Appendix 1

The Geometry of Vectors and Matrices:

Eigenvalues and Eigenvectors

Please email any comments/corrections to: jbwalsh@u.arizona.edu

This chapter introduces a variety of tools from matrix algebra and multivariate statistics useful in the analysis of selection on multiple characters. Our primary intent is to introduce the reader to the idea of vectors and matrices as geometric structures, viewing matrix operations as transformations converting a vector into a new vector by a change in geometry (rotation and scaling). The eigenvalues and their associated eigenvectors of a matrix describe the geometry of the transformation associated with that matrix.

THE GEOMETRY OF VECTORS AND MATRICES

There are numerous excellent texts on matrix algebra, so we will make little effort to prove most of the results given below. For statistical applications, concise introductions can be found in the chapters on matrix methods in Johnson and Wichern (1988) and Morrison (1976), while Dhrymes (1978) and Searle (1982) provide a more extended treatment. Wilf’s (1978) short chapter on matrix methods provides a very nifty review of methods useful in applied mathematics. Franklin (1968), Horn and Johnson (1985), and Gantmacher (1960), respectively, give increasingly sophisticated treatments of matrix analysis.

Comparing Vectors: Lengths and Angles

As Figure A1.1 shows, a vector $\mathbf{x}$ can be treated as a geometric object, an arrow leading from the origin to the $n$-dimensional point whose coordinates are given by the elements of $\mathbf{x}$. By changing coordinate systems, we change the resulting vector, potentially changing both its direction (rotating the vector) and length (scaling the vector). This geometric interpretation suggests several ways for comparing vectors, such as the angle between two vectors and the projection of one vector onto another.
Figure A1.1. Some basic geometric concepts of vectors. While we use examples from two dimensions, these concepts easily extend to \( n \) dimensions. A: A vector \( \mathbf{x} \) can be thought of as an arrow from the origin to a point in space whose coordinates are given by the elements of \( \mathbf{x} \). B: Multiplying a vector by \(-1\) results in a reflection about the origin. C: One measure of the difference in direction between two vectors is the angle \( \theta \) between them. D: \( \text{Proj}(\mathbf{b} \text{ on } \mathbf{A}) \) is the vector resulting from the projection of \( \mathbf{b} \) on \( \mathbf{A} \). Note that the resulting projection vector is either in the same direction as \( \mathbf{A} \) or in the direction of the reflection of \( \mathbf{A} \), as seen for \( \text{Proj}(\mathbf{c} \text{ on } \mathbf{A}) \).

Consider first the length (or norm) of a vector. The most common length measure is the Euclidean distance of the vector from the origin, \( ||\mathbf{x}|| \), which is defined by

\[
||\mathbf{x}|| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} = \sqrt{\mathbf{x}^T \mathbf{x}} \tag{A1.1a}
\]

Hence for any scalar \( a \), \( ||a \mathbf{x}|| = |a| ||\mathbf{x}|| \). If \( a < 0 \), the vector \( a \mathbf{x} \) is scaled by \( |a| \) and reflected about the origin as is shown in Figure A1.1. Similarly, the Euclidean distance between \( \mathbf{x} \) and \( \mathbf{y} \) is

\[
||\mathbf{x} - \mathbf{y}||^2 = \sum_{i=1}^{n} (x_i - y_i)^2 = (\mathbf{x} - \mathbf{y})^T(\mathbf{x} - \mathbf{y}) = (\mathbf{y} - \mathbf{x})^T(\mathbf{y} - \mathbf{x}) \tag{A1.1b}
\]

Vectors can differ by length, direction, or both. The angle \( \theta \) between two vectors \( \mathbf{x} \) and \( \mathbf{y} \) provides a measure of how much they differ in direction (Figure A1.1). If the vectors satisfy \( a \mathbf{x} = \mathbf{y} \) where \( a > 0 \) they point in exactly the same direction, and they are defined to be zero degrees apart. If \( a < 0 \), they are exactly 180 degrees apart and differ in direction only by a reflection of the axes about the origin. At the other extreme, two vectors can be at right angles to each other \( (\theta = 90^\circ \text{ or } 270^\circ) \), in which case the vectors are said to be orthogonal. Orthogonal vectors of unit length are further said to be orthonormal. For any two \( n \) dimensional vectors, \( \theta \) satisfies

\[
\cos(\theta) = \frac{\mathbf{x}^T \mathbf{y}}{||\mathbf{x}|| \ ||\mathbf{y}||} = \frac{\mathbf{y}^T \mathbf{x}}{||\mathbf{x}|| \ ||\mathbf{y}||} \tag{A1.2}
\]

