), the form is said to be *negative semidefinite.*
5. If the λ_i represent a mixture of positive, zero, and negative values, the form is said to be *indefinite*

In multivariate analysis we are generally interested in forms that are either positive definite or positive semidefinite. For example, if a symmetric matrix is of product-moment form (either $\mathbf{A}'\mathbf{A}$ or $\mathbf{A}\mathbf{A}'$), then it is *either positive definite or positive semidefinite.* Since various types of cross-products matrices are of this form, the cases of positive definite or positive semidefinite are of most interest to us in multivariate analysis.

5.8.7 Relating Quadratic Forms to Matrix Transformations

As might be surmised from our earlier discussion of matrix eigenstructure and basic structure, quadratic forms are intimately connected with much of the preceding material. For example, suppose we have the point transformation

$$\mathbf{u} = \mathbf{X}\mathbf{v}$$

where **X**, whose rows are sets of direction cosines, maps **v**, considered as a column vector, onto **u** in some space of interest.

To illustrate, we let

$$\mathbf{X} = \begin{bmatrix} 0.8 & 0.6 \\ 0.71 & 0.71 \end{bmatrix}$$

Hence, if $\mathbf{v} = \left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]$, we have

$$\mathbf{u} = \mathbf{X}\mathbf{v} = \begin{bmatrix} 0.8 & 0.6 \\ 0.71 & 0.71 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.8 \\ 0.71 \end{bmatrix}$$

Now suppose we want to find the squared length of **u**.

The squared length of **u** is defined to be $\mathbf{u}'\mathbf{u}$. Given **v** and the linear transformation **X**, we set up the expression

$$\mathbf{u}'\mathbf{u} = (\mathbf{X}\mathbf{v})'(\mathbf{X}\mathbf{v}) = \mathbf{v}'\mathbf{X}'\mathbf{X}\mathbf{v}$$

But now we see that $\mathbf{X}'\mathbf{X}$ is just the minor product moment of **X** which we have already discussed. We can denote this as **A**. Thus, we have

$$\mathbf{u}'\mathbf{u} = \mathbf{v}'\mathbf{A}\mathbf{v} = (1,0)\overbrace{\begin{bmatrix} 1.14 & 0.98 \\ 0.98 & 0.86 \end{bmatrix}}^{\mathbf{A}} \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\mathbf{u}'\mathbf{u} = 1.14$$

Hence, product-moment matrices, which were discussed earlier in the context of eigenstructure and SVD, also appear in the present context as matrices defining quadratic forms. That is, $\mathbf{A} = \mathbf{X}'\mathbf{X} = \mathbf{S}(X)$ is *the matrix of the quadratic form that finds the squared length of* **v** *under the linear transformation* **X**.

Up to this point we have said relatively little about the process of finding eigenstructures of *nonsymmetric* matrices. We did indicate, however, that for the matrices of interest to us in multivariate analysis their eigenstructures will involve real-valued eigenvalues and eigenvectors. Be that as it may, it is now time to discuss their eigenstructure computation.

5.9 EIGENSTRUCTURES OF NONSYMMETRIC MATRICES IN MULTIVARIATE ANALYSIS

In multivariate analysis it is not infrequently the case that we encounter various types of nonsymmetric matrices for which we desire to find an eigenstructure. Canonical correlation, multiple discriminant analysis, and multivariate analysis of variance are illustrative of techniques where this may occur.

As a case in point, let us examine the third sample problem presented in Section 1.6.4. As recalled, the twelve employees were divided into three groups on the basis of level of absenteeism. While an underlying variable, degree of absenteeism, is present in this example, let us assume that the three groups represent only an unordered polytomy.

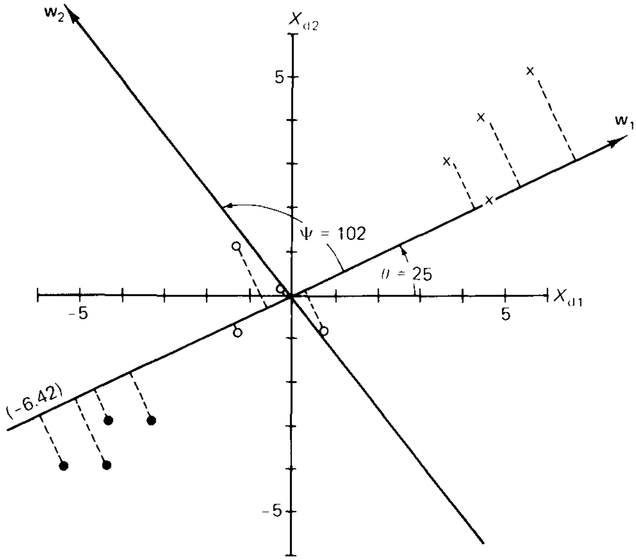

Fig. 5.10 Mean-corrected predictor variables (from Fig. 1.5). Key: • Group 1; ○ Group 2; x Group 3.

