Appendix 4

The Geometry of Vectors and Matrices:

Eigenvalues and Eigenvectors

An unspeakable horror seized me. There was a darkness; then a dizzy, sickening sensation of sight that was not like seeing; I saw a Line that was no Line; Space that was not space: I was myself, and not myself. When I could find voice, I shrieked loud in agony, Either this is madness or it is Hell. "It is neither," calmly replied the voice of the Sphere, "It is Knowledge." —Flatland (Edwin Abbott Abbott, 1884)

Draft Version 18 November 2008

This appendix reviews 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 one vector into another 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 A4.1A 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.

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

$$||\mathbf{x}|| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} = \sqrt{\mathbf{x}^T \mathbf{x}} \qquad \text{(A4.1a)}$$

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 A4.1B. Similarly (Figure A4.1C), the Euclidean distance between

x and **y** is

$$||\mathbf{x} - \mathbf{y}||^2 = \sum_{i=1}^{n}(x_i - y_i)^2 = (\mathbf{x} - \mathbf{y})^T(\mathbf{x} - \mathbf{y}) = (\mathbf{y} - \mathbf{x})^T(\mathbf{y} - \mathbf{x}) \qquad (A4.1b)$$

Vectors can differ by length, direction, or both. The angle θ between two vectors **x** and **y** provides a measure of how much they differ in direction (Figure A4.1C). 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 ($\theta = 90°$ or $270°$), 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, θ satisfies

$$\cos(\theta) = \frac{\mathbf{x}^T\mathbf{y}}{||\mathbf{x}||\,||\mathbf{y}||} = \frac{\mathbf{y}^T\mathbf{x}}{||\mathbf{x}||\,||\mathbf{y}||} \qquad (A4.2)$$

Note that since $\cos(90°) = \cos(270°) = 0$, two vectors are orthogonal if and only if their inner product is zero, $\mathbf{x}^T\mathbf{y} = 0$.
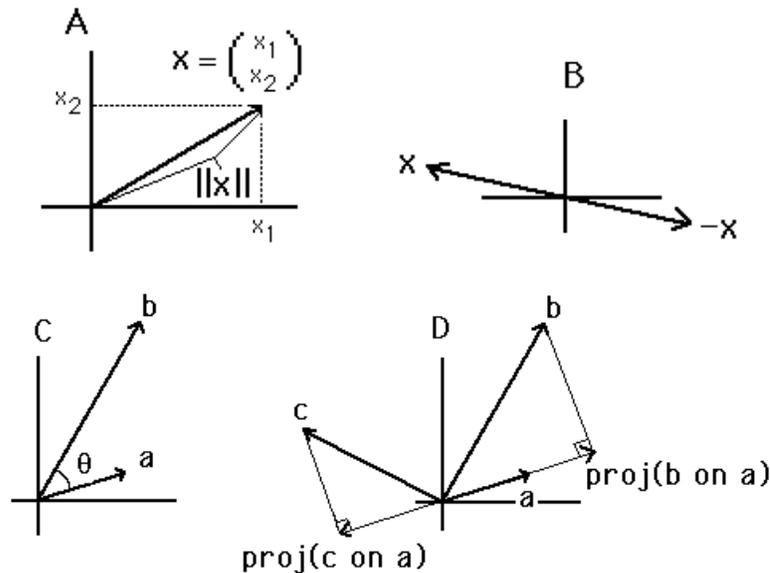

Figure A4.1.   Some basic geometric concepts of vectors. While we use examples from two dimensions, these concepts easily extend to n dimensions. **A**: A vector **x** can be thought of as an arrow from the origin to a point in space whose coordinates are given by the elements of **x**. **B**: Multiplying 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 θ between them. **D**: Proj(**b** on **a**) is the vector resulting from the projection of **b** on **a**. Note that the resulting projection vector is either in the same direction as **a** or in the direction of the reflection of **a**, as seen for Proj(**c** on **a**).

Another way to compare vectors, illustrated in Figure A4.1D, is to consider the **projection** of one vector on another. Specifically, the projection of **x** on **y**, denoted Proj(**x** on **y**),

measures how much of vector $\mathbf{x}$ lines along the direction of vector $\mathbf{y}$. For any two n dimensional vectors, the projection of $\mathbf{x}$ on $\mathbf{y}$ generates a vector defined by

$$\text{Proj}(\mathbf{x} \text{ on } \mathbf{y}) = \frac{\mathbf{x}^T\mathbf{y}}{\mathbf{y}^T\mathbf{y}} \, \mathbf{y} = \frac{\mathbf{x}^T\mathbf{y}}{||\mathbf{y}||^2} \, \mathbf{y} = \left(\cos(\theta) \, \frac{||\mathbf{x}||}{||\mathbf{y}||} \right) \mathbf{y} \qquad \text{(A4.3 a)}$$

Thus $\text{Proj}(\mathbf{x} \text{ on } \mathbf{y})$ a scaled version of the vector we are projecting onto, $\mathbf{y}$. If $||\mathbf{y}|| = 1$, then

$$\text{Proj}(\mathbf{x} \text{ on } \mathbf{y}) = (\mathbf{x}^T\mathbf{y}) \, \mathbf{y} = (\cos(\theta) \, ||\mathbf{x}||) \, \mathbf{y} \qquad \text{(A4.3b)}$$