Note that since \( \cos(90^\circ) = \cos(270^\circ) = 0 \), two vectors are orthogonal if and only if their inner product is zero, \( \mathbf{x}^T \mathbf{y} = 0 \).
Another way to compare vectors, illustrated in Figure A1.1, is to consider the projection of one vector on another. For any two \( n \) dimensional vectors, the projection of \( x \) on \( y \) generates a vector defined by

\[
\text{Proj}(x \text{ on } y) = \frac{x^T y}{y^T y} y = \left( \cos(\theta) \frac{||x||}{||y||} \right) y
\]

If \( ||y|| = 1 \), then

\[
\text{Proj}(x \text{ on } y) = (x^T y) y = (\cos(\theta) ||x||) y
\]

Note that since the term involving cosines in Equations A1.3a/b is a scalar, the vector resulting from the projection of \( x \) on \( y \) is in the same direction as \( y \), unless \( 90^\circ < \theta < 270^\circ \) in which case \( \cos(\theta) < 0 \) and the projection vector is in exactly the opposite direction (the reflection of \( y \) about the origin). The length of the projection vector is

\[
||\text{Proj}(x \text{ on } y)|| = |\cos(\theta)| ||x||
\]

If two vectors lie in exactly the same direction, the projection of one on the other just recovers the vector (\( \text{Proj}(x \text{ on } y) = x \)). Conversely, if two vectors are orthogonal, then the projection of one on the other yields a vector of length zero. An important use of projection vectors is that if \( y_1, y_2, \ldots, y_n \) is any set of mutually orthogonal \( n \) dimensional vectors, then any \( n \) dimensional vector \( x \) can be represented as the sum of projections of \( x \) onto the members of this set,

\[
x = \sum_{i=1}^{n} \text{Proj}(x \text{ on } y_i)
\]

Matrices Describe Vector Transformations

When we multiply a vector \( x \) by a matrix \( A \) to create a new vector \( y = Ax \), \( A \) rotates and scales the original vector \( x \) to give \( y \). Thus \( A \) describes a transformation of the original coordinate system of \( x \) into a new coordinate system \( y \) (which has a different dimensions than \( x \) unless \( A \) is square).

**Example 1.** Consider the multivariate version of the breeders’ equation, \( \Delta \mu = G \beta \), which holds when certain normality assumptions are satisfied (we will discuss this in detail shortly). Here \( \Delta \mu \) is the resulting change in the vector of phenotypic means, \( G \) the variance-covariance matrix of additive genetic values (breeding values) of these characters, and \( \beta \) the directional selection gradient (the direction of change in character means that results in the greatest increase in mean population fitness when phenotypes are multivariate normally distributed). Suppose

\[
G = \begin{pmatrix}
4 & -2 \\
-2 & 2
\end{pmatrix}
\quad \beta = \begin{pmatrix}
1 \\
3
\end{pmatrix}
\]
hence
\[ \Delta \mu = G\beta = \begin{pmatrix} -2 \\ 4 \end{pmatrix} \]

The resulting change in character means are different from those most favored by natural selection. Selection favors an increase in \( z_1 \), but when the genetic variance-covariance structure is taken into account, the resulting change in \( z_1 \) is negative. Taking the appropriate innerproducts, we find \( ||\beta|| = \sqrt{10}, ||\Delta \mu|| = \sqrt{20} \), and \( \beta^T \Delta \mu = 10 \). Applying Equation A1.2,

\[ \cos \theta = \frac{\beta^T \Delta \mu}{||\Delta \mu|| \cdot ||\beta||} = \frac{1}{\sqrt{2}} \]

Thus the response vector is \( \cos^{-1}(1/\sqrt{2}) = 45^\circ \) from the selection gradient, implying that the constraints introduced by the genetic variance-covariance matrix moves the response vector considerably away from the direction most favored by natural selection (see Figure A1.2).

---

**Figure A1.2.** Under appropriate normality assumptions, the genetic variance-covariance matrix \( G \) transforms the selection gradient \( \beta \) into the response vector \( \Delta \mu \). Using the values of \( \beta \) and \( G \) from Example 1, observe that \( G \) translates the directional selection gradient in a counter-intuitive fashion. This behavior results from the strong negative additive genetic covariance between \( z_1 \) and \( z_2 \), as will become more obvious when we consider the eigenvectors of \( G \) (Figure A1.3).

---

**Orthonormal Matrices**

Matrix transformations consist of two basic operations, rotations (changes in the direction of a vector) and scalings (changes in its length). We can partition a matrix transformation into these two basic operations by using orthonormal matrices. Writing a square matrix \( U \) as \( U = (u_1, u_2, \ldots, u_n) \) where each \( u_i \) is an \( n \) dimen-
sional column vector, \( U \) is orthonormal if
\[
\begin{cases} 
1 & \text{if } i = j \\
0 & \text{if } i \neq j 
\end{cases}
\]

In other words, each column of \( U \) is independent from every other column and has unit length. Matrices with this property are also referred to as unitary, or orthogonal and satisfy
\[
U^T U = UU^T = I \tag{A1.5a}
\]

Hence,
\[
U^T = U^{-1} \tag{A1.5b}
\]