The two predictor variables were X_1 (attitude toward the firm) and X_2 (number of years employed by the firm). Figure 5.10 reproduces the scatter plot of the mean-corrected values of X_{d2} versus X_{d1}, as first shown in Fig. 1.5. The three groups have been appropriately coded by dots, circles, and small x's. We note from the figure that the individuals in the three groups show some tendency to cluster.

However, we wonder if a linear composite of X_{d1} and X_{d2} could be found that would have the property of maximally separating the three groups in the sense of maximizing their among-group variation relative to their within-group variation on this composite. Somewhat more formally, we seek a linear composite with values

$$w_{i(1)} = v_1 X_{di1} + v_2 X_{di2}$$

with the intent of maximizing the ratio

$$\boxed{\lambda_1 = \frac{SS_A(\mathbf{w}_1)}{SS_W(\mathbf{w}_1)}}$$

where SS_A and SS_W denote the among-group and within-group sums of squares of the linear composite $\mathbf{w}_1$.

We can rewrite the preceding expression in terms of quadratic forms by means of

$$\boxed{\lambda_1 = \frac{\mathbf{v}_1'\mathbf{A}\mathbf{v}_1}{\mathbf{v}_1'\mathbf{W}\mathbf{v}_1}}$$

where $\mathbf{A}$ and $\mathbf{W}$ denote among-group and (pooled) within-group SSCP matrices, respectively. Thus, we wish to find a new axis in Fig. 5.10, that can be denoted $\mathbf{w}_1$, with the property of maximizing the among- to within-group variation of the twelve points, when they are projected onto it.

The reader will note the similarity of this problem to the motivation underlying principal components analysis. Again we wish to maximize a quantity λ_1, with respect to $\mathbf{v}_1$. However, λ_1 is now considered as a *ratio of two different quadratic forms.* As such, this problem differs from principal components analysis in several significant ways.

As shown in Appendix A, the following matrix equation

$$\boxed{(\mathbf{A}-\lambda_1\mathbf{W})\mathbf{v}=\mathbf{0}}$$

is involved in the present maximization task. However, if $\mathbf{W}$ is nonsingular and hence $\mathbf{W}^{-1}$ exists, we can multiply both sides by $\mathbf{W}^{-1}$:

$$(\mathbf{W}^{-1}\mathbf{A}-\lambda_1\mathbf{I})\mathbf{v}=\mathbf{0}$$

with the resulting characteristic equation

$$\boxed{|\mathbf{W}^{-1}\mathbf{A}-\lambda_1\mathbf{I}|=0}$$

and the problem now is to solve for the eigenstructure of $\mathbf{W}^{-1}\mathbf{A}$.

So far, nothing new except for the important fact that $\mathbf{W}^{-1}\mathbf{A}$ is nonsymmetric, even though both $\mathbf{W}^{-1}$ and $\mathbf{A}$ are symmetric. Up to this point relatively little has been said about finding the eigenstructure of a nonsymmetric matrix. We can, however, proceed in two different, but related, ways.[18] First, we solve directly for the eigenstructure of the matrix involved in the current sample problem. This approach is a straightforward extension of earlier discussion involving the eigenstructure of symmetric matrices (as well as material covered in Section 5.3).

Second, we can show geometrically and algebraically an equivalent approach that involves the simultaneous diagonalization of two different quadratic forms. This presentation ties in some of the material here with previous comments on principal components analysis.

5.9.1 The Eigenstructure of $\mathbf{W}^{-1}\mathbf{A}$

Probably the most popular approach to solving for the eigenstructure of $\mathbf{W}^{-1}\mathbf{A}$ is to find the eigenvalues and eigenvectors directly, in the same general way as discussed earlier for symmetric matrices. In this case, however, $\mathbf{W}^{-1}\mathbf{A}$ is nonsymmetric; hence $\mathbf{V}$, the matrix of eigenvectors, will *not* be orthogonal.

Table 5.3 shows the preliminary calculations needed for finding the (pooled) within-group SSCP matrix $\mathbf{W}$ and the among-group SSCP matrix $\mathbf{A}$.

Table 5.4 shows the various quantities needed to solve for the eigenstructure of $\mathbf{W}^{-1}\mathbf{A}$ in terms of the characteristic equation

[18] As a matter of fact, still other ways are available to solve this problem. The interested reader can see McDonald (1968).

TABLE 5.3

Preliminary Calculations for Multiple Discriminant Analysis

Employee		X_{d1}	X_{d2}	$\mathbf{X}_k - \bar{\mathbf{X}}_k$ Within-group deviations		$\bar{\mathbf{X}}_k - \bar{\bar{\mathbf{X}}}$ Among-group deviations	
1	a	−5.25	−3.92	−1	−0.5	−4.25	−3.42
	b	−4.25	−3.92	0	−0.5	−4.25	−3.42
	c	−4.25	−2.92	0	0.5	−4.25	−3.42
	d	−3.25	−2.92	1	0.5	−4.25	−3.42
Mean		−4.25	−3.42				
2	e	−1.25	−0.92	−0.75	−0.75	−0.5	−0.17
	f	−1.25	1.08	−0.75	1.25	−0.5	−0.17
	g	−0.25	0.08	0.25	0.25	−0.5	−0.17
	h	0.75	−0.92	1.25	−0.75	−0.5	−0.17
Mean		−0.50	−0.17				
3	i	3.75	3.08	−1	−0.50	4.75	3.58
	j	4.75	2.08	0	−1.50	4.75	3.58
	k	4.75	4.08	0	0.50	4.75	3.58
	l	5.75	5.08	1	1.50	4.75	3.58
Mean		4.75	3.58				
Within-group SSCP matrix $\mathbf{W} = (\mathbf{X}_k - \bar{\mathbf{X}}_k)'(\mathbf{X}_k - \bar{\mathbf{X}}_k)$				Among-group SSCP matrix $\mathbf{A} = (\bar{\mathbf{X}}_k - \bar{\bar{\mathbf{X}}})'(\bar{\mathbf{X}}_k - \bar{\bar{\mathbf{X}}})$			

$$|\mathbf{W}^{-1}\mathbf{A} - \lambda_i \mathbf{I}| = 0$$

As noted in Table 5.4, we first compute the (pooled) within-group SSCP matrix **W** and the among-group SSCP matrix **A**.

One then finds $\mathbf{W}^{-1}$ and the matrix product $\mathbf{W}^{-1}\mathbf{A}$. From here on, the same general procedure applies for finding the eigenvalues. These turn out to be

$$\lambda_1 = 29.444; \qquad \lambda_2 = 0.0295$$

which appear in Table 5.4 along with the matrix **V** whose columns are eigenvectors of $\mathbf{W}^{-1}\mathbf{A}$. And, as indicated earlier, **V** is, in general, not orthogonal.

Returning to Fig. 5.10, we note that the first column of **V** entails direction cosines related to a 25° angle with the horizontal axis. The resulting linear composite $\mathbf{w}_1$ has scores that maximize among- to within-group variation. The second discriminant axis $\mathbf{w}_2$ (with an associated eigenvalue of only 0.0295) produces very little separation and, in cases of practical interest, would no doubt be discarded.

Other parallels with the principal components analysis of Sections 5.3 and 5.4 are found here. For example, discriminant scores—analogous to component scores—are found by projecting the points onto the discriminant axes. The discriminant score of the first observation on $\mathbf{w}_1$ is

$$w_{1(1)} = 0.905(-5.25) + 0.425(-3.92) = -6.42$$

as shown in Fig. 5.10.

TABLE 5.4

Finding the Eigenstructure of $\mathbf{W}^{-1}\mathbf{A}$

SSCP matrices of sample problem

Within-group SSCP matrix

$$\mathbf{W} = \begin{bmatrix} 6.75 & 1.75 \\ 1.75 & 8.75 \end{bmatrix}$$

Among-group SSCP matrix

$$\mathbf{A} = \begin{bmatrix} 163.50 & 126.50 \\ 126.50 & 98.17 \end{bmatrix}$$

Solving for the eigenstructure of $\mathbf{W}^{-1}\mathbf{A}$

$$\mathbf{W}^{-1} = \begin{bmatrix} 0.156 & -0.031 \\ -0.031 & 0.121 \end{bmatrix}; \qquad \mathbf{W}^{-1}\mathbf{A} = \begin{bmatrix} 21.594 & 16.698 \\ 10.138 & 7.880 \end{bmatrix}$$