The coordinate transformation induced by an orthonormal matrix has a very simple geometric interpretation in that it is a rigid rotation of the original coordinate system — all axes of the original coordinate are simply rotated by the same angle to create the new coordinate system. To see this, note first that orthonormal matrices preserve all innerproducts. Taking \( y_1 = Ux_1 \) and \( y_2 = Ux_2 \),
\[
y_1^T y_2 = x_1^T (U^T U)x_2 = x_1^T x_2
\]

A special case of this is that orthonormal matrices do not change the length of vectors, as \( ||y_1|| = y_1^T y_1 = x_1^T x_1 = ||x_1|| \). If \( \theta \) is the angle between vectors \( x_1 \) and \( x_2 \), then following transformation by an orthonormal matrix,
\[
\cos(\theta | y_1, y_2) = \frac{y_1^T y_2}{\sqrt{||y_1|| \cdot ||y_2||}} = \frac{x_1^T x_2}{\sqrt{||x_1|| \cdot ||x_2||}} = \cos(\theta | x_1, x_2)
\]

and the angle between any two vectors remains unchanged following their transformation by the same orthonormal matrix.

**Eigenvalues and Eigenvectors**

The eigenvalues and their associated eigenvectors of a square matrix describe the geometry of the transformation induced by that matrix. Eigenvalues describe how the original coordinate axes are scaled in the new coordinate system while eigenvectors describe how the original axes are rotated.

Suppose that the vector \( y \) satisfies the matrix equation
\[
Ay = \lambda y \tag{A1.6}
\]

for some scalar value \( \lambda \). Geometrically, this means that the new vector resulting from transformation of \( y \) by \( A \) points in the same direction (or is exactly reflected about the origin if \( \lambda < 0 \)) as \( y \). For such vectors, the only action of the matrix transformation is to scale them by some amount \( \lambda \). Hence, it is natural that the
system of such vectors along with their corresponding scalar multipliers completely describes the geometry of the transformation associated with $A$. Vectors satisfying Equation A1.6 are referred to as **eigenvectors** and their associated scaling factors are **eigenvalues**. If $y$ is an eigenvector, then $ay$ is also an eigenvector as $A(ay) = a(Ay) = \lambda(ay)$. Note, however, that the associated eigenvalue remains unchanged. Hence, we typically scale eigenvectors to be of unit length to give unit or **normalized** eigenvectors. In particular, if $u_i$ is the eigenvector associated with the $i$th eigenvalue, then the associated normalized eigenvector is $e_i = u_i/||u_i||$.

The eigenvalues of square matrix $A$ of dimension $n$ are solution of Equation A1.6, which is usually expressed as the **characteristic equation** $|A - \lambda I| = 0$. This can be also be expressed using the **Laplace expansion** as

$$
|A - \lambda I| = (-\lambda)^n + S_n(-\lambda)^{n-1} + \cdots + S_1(-\lambda) + S_0 = 0
$$

(A1.7)

where $S_i$ is the sum of all **principal minors** (minors including diagonal elements of the original matrix) of order $i$. Minors were defined in Chapter 7. Finding the eigenvalues thus requires solving a polynomial equation of order $n$. In practice, for $n > 2$ this is usually done numerically, and most statistical and numerical analysis packages offer routines to accomplish this task.

Two of these principal minors are easily obtained and provide some information on the nature of the eigenvalues. The only principal minor having the same order of the matrix is the full matrix itself, so that $S_n = |A|$, the determinant of $A$. $S_1$ is also related to an important matrix quantity, the **trace**. This is denoted by $\text{tr}(A)$ and is the sum of the diagonal elements of the matrix,

$$
\text{tr}(A) = \sum_{i=1}^{n} a_{ii}
$$

Observe that $S_1 = \text{tr}(A)$ as the only principal minors of order one are the diagonal elements themselves, the sum of which equals the trace. The trace and determinant can be expressed as functions of the eigenvalues,

$$
\text{tr}(A) = \sum_{i=1}^{n} \lambda_i \quad \text{and} \quad |A| = \prod_{i=1}^{n} \lambda_i
$$

(A1.8)

Hence $A$ is singular ($|A| = 0$) if and only if at least one eigenvalue is zero.

Let $e_i$ be the (unit-length) eigenvector associated with eigenvalue $\lambda_i$. If the eigenvectors of $A$ can be chosen to be mutually orthogonal, e.g., $e_i^T e_j = 0$ for $i \neq j$, we can express $A$ as

$$
A = \lambda_1 e_1 e_1^T + \lambda_2 e_2 e_2^T + \cdots + \lambda_n e_n e_n^T
$$

(A1.9b)

This decomposition is called the **spectral decomposition** of $A$. Hence,

$$
A x = \lambda_1 e_1 e_1^T x + \lambda_2 e_2 e_2^T x + \cdots + \lambda_n e_n e_n^T x
= \lambda_1 \text{Proj}(x \text{ on } e_1) + \lambda_2 \text{Proj}(x \text{ on } e_2) + \cdots + \lambda_n \text{Proj}(x \text{ on } e_n)
$$

(A1.9b)
The last equality follows since $e_i(e_i^T x) = (e_i^T x)e_i$ as $e_i^T x$ is a scalar.