Eigenvalues of $\mathbf{W}^{-1}\mathbf{A}$

$$\mathbf{\Lambda} = \begin{bmatrix} 29.444 & 0 \\ 0 & 0.0295 \end{bmatrix};$$

Eigenvectors of $\mathbf{W}^{-1}\mathbf{A}$

$$\mathbf{V} = \begin{bmatrix} 0.905 & -0.612 \\ 0.425 & 0.791 \end{bmatrix}$$

However, unlike principal components analysis, we can observe from the figure that $\mathbf{v}_1$ and $\mathbf{v}_2$ are *not* orthogonal, even though the scores on $\mathbf{w}_1$ versus $\mathbf{w}_2$ are uncorrelated. From the $\mathbf{V}$ matrix in Table 5.4 we can compute the cosine between $\mathbf{v}_1$ and $\mathbf{v}_2$ as follows:

$$\cos \Psi = (0.905 \quad 0.425)\begin{bmatrix} -0.612 \\ 0.791 \end{bmatrix} = -0.21$$

The angle Ψ separating $\mathbf{v}_1$ and $\mathbf{v}_2$ is $90^\circ + 12^\circ = 102^\circ$, as shown in Fig. 5.10.

In summary, finding the eigenstructure of the nonsymmetric matrix $\mathbf{W}^{-1}\mathbf{A}$ proceeds in an analogous fashion to the procedure followed in the case of symmetric matrices. Note, however, that the matrix of eigenvectors $\mathbf{V}$ is not orthogonal even though the discriminant scores on $\mathbf{w}_1$ and $\mathbf{w}_2$ are uncorrelated.

5.9.2 Diagonalizing Two Different Quadratic Forms

The preceding solution, while straightforward and efficient, does not provide much in the way of an intuitive guide to what goes on in the simultaneous diagonalization of two different quadratic forms:

$$\mathbf{v}_1'\mathbf{W}\mathbf{v}_1; \qquad \mathbf{v}_1'\mathbf{A}\mathbf{v}_1$$

However, we can sketch out briefly a complementary geometric and algebraic approach that relates this diagonalization problem to the earlier discussion of principal components analysis.

As recalled from Chapter 2, variances and covariances can be represented as vector lengths and angles in person space. Moreover, as shown in Fig. 5.9, quadratic forms can be pictured geometrically as ellipses in two variables, or ellipsoids in three variables, or hyperellipsoids in more than three variables.[19] The thinner the ellipse, the greater the correlation between the two variables. The tilt of the ellipse and the relative lengths of its

[19] Of course, if more than three dimensions are involved, a literal "picture" is not possible.

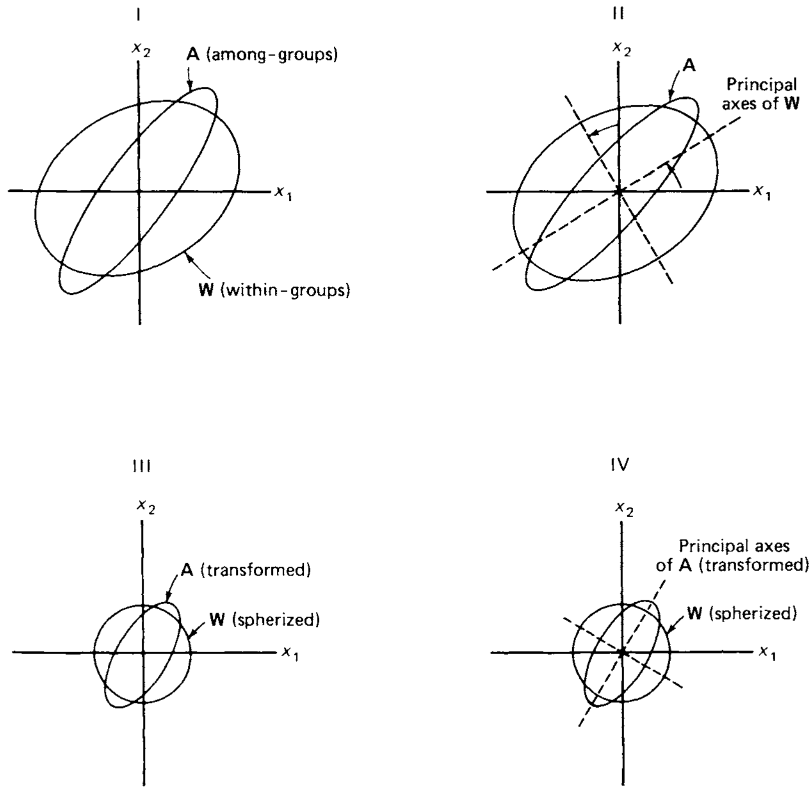

Fig. 5.11 Simultaneous diagonalization of two different quadratic forms.

axes are functions of the covariances and variances of the two variables. As we know, larger variances lead to longer (squared) lengths and also tilt the ellipse in the direction of the variable with the larger variance.