Example 2. Determine the eigenstructure (the set of eigenvalues and their associated unit eigenvectors) of the genetic variance-covariance matrix $G$ given in Example 1. Writing the characteristic equation,

$$|G - \lambda I| = \begin{vmatrix} 4 - \lambda & -2 \\ -2 & 2 - \lambda \end{vmatrix} = (4 - \lambda)(2 - \lambda) - (-2)^2 = \lambda^2 - 6\lambda + 4 = 0$$

Alternatively, using the Laplace expansion, $\text{tr}(G) = 4 + 2 = 6$ and $|G| = 4 \times 2 - (-2)^2 = 4$ also recovers the characteristic equation, which has solutions

$$\lambda_1 = 3 + \sqrt{5} \simeq 5.236 \quad \lambda_2 = 3 - \sqrt{5} \simeq 0.764$$

The associated unit eigenvectors are

$$e_1 \simeq \begin{pmatrix} -0.851 \\ 0.526 \end{pmatrix} \quad e_2 \simeq \begin{pmatrix} 0.526 \\ 0.851 \end{pmatrix}$$

These are orthogonal as $e_1^T e_2 = 0$. From Example 1, $\|\beta\| = \sqrt{10}$, while $e_1^T \beta \simeq 0.727$ and $e_2^T \beta \simeq 3.079$. Since $\|e_1\| = \|e_2\| = 1$,

$$\cos(\theta |e_1, \beta|) \simeq \frac{0.727}{\sqrt{10}} \simeq 0.201 \quad \text{and} \quad \cos(\theta |e_2, \beta|) \simeq \frac{3.079}{3.162} \simeq 0.974$$

giving the angle between $e_1$ and $\beta$ as $\theta |e_1, \beta| \simeq 78.4^\circ$, while $\theta |e_2, \beta| \simeq 13.2^\circ$. The corresponding scaled projections of $\beta$ on these eigenvectors are

$$\lambda_1 \text{Proj}(\beta \text{ on } e_1) \simeq \begin{pmatrix} -3.236 \\ 2 \end{pmatrix} \quad \text{and} \quad \lambda_2 \text{Proj}(\beta \text{ on } e_2) \simeq \begin{pmatrix} 1.236 \\ 2 \end{pmatrix}$$

From Equation A1.9b, $\beta$ is the sum of these two projections. As Figure A1.3 shows, the eigenstructure of $G$ explains the unusual behavior of response seen in Figure A1.2. The eigenvector associated with the leading eigenvalue $\lambda_1$ accounts for most of the variation inherent in $G$, and this eigenvector corresponds to a strong negative correlation between the additive genetic values of $z_1$ and $z_2$. Hence, even though $\beta$ points in very much the same direction as $e_2$, because $\lambda_1 >> \lambda_2$ the projection of $\beta$ on $e_1$ gives a vector of greater length than the projection on $e_2$, and it is this projection vector that results in the decrease in $\mu_{z_1}$. 
Properties of Symmetric Matrices

Many of the matrices we will encounter are symmetric, satisfying $A = A^T$. Examples include variance-covariance matrices and the $\gamma$ matrix of quadratic coefficients in the Pearson-Lande-Arnold fitness regression. Here we give some of the more useful properties of symmetric matrices. Proofs can be found in Dhrymes (1978), Horn and Johnson (1985), and Wilf (1978).

1. If $A$ is symmetric, then if $A^{-1}$ exists, it is also symmetric.
2. The eigenvalues and eigenvectors of a symmetric matrix are all real.
3. For any $n$-dimensional symmetric matrix, a corresponding set of orthonormal eigenvectors can be constructed, i.e., we can obtain a set of eigenvalues $e_i$ for $1 \leq i \leq n$ that satisfies

$$e_i^T e_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

In particular, this guarantees that a spectral decomposition of $A$ exists. This can be restated as:

4. A symmetric matrix $A$ can be diagonalized as

$$A = U \Lambda U^T \quad (A1.10a)$$

where $\Lambda$ is a diagonal matrix, and $U$ is an orthonormal matrix ($U^{-1} = U^T$). If $\lambda_i$ and $e_i$ are the $i$th eigenvalue and its associated unit eigenvector.