For what follows we shall use the matrix $\mathbf{Q}_1$ to refer to the matrix of eigenvectors of $\mathbf{W}^{-1/2}$ and the matrix $\mathbf{Q}_2$ to refer to the matrix of eigenvectors of the transformed matrix $\mathbf{W}^{-1/2}\mathbf{A}\mathbf{W}^{-1/2}$(as will be explained).

As motivation for this discussion, suppose we wished to find a single change of basis vectors in Panel I of Fig. 5.11 that diagonalizes both quadratic forms.[20] One quadratic form could involve a pooled within-group SSCP matrix $\mathbf{W}$. Similarly, the second form could be represented by an among-group SSCP matrix $\mathbf{A}$. We assume that the quadratic form denoting the within-group variation is positive definite. (One of the two forms *must* be positive definite for what follows.)

Geometrically, what is involved is first to rotate the within-group ellipse in Panel I to principal axes, as shown in Panel II. We then change scale, deforming the reoriented ellipse $\mathbf{W}$ to a circle in Panel III. This, in general, is called "spherizing." After this is done, *any* direction can serve as a principal axis. Hence, we can rotate the new axes to line them

[20] By diagonalization we mean a transformation in which off-diagonal elements vanish in the matrix of the quadratic form.

up with the principal axes of the *second* ellipse **A**, representing the among-group SSCP matrix in Panel IV. And that, basically, is what simultaneous diagonalization is all about.

Let us see what these geometric operations mean algebraically. First, any observation X_{dij} in Table 5.3 can be represented as the sum of

$$X_{dij} = (\mathbf{X}_k - \bar{\mathbf{X}}_k) + (\bar{\mathbf{X}}_k - \bar{\bar{\mathbf{X}}}_k)$$

$$\mathbf{X}_d = \mathbf{J} + \mathbf{G}$$

where **J** denotes the matrix of within-group deviations and **G** the matrix of among-group deviations. For example, the first observation on variable X_1 in Table 5.3 is

$$-5.25 = -1 + (-4.25)$$

where -1 indicates that it is one unit less than its group mean, and -4.25 indicates that its group mean is 4.25 units less than the grand mean. If we can find a transformation of $\mathbf{X}_d$ that spherizes the **J** portion (the within-group variation), we could then find the eigenstructure of the *adjusted* cross-products matrix.

The **J** portion can be readily spherized by the transformation:

$$\boxed{\mathbf{X}_d\mathbf{W}^{-1/2}}$$

where $\mathbf{W}^{-1/2}$, in turn, can be written as $\mathbf{Q}_1\mathbf{\Lambda}^{-1/2}\mathbf{Q}_1'$. In this case $\mathbf{\Lambda}^{-1/2}$ is a diagonal matrix of the reciprocals of the square roots of the eigenvalues of **W**, and $\mathbf{Q}_1$ is an orthogonal matrix of associated eigenvectors (since **W** is symmetric).[21]

Note, then, that what is being done here is to find the "square root" of $\mathbf{W}^{-1}$, the inverse of the (pooled) within-group SSCP matrix. To do so we recall that if **W** is symmetric and possesses an inverse $\mathbf{W}^{-1}$, we can write

$$\mathbf{W}^{-1/2}\mathbf{W}^{-1/2} = \mathbf{W}^{-1}$$

where

$$\mathbf{W}^{-1/2} = \mathbf{Q}_1\mathbf{\Lambda}^{-1/2}\mathbf{Q}_1'$$

Geometrically, the multiplication of $\mathbf{X}_d$ by $\mathbf{W}^{-1/2}$ has the effect of normalizing the within-group portion of the vectors in $\mathbf{X}_d$ to unit length, *after* rotation to the principal axes of **W** by means of the direction cosines represented by $\mathbf{Q}_1$. Subsequent rotation by $\mathbf{Q}_1'$ has no effect on what happens next, since the spherizing has already occurred.