Figure A1.3. Left: The scaled eigenvectors associated with the variance-covariance matrix $G$ from Example 1. Note that $e_1$ and $e_2$ are orthogonal and hence can be thought of as describing a new coordinate system. Right: Since $\lambda_1 >> \lambda_2$, the leading eigenvector largely dominates the transformation. This is shown by taking the projections of $\beta$ on each of these eigenvectors. Even though $\beta$ is nearly orthogonal to $e_2$, the projection of $\beta$ on $e_1$ yields a vector of greater length than the projection of $\beta$ on $e_2$. From Equation A1.16b, the response to selection $\Delta \mu$ is the sum of these two projections.
of $\mathbf{A}$, then
\[
\mathbf{A} = \text{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n) = \begin{pmatrix}
\lambda_1 & 0 & \cdots & 0 \\
0 & \lambda_2 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & \cdots & \cdots & \lambda_n
\end{pmatrix}
\] (A1.10b)

and
\[
\mathbf{U} = (\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_n)
\] (A1.10c)

Geometrically, $\mathbf{U}$ describes a rigid rotation of the original coordinate system while $\mathbf{A}$ is the amount by which unit lengths in the original coordinate system are scaled in the transformed system. Using Equation A1.10a, it is easy to show that
\[
\mathbf{A}^{-1} = \mathbf{U}\mathbf{A}^{-1}\mathbf{U}^T
\] (A1.11a)

\[
\mathbf{A}^{1/2} = \mathbf{U}\mathbf{A}^{1/2}\mathbf{U}^T
\] (A1.11b)

where the square root matrix $\mathbf{A}^{1/2}$ (which is also symmetric) satisfies $\mathbf{A}^{1/2}\mathbf{A}^{1/2} = \mathbf{A}$. Since $\mathbf{A}$ is diagonal, the $i$th diagonal elements of $\mathbf{A}^{-1}$ and $\mathbf{A}^{1/2}$ are $\lambda_i^{-1}$ and $\lambda_i^{1/2}$ respectively, implying that if $\lambda_i$ is an eigenvalue of $\mathbf{A}$, then $\lambda_i^{-1}$ and $\sqrt{\lambda_i}$ are eigenvalues of $\mathbf{A}^{-1}$ and $\mathbf{A}^{1/2}$. Note that Equations A1.11a/b imply that $\mathbf{A}$, $\mathbf{A}^{-1}$ and $\mathbf{A}^{1/2}$ all have the same eigenvectors. Finally, using Equation A1.10a we see that premultiplying $\mathbf{A}$ by $\mathbf{U}^T$ and then postmultiplying by $\mathbf{U}$ gives a diagonal matrix whose elements are the eigenvalues of $\mathbf{A}$,

\[
\mathbf{U}^T \mathbf{A} \mathbf{U} = \mathbf{U}^T (\mathbf{U} \mathbf{A} \mathbf{U}^T) \mathbf{U} = (\mathbf{U}^T \mathbf{U}) \mathbf{A} (\mathbf{U}^T \mathbf{U})
\]

\[
= \mathbf{A}
\] (A1.12)

---

**Example 3.** Consider the additive genetic variance-covariance matrix $\mathbf{G}$ from Examples 1 and 2. Recalling the results from Example 2 and using Equation A1.10a, we can express $\mathbf{G}$ as $\mathbf{U} \mathbf{A} \mathbf{U}^T$, where

\[
\mathbf{A} = \begin{pmatrix}
5.241 & 0 \\
0 & 0.765
\end{pmatrix}
\]

and

\[
\mathbf{U} = (\mathbf{e}_1, \mathbf{e}_2) = \begin{pmatrix}
-0.851 & 0.526 \\
0.526 & 0.851
\end{pmatrix}
\]

From Equation A1.11a, the eigenvalues of $\mathbf{A}^{-1}$ are $(5.241)^{-1} \approx 0.191$ and $(0.765)^{-1} \approx 1.307$, while from Equation A1.11b, the eigenvalues of $\mathbf{A}^{1/2}$ are $\sqrt{5.241} \approx 2.289$ and $\sqrt{0.765} \approx 0.875$. 
5. The Rayleigh-Ritz theorem gives useful bounds on quadratic products associated with the symmetric matrix $A$: if the eigenvalues of $A$ are ordered as $\lambda_{\text{max}} = \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n = \lambda_{\text{min}}$, then for any vector of constants $c$,

$$
\lambda_1 ||c|| \geq c^T A c \geq \lambda_n ||c|| \quad (A1.13a)
$$

Alternatively, if $c$ is of unit length

$$
\max_{||c||=1} c^T A c = \lambda_1 \quad (A1.13b)
$$
$$
\min_{||c||=1} c^T A c = \lambda_n \quad (A1.13c)
$$

Where the maximum and minimum occur when $c = e_1$ and $c = e_n$, the eigenvectors associated with $\lambda_1$ and $\lambda_n$.

This is an especially useful result for bounding variances. Consider a univariate random variable $y = c^T x$ formed by a linear combination of the elements of a random vector $x$. From Equation 7.20, $\sigma^2(y) = c^T \Sigma_x c$ and applying Equation A1.13a,

$$
\lambda_1 ||c||^2 \geq \sigma^2(y) \geq \lambda_n ||c||^2 \quad (A1.14)
$$

where $\lambda_1$ is the largest (leading or dominant) and $\lambda_n$ the smallest eigenvalue of $\Sigma_x$.

Correlations can be Removed by a Matrix Transformation

A particularly powerful use of diagonalization is that it allows us to extract a set of $n$ uncorrelated variables when the variance-covariance matrix $\Sigma_x$ is nonsingular and of dimension $n$. Consider the transformation

$$
y = U^T x \quad (A1.15a)
$$

where $U = (e_1, e_2, \cdots, e_n)$ contains the normalized eigenvectors of $\Sigma_x$. Since $U$ is an orthonormal matrix, this transformation is a rigid rotation of the axes of the original $(x_1, \cdots, x_n)$ coordinate system to a new system given by $(y_1, \cdots, y_n)$. Applying Equations 7.22b and A1.12, the variance-covariance matrix for $y$ is

$$
\Sigma_y = U^T \Sigma_x U = \Lambda \quad (A1.15b)
$$

where $\Lambda$ is a diagonal matrix whose elements are the eigenvalues of $\Sigma_x$, so that

$$
\sigma(y_i, y_j) = \begin{cases} 
\lambda_i & \text{if } i = j \\
0 & \text{if } i \neq j 
\end{cases}
$$
The rigid rotation introduced by $U$ thus creates a set of $n$ uncorrelated variables, the $i$th of which is

$$y_i = e_i^T x$$  \hspace{1cm} (A1.15c)

Since $e_i$ are defined to be of unit length, from Equation A1.3b we have $y_i = e_i^T x = \text{Proj}(x \text{ on } e_i)$, so that this new variable is the projection of $x$ onto the $i$th eigenvector of $\Sigma_x$, implying that the axes of new coordinate system are given by the orthogonal set of eigenvectors of $\Sigma_x$.

**Example 4.** Applying the change of variables suggested by Equation A1.15a to the vector $z$ of characters with associated $G$ matrix used in Example 1 and using the eigenvalues and vectors obtained in Example 2 gives

$$y = U^T z = \begin{pmatrix} e_1^T \\ e_2^T \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} -0.851 & 0.526 \\ 0.526 & 0.851 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} -0.851z_1 + 0.526z_2 \\ 0.526z_1 + 0.851z_2 \end{pmatrix}$$

From Equation A1.15b, $\Sigma_y = \Lambda$ as given in Example 3, so that $y_1$ and $y_2$ are uncorrelated with $\sigma^2(y_1) = \lambda_1 = 5.241$ and $\sigma^2(y_2) = \lambda_2 = 0.765$. Hence, by considering the new coordinate system with $y_1 = e_1^T z = -0.851z_1 + 0.526z_2$ and $y_2 = e_2^T z = 0.526z_1 + 0.851z_2$ we can transform the original coordinate system into a new system on which there are no additive genetic correlations between these new characters. Figure A1.4 shows that this transformation is just a rigid rotation of the axes.
eigenvectors $e_1$ and $e_2$. The angle between the new axis $e_1$ and the original $z_1$ axis is given by taking the angle between $e_1$ and $z_1 = (1, 0)^T$. $||e_1|| = ||z_1|| = 1$ and $e_1^T z_1 = 0.851$, giving $\theta = \cos^{-1}(0.851) \approx 32^\circ$. The angle between $e_2$ and the $z_2$ axis is also 32 degrees.

Canonical Axes of Quadratic Forms

The transformation $y = U^T x$ given by Equation A1.15a applies to any symmetric matrix, and is referred to as the canonical transformation associated with that matrix. The canonical transformation simplifies the interpretation of the quadratic form $x^T A x$ as rotation of the original axes to align them with the eigenvalues of $A$ removes all cross-product terms on this new coordinate system. Applying Equations A1.15a and A1.12 transforms the quadratic form to one in which the square matrix is diagonal,

$$x^T A x = (Uy)^T AUy = y^T (U^T AU)y = \sum_{i=1}^{n} \lambda_i y_i^2$$

(A1.16)

where $\lambda_i$ and $e_i$ are the eigenvalues and associated (normalized) eigenvectors of $A$ and $y_i = e_i^T x$. The new axes defined by $e_i$ are the canonical (or principal) axes. Since the $y_i^2 \geq 0$, Equation A1.16 immediately shows the connection between the signs of the eigenvalues of a matrix and whether that matrix is positive definite, negative definite, or indefinite. If all eigenvalues are positive (all $\lambda_i > 0$), then the quadratic form is always positive (unless all the $y_i$ are zero) and hence $A$ is positive definite. If all eigenvalues are negative (all $\lambda_i < 0$), then $A$ is negative definite as the quadratic form is always negative. If at least one eigenvalue is zero, then $A$ is at most semidefinite, while if $A$ has both positive and negative eigenvalues it is indefinite.