Next we set up the equation

$$\boxed{[\mathbf{W}^{-1/2}\mathbf{A}\mathbf{W}^{-1/2}]\mathbf{Q}_2 = \mathbf{Q}_2\mathbf{\Lambda}}$$

where $\mathbf{Q}_2$ is the matrix of eigenvectors, and $\mathbf{\Lambda}$ the matrix of eigenvalues of $[\mathbf{W}^{-1/2}\mathbf{A}\mathbf{W}^{-1/2}]$. This, in turn, follows from

$$\begin{aligned}(\mathbf{X}_d\mathbf{W}^{-1/2})'(\mathbf{X}_d\mathbf{W}^{-1/2}) &= \mathbf{W}^{-1/2}\mathbf{X}_d'\mathbf{X}_d\mathbf{W}^{-1/2} = \mathbf{W}^{-1/2}\mathbf{J}'\mathbf{J}\mathbf{W}^{-1/2} + \mathbf{W}^{-1/2}\mathbf{G}'\mathbf{G}\mathbf{W}^{-1/2}\\ &= \mathbf{I} + \mathbf{W}^{-1/2}\mathbf{G}'\mathbf{G}\mathbf{W}^{-1/2} = \mathbf{I} + \mathbf{W}^{-1/2}\mathbf{A}\mathbf{W}^{-1/2}\end{aligned}$$

[21] The square root of a symmetric matrix was discussed in Section 5.5.2.

where the within-group portion has been transformed to an identity matrix **I**, as desired.[22]

We then find the eigenstructure of $[\mathbf{W}^{-1/2}\mathbf{A}\mathbf{W}^{-1/2}]$ which, given the preliminary spherization, is tantamount to a rotation to principal axes orientation. A nice feature of this procedure is that $[\mathbf{W}^{-1/2}\mathbf{A}\mathbf{W}^{-1/2}]$ is *also symmetric*. The final transformation to be applied to the *original matrix* of mean-corrected scores $\mathbf{X}_\mathrm{d}$ involves

$$\boxed{\mathbf{Y} = \mathbf{X}_\mathrm{d}\mathbf{W}^{-1/2}\mathbf{Q}_2}$$

which effects the desired simultaneous diagonalization of **W** and **A**. Note, however, that $\mathbf{W}^{-1/2}\mathbf{Q}_2$ is *not* a rotation since the data are rescaled so that the **J** portion is spherized. In summary, then, a principal components analysis of data that are first spherized in terms of pooled within-group variation provides a counterpart approach to the direct attack on finding the eigenstructure of $\mathbf{W}^{-1}\mathbf{A}$.

5.9.3 Recapitulation

While two methods have been discussed for solving

$$(\mathbf{A} - \lambda\mathbf{W})\,\mathbf{v} = \mathbf{0}$$

the first method, utilizing a direct approach to computing the eigenstructure of a nonsymmetric matrix, is probably the better known. The second procedure appears useful in its own right, however, as well as serving as an alternative method to the more usual decomposition.

If we return to the ratio of quadratic forms, stated earlier:

$$\lambda_1 = \frac{\mathbf{v}_1'\mathbf{A}\mathbf{v}_1}{\mathbf{v}_1'\mathbf{W}\mathbf{v}_1}$$

the problem of multiple discriminant analysis can be stated as one of finding extreme values of the above function where **V**, the matrix of discriminant weights, exhibits the properties:

$$\boxed{\begin{array}{l}\mathbf{V}'\mathbf{A}\mathbf{V} = \mathbf{\Lambda} \\ \mathbf{V}'\mathbf{W}\mathbf{V} = \mathbf{I}\end{array}}$$

Thus, **V** diagonalizes **A** (since $\mathbf{\Lambda}$ is diagonal) and converts **W** to an identity matrix. Notice, then, that the correlation of group means on any of the linear composite(s) is zero, since $\mathbf{\Lambda}$ is diagonal. Similarly, the *average* within-group correlation of individuals on each discriminant function (linear composite) is also zero, since **I** is an identity.

However, it should be remembered, as shown in Fig. 5.10, that **V** is *not* orthogonal since $\mathbf{W}^{-1}\mathbf{A}$, the matrix to be diagonalized, is not symmetric. Moreover, a preliminary transformation such as that applied in the second method described above, *still* ends up with a **V** that is not orthogonal.

[22] Since $\mathbf{W}^{-1/2}$ is symmetric, $\mathbf{W}^{-1/2} = (\mathbf{W}^{-1/2})'$. Since $\mathbf{J}'\mathbf{J} = \mathbf{W}$, $\mathbf{W}^{-1/2}\mathbf{J}'\mathbf{J}\mathbf{W}^{-1/2} = \mathbf{I}$.

5.10 SUMMARY

This chapter has primarily been concerned with various types of matrix decompositions—eigenstructures, singular value decompositions, and quadratic forms. The common motivation has been to search for special kinds of basis vector transformations that can be expressed in simple ways, for example, as the product of a set of matrices that individually permit straightforward geometric interpretations. In addition, such decomposition provides new perspectives on the concepts of matrix singularity and rank.

The topic was introduced by first reviewing the nature of point and basis vector changes. This introduction led to a discussion of the role of eigenstructures in rendering a given matrix (not necessarily symmetric) diagonal via nonsingular transformations. The geometric aspects of eigenstructures were stressed at this point.

We next discussed eigenstructures from a complementary view, one involving the development of linear composites with the property of maximizing the variance of point projections, subject to all composites being orthogonal with previously found composites. This time we discussed the eigenstructure of symmetric matrices with real-valued entries. In such cases all eigenvalues and eigenvectors of the matrix are real.

Since eigenstructures are not defined for rectangular matrices and their computation can present problems in the case of square, nonsymmetric matrices, we discussed these cases next in the context of the SVD. Matrix decomposition in this case involves finding a triple matrix product by which *any* matrix can be expressed as

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

where $\mathbf{P}$ and $\mathbf{Q}$ are orthonormal by columns and Δ is diagonal. This form of matrix decomposition represents a powerful organizing concept in matrix algebra, since it can be applied to any matrix of full, or less than full, rank. Furthermore, it shows that any matrix transformation can be considered as the product of a rotation–stretch–rotation or rotation–reflection–stretch–rotation under a suitable change in basis vectors.

By using product-moment matrices—$\mathbf{A}'\mathbf{A}$ or $\mathbf{A}\mathbf{A}'$, whichever has the smaller order—we were able to relate the SVD of a matrix to earlier ideas involving symmetric matrices. One can solve for the eigenstructure of $\mathbf{A}'\mathbf{A}$ (or $\mathbf{A}\mathbf{A}'$) in order to find $\mathbf{Q}'$ and Δ and then solve finally for $\mathbf{P}$. The net result is the determination of matrix rank as well as the specific geometric character of the transformation. And, if $\mathbf{A}$ is symmetric to begin with, the general procedure leads to the special case of $\mathbf{Q}'\mathbf{A}\mathbf{Q} = \mathbf{D} = \Delta^2$.

Related ideas were presented in our discussion of quadratic forms, a function that maps n-dimensional vectors into one dimension. Again, the motivation is to find a new set of basis vectors, via rotation, in which the function assumes a particularly simple form, namely, one in which cross-product terms vanish.

The last main section of the chapter dealt with ways of finding the eigenstructure of nonsymmetric matrices as they may arise in the simultaneous diagonalization of two different quadratic forms. The geometric character of this type of transformation was described and illustrated graphically.

The material of this chapter represents a major part of the more basic mathematical aspects of multivariate procedures. Typically, in multivariate analysis we are trying to find linear combinations of the original variables that optimize some quantity of interest to the researcher. As Appendix A shows, function optimization subject to certain constraints, such as Lagrange multipliers, is used time and time again in many of the statistical techniques that appear in multivariate data analysis.

REVIEW QUESTIONS

1. Form the characteristic equations of the following matrices and determine their eigenvalues and eigenvectors:

a.
$$\mathbf{A} = \begin{bmatrix} 5 & 0 \\ 7 & 8 \end{bmatrix}$$

b.
$$\mathbf{A} = \begin{bmatrix} -8 & 8 \\ -1 & -2 \end{bmatrix}$$

c.
$$\mathbf{A} = \begin{bmatrix} -3 & 5 \\ 4 & -2 \end{bmatrix}$$

d.
$$\mathbf{A} = \begin{bmatrix} 2 & -1 \\ 6 & 7 \end{bmatrix}$$

e.
$$\mathbf{A} = \begin{bmatrix} 2 & 2 & -4 \\ -2 & 4 & 2 \\ 0 & 2 & -2 \end{bmatrix}$$

f.
$$\mathbf{A} = \begin{bmatrix} 0 & 3 & 0 \\ 0 & 0 & 3 \\ 3 & -9 & 9 \end{bmatrix}$$

2. Calculate the trace and determinant of each of the first four of the matrices above and verify that

$$\operatorname{tr}(\mathbf{A}) = \sum_{i=1}^{n} \lambda_i; \qquad |\mathbf{A}| = \prod_{i=1}^{n} \lambda_i$$