Equations of the form

$$x^T A x = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j = c^2$$

(A1.17a)

arise fairly frequently. For example, they describe surfaces of constant variance (tracing out the surface created by vectors $b$ such that $b^T x$ has constant variance $c^2$, see Figure A1.5) or surfaces of constant fitness in quadratic fitness regressions (those vectors of phenotypic values $z$ such that $w(z) = a + (z - \mu)^T \gamma (z - \mu)$ is constant). Solutions to Equation A1.17a describe quadratic surfaces — for two dimensions these are the familiar conic sections (ellipses, parabolas, or hyperbolas).
Equation A1.16 greatly simplifies the interpretation of these surfaces by removing all cross product terms, giving

\[ x^T A x = \sum_{i=1}^{n} \lambda_i y_i^2 = c^2 \]  

(A1.17b)

Since \((y_i)^2\) and \((-y_i)^2\) have the same value, the canonical axes of \(A\) are also the axes of symmetry for the quadratic surface generated by quadratic forms involving \(A\). When all the eigenvalues of \(A\) are positive (as occurs with nonsingular variance-covariance and other positive definite matrices), Equation A1.17b describes an ellipsoid whose axes of symmetry are given by the eigenvectors of \(A\). The distance from the origin to the surface along the axis given by \(e_i\) is \(\lambda_i y_i^2 = c^2\) or \(y_i = c\lambda_i^{-1/2}\), as can be seen by setting all the \(y_k\) equal to zero except for \(y_i\), giving \(x^T A x = \lambda_i y_i^2 = c^2\). Figure A1.5 shows an example of a two-dimensional constant-variance surface: if we plot the entire set of vectors \(b\) such that the variable \(y = b^T x\) has variance \(c^2 = b^T \Sigma x b\), the tips of these vectors sweep out the ellipse.

![Figure A1.5](image)

**Figure A1.5.** The general shape of surfaces of constant variance for the additive genetic variance-covariance matrix \(G\) given in Example 1. Defining a new composite character \(y = az_1 + bz_2\), the rotated ellipse represents the set of \((a, b)\) values that give \(y\) the same additive genetic variance \(c^2\). The major axis of the ellipse is along \(e_2\), the eigenvector associated with the smallest eigenvalue of \(G\), where \(\lambda_2 \approx 0.765\), giving \(\lambda_2^{-1/2} \approx 1.143\). The minor axis of the ellipse is along \(e_1\), the eigenvector associated with the largest eigenvalue of \(G\), where \(\lambda_1 \approx 5.241\), giving \(\lambda_1^{-1/2} \approx 0.437\).

**Principal Components of the Variance-Covariance Matrix**

We are very interested in how the variance of a random vector can be decomposed into independent components. For example, even though we may be measuring \(n\) variables, only one or two of these may account for the majority of the variation. If this is the case we may wish to exclude those variables contributing very
little variation from further analysis. More generally, if the random variables are correlated, then certain linear combinations of the elements of \( x \) may account for most of the variance. The procedure of **principal component analysis** extracts these combinations by decomposing the variance of \( x \) into a series of orthogonal vectors, the first of which explains the most variation possible for any single vector, the second the next possible amount, and so on until the entire variance of \( x \) has been accounted for.

Consider Figure A1.5. Since the set of points comprising the ellipse represents those linear combinations of the random variables of \( z \) giving equal variance, we see that the closer a point on this curve is to the origin, the more variance there is in that direction. The points closest to the origin are those that lie along the axis defined by \( e_1 \), while those furthest away lie along the axis defined by \( e_2 \). Here \( e_1 \) and \( e_2 \) are the principal components of \( G \), with the first principal component accounting for most of the variation of \( G \). In particular, the ratio of additive variances for the characters \( y_1 = e_1^T z \) and \( y_2 = e_2^T z \) is

\[
\frac{\sigma^2(y_1)}{\sigma^2(y_2)} = \frac{\sigma^2(e_1^T z)}{\sigma^2(e_2^T z)} = e_1^T G e_1 / e_2^T G e_2 = \frac{\lambda_1}{\lambda_2} \approx \frac{5.241}{0.765} \approx 6.85,
\]

so that a character in the direction of \( e_1 \) has almost seven times as much additive variance as a character lying in the direction of \( e_2 \).

In general, suppose we have an \( n \)-dimensional variance-covariance matrix \( \Sigma_x \). Ordering the eigenvalues of \( \Sigma_x \) as \( \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \), then from Equation A1.13b the maximum variance for any linear combination \( y = c_1^T x \) (subject to the constraint that \( ||c_1|| = 1 \)) is

\[
\max \sigma^2(y) = \max_{||c_1||=1} \sigma^2(c_1^T x) = c_1^T \Sigma_x c_1 = \lambda_1
\]

which occurs when \( c_1 = e_1 \). This vector the first principal component (often abbreviated as PC1). Excluding PC1, consider how the remaining variance can be explained. The vector \( c_2 \) orthogonal to PC1 (e.g., \( c_2^T c_1 = 0 \)) that maximizes the remaining variance, e.g., maximizes \( \sigma^2(c_2^T x) \), can be shown to be \( e_2 \) and that the amount of the remaining variation it explains is \( \lambda_2 \) (e.g., Morrison 1976, Johnson and Wichern 1988). Proceeding in this fashion, we see that the \( i \)th PC is given by \( e_i \) and that the amount of variation it accounts for is

\[
\lambda_i / \sum_{k=1}^{n} \lambda_k = \frac{\lambda_i}{\text{tr}(\Sigma_x)}
\]

(A1.18)

**Example 5.** Again consider the additive genetic variance-covariance matrix \( G \) as given in Example 1. Since \( \lambda_1 \approx 5.241, \lambda_2 \approx 0.765 \) and \( \text{tr}(G) = 4 + 2 = 6 \), the first PC explains \( 5.241/6 \approx 0.8735 \) or 87 percent of the variance in \( G \). Note, however, that although the first PC accounts for the majority of variation,
the amount of variation explained by PC1 for any particular variable \( y = b^T x \) depends on the projection of \( b \) onto PC1. For example, if \( b = e_2 \), then the projection of \( b \) onto PC1 has length zero and hence PC1 accounts for no variation of \( y \).

Example 6. Jolicoeur and Mosimann (1960) measured three carapace characters in 24 males of the painted turtle *Chrysemys picta marginata*. Letting \( z_1 \) be carapace length, \( z_2 \) maximum carapace width, and \( z_3 \) carapace height, the resulting sample variance-covariance matrix (\( S_z \), the sample estimate of \( \Sigma_z \)) for these data is

\[
S_z = \begin{pmatrix}
13.77 & 79.15 & 37.13 \\
79.15 & 50.04 & 21.65 \\
37.13 & 21.65 & 13.26
\end{pmatrix}
\]

Hence, \( \text{tr}(S_z) = 13.77 + 50.04 + 13.26 = 200.07 \). The eigenvalues for this matrix are found to be

\[
\lambda_1 = 195.281, \quad \lambda_2 = 3.687, \quad \lambda_3 = 1.103
\]

and the associated normalized eigenvectors are

\[
e_1 = \begin{pmatrix}
0.840 \\
0.492 \\
0.229
\end{pmatrix}, \quad e_2 = \begin{pmatrix}
0.488 \\
-0.870 \\
0.079
\end{pmatrix}, \quad e_3 = \begin{pmatrix}
0.213 \\
0.043 \\
-0.971
\end{pmatrix}
\]

PC1 accounts for 97.6% of the variation (as \( 195.281/200.07 = 0.976 \)), while PC2 and PC3 account for 1.84% and 0.55%, respectively. Jolicoeur and Mosimann interpret PC1 as measuring overall size as the new variable is

\[
y_1 = e_1^T z = 0.840z_1 + 0.492z_2 + 0.229z_3
\]

which corresponds to a simultaneous change in all three variables, as is expected to occur as individuals change their overall size. Likewise PC2 and PC3 are

\[
y_2 = e_2^T z = 0.488z_1 - 0.870z_2 + 0.079z_3 \\
y_3 = e_3^T z = 0.213z_1 + 0.043z_2 - 0.971z_3
\]

which Jolicoeur and Mosimann interpret as measures of shape. Since the coefficient on \( z_3 \) is small relative to the others in PC2, they interpret PC2 as measuring the tradeoff between length (\( z_1 \)) and width (\( z_2 \)). After removing the variation in size, 1.84% of the remaining variation can be accounted for by differences in the shape measured by length versus width. Likewise, since the coefficient in \( z_2 \) is very small in PC3, it measures shape differences due to length (\( z_1 \)) versus height (\( z_3 \)).
This example points out some of the advantages, and possible pitfalls, of using principal components analysis to reduce the data. Essentially all (over 97 percent) of the variance in the three measured characters is accounted for by variation in overall size, with the remaining variation accounted for by differences in shape. While the temptation is strong to simply consider overall size and ignore all shape information, it might be the case that selection is largely ignoring variation in size and instead is focusing on (size-independent) shape differences. In this case, an analysis ignoring shape (such as would occur if only the new character generated by PC1 is considered) would be very misleading. A further complication with principal component analysis is that it can often be very difficult to give biological interpretations to the new characters resulting from the rotation of the coordinate system. This example serves as a brief introduction to the important field of morphometrics, which is concerned with how to quantify and compare the size and shape of organisms. The reader is referred to Pimentel (1979), Reyment et al. (1984), and especially Bookstein et al. (1985), Rohlf and Bookstein (1990), and Reyment (1991) for detailed treatments.

Mahalanobi’s Distance