3. Find the invariant vectors (i.e., eigenvectors) under the following transformations:

a. A shear
$$\begin{bmatrix} x_1^* \\ x_2^* \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

b. A stretch
$$\begin{bmatrix} x_1^* \\ x_2^* \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

c. A central dilation
$$\begin{bmatrix} x_1^* \\ x_2^* \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

d. A rotation
$$\begin{bmatrix} x_1^* \\ x_2^* \end{bmatrix} = \begin{bmatrix} 0.707 & 0.707 \\ -0.707 & 0.707 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

4. Starting with the matrix

$$\mathbf{A} = \begin{bmatrix} 5 & 1 \\ 1 & 3 \end{bmatrix}$$

find the eigenstructures of the following matrices:

a. $3\mathbf{A}$ b. $\mathbf{A} + 2\mathbf{I}$ c. $\mathbf{A} - 3\mathbf{I}$

d. $\mathbf{A}^3$ e. $\mathbf{A}^{-1}$ f. $\mathbf{A}^{1/2}$ g. $\mathbf{A}^{-1/2}$

5. Given the set of linearly independent vectors

$$\mathbf{a}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}; \qquad \mathbf{a}_2 = \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix}; \qquad \mathbf{a}_3 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$$

and the matrix

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

show that **B** can be made diagonal via the 3 x 3 matrix **A** (made up from the linearly independent vectors) and its inverse.

6. Find an orthogonal matrix **U** such that $\mathbf{U}^{-1}\mathbf{AU}$ is diagonal, where

$$\mathbf{A} = \begin{bmatrix} 7 & -2 & 1 \\ -2 & 10 & -2 \\ 1 & -2 & 7 \end{bmatrix}$$

7. Using the minor product-moment procedure, and subsequent calculation of eigenstructure, what is the rank of the following matrices:

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

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

c.
$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 2 & 2 \\ 3 & 0 \\ 4 & 0 \\ 5 & 3 \end{bmatrix}$$

d.
$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \\ 3 & 6 \\ 5 & 10 \end{bmatrix}$$

8. Find the SVD and rank of each of the following matrices:

a.
$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

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

c.
$$\mathbf{A} = \begin{bmatrix} 0.707 & 0.707 \\ -0.707 & 0.707 \end{bmatrix}$$

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

9. In the first two examples of Question 8, what is the SVD of

a. $[\mathbf{A}'\mathbf{A}]^2$ b. $[\mathbf{A}'\mathbf{A}]^{1/2}$

10. Compute $\mathbf{A}^{1/2}$ and $\mathbf{A}^{-1/2}$, by finding eigenstructures for the matrices:

a. $\mathbf{A} = \begin{bmatrix} 4 & 3 \\ 3 & 4 \end{bmatrix}$ b. $\mathbf{A} = \begin{bmatrix} 7 & 1 \\ 1 & 7 \end{bmatrix}$ c. $\mathbf{A} = \begin{bmatrix} 8 & 2 \\ 2 & 5 \end{bmatrix}$ d. $\mathbf{A} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$

11. Represent each of the following quadratic forms by a real symmetric matrix and determine its rank

a. $x_1^2 + 2x_1x_2 + x_2^2$ b. $x_1^2 + 2x_2^2 - 4x_1x_2$

c. $9x_1^2 - 6x_1x_2 + x_2^2$ d. $2x_1^2 - 3x_1x_2 + 3x_2^2$

12. Diagonalize the matrix of each quadratic form in Question 11 and describe its geometric character.

13. In the sample problem, whose mean-corrected data appear in Table 1.2:

a. Find the covariance matrix of the full set of three variables.

b. Compute the principal components of the three-variable covariance matrix and the matrix of component scores.

c. Compare these results with those found in the present chapter.

14. Simplify the following quadratic forms and indicate the type of definiteness of each form:

a. $\mathbf{x}'\begin{bmatrix} 9 & -3 \\ -3 & 1 \end{bmatrix}\mathbf{x}$ b. $\mathbf{x}'\begin{bmatrix} -4 & 2 & -1 \\ 2 & -4 & 2 \\ -1 & 2 & -4 \end{bmatrix}\mathbf{x}$

c. $\mathbf{x}'\begin{bmatrix} 2 & -1.5 \\ -1.5 & 3 \end{bmatrix}\mathbf{x}$ d. $\mathbf{x}'\begin{bmatrix} 4 & 2 & -2 \\ 2 & 4 & 2 \\ -2 & 2 & 4 \end{bmatrix}\mathbf{x}$

15. Returning to the multiple discriminant function problem considered in Section 5.9:

a. Spherize the (pooled) within-group SSCP matrix and compute the eigenstructure in accordance with the procedure outlined in Section 5.9.2.

b. Compare these results with those found from the procedure used in Section 5.9.1.