The vector resulting from the projection of $\mathbf{x}$ on $\mathbf{y}$ is in the same direction as $\mathbf{y}$, unless $90° < \theta < 270°$ in which case $\cos(\theta) < 0$ and the projection vector is in exactly the opposite direction (the reflection of $\mathbf{y}$ about the origin). The length of the projection vector is

$$||\text{Proj}(\mathbf{x} \text{ on } \mathbf{y})|| = |\cos(\theta)| \, ||\mathbf{x}|| \qquad \text{(A4.3c)}$$

If two vectors lie in exactly the same direction, the projection of one on the other just recovers the vector ($\text{Proj}(\mathbf{x} \text{ on } \mathbf{y}) = \mathbf{x}$). 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 $\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_n$ is any set of mutually orthogonal n dimensional vectors, then any n dimensional vector $\mathbf{x}$ can be represented as the sum of projections of $\mathbf{x}$ onto the members of this set,

$$\mathbf{x} = \sum_{i=1}^{n} \text{Proj}(\mathbf{x} \text{ on } \mathbf{y}_i) \qquad \text{(A4.4)}$$

One way to think about such a decomposition is the transformation from one set of axes (or coordinates) into another (defined by the vectors $\mathbf{y}_i$ that **span** the vector space). We can also examine the projection of a vector into some **subspace** of a matrix (say $\mathbf{y}_1, \cdots, \mathbf{y}_k$, where $k < n$), namely the projection onto some subset of the vectors than span the space of the original matrix. This is a key point developed in Chapter 30.

Matrices Describe Vector Transformations

When we multiply a vector $\mathbf{x}$ by a matrix $\mathbf{A}$ to create a new vector $\mathbf{y} = \mathbf{A}\mathbf{x}$, $\mathbf{A}$ rotates and scales the original vector. Thus $\mathbf{A}$ describes a *transformation* of the original coordinate system of $\mathbf{x}$ into a new coordinate system $\mathbf{y}$ (which has a different dimension from $\mathbf{x}$ unless $\mathbf{A}$ is square).

Example A4.1.   Consider the multivariate version of the breeders' equation, $\mathbf{R} = \mathbf{G}\boldsymbol{\beta}$, which holds when certain normality assumptions are satisfied (Chapter 30). Here $\mathbf{R}$ is the resulting change in the vector of phenotypic means, $\mathbf{G}$ the variance-covariance matrix of additive genetic values (breeding values) of these characters, and $\boldsymbol{\beta}$ 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

$$\mathbf{G} = \begin{pmatrix} 4 & -2 \\ -2 & 2 \end{pmatrix} \qquad \boldsymbol{\beta} = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$$

hence

$$\mathbf{R} = \mathbf{G}\boldsymbol{\beta} = \begin{pmatrix} -2 \\ 4 \end{pmatrix}$$

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 inner products, we find $||\boldsymbol{\beta}|| = \sqrt{10}$, $||\mathbf{R}|| = \sqrt{20}$, and $\boldsymbol{\beta}^T \mathbf{R} = 10$. Applying Equation A4.2,

$$\cos \theta = \frac{\boldsymbol{\beta}^T \mathbf{R}}{||\mathbf{R}|| \, ||\boldsymbol{\beta}||} = \frac{1}{\sqrt{2}}$$

Thus the response vector is $\cos^{-1}(1/\sqrt{2}) = 45°$ 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 (Figure A4.2).
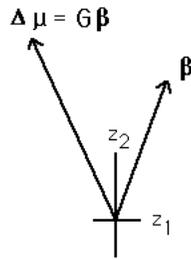

Figure A4.2.   Using the values of $\boldsymbol{\beta}$ and $\mathbf{G}$ from Example A4.1, observe that $\mathbf{G}$ translates the directional selection gradient in a counter-intuitive fashion. Specifically, $\boldsymbol{\beta}$ shows that fitness is maximized by increasing both traits 1 and 2. While the resulting response $\mathbf{R}$ also increases trait 2, it *decreases* trait 1. 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 $\mathbf{G}$ (Figure A4.3).

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 $\mathbf{U}$ as $\mathbf{U} = (\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n)$ where each $\mathbf{u}_i$ is an n dimensional column vector, $\mathbf{U}$ is orthonormal if

$$\mathbf{u}_i{}^T \mathbf{u}_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

In other words, each column of $\mathbf{U}$ is independent from every other column and has unit length. Matrices with this property are also referred to as **unitary** and satisfy

$$\mathbf{U}^T \mathbf{U} = \mathbf{U} \mathbf{U}^T = \mathbf{I} \tag{A4.5a}$$

Hence, the inverse of a unitary matrix is simply its transpose,

$$\mathbf{U}^T = \mathbf{U}^{-1} \tag{A4.5b}$$

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 inner products*. Taking $\mathbf{y}_1 = \mathbf{U}\mathbf{x}_1$ and $\mathbf{y}_2 = \mathbf{U}\mathbf{x}_2$,

$$\mathbf{y}_1{}^T\mathbf{y}_2 = \mathbf{x}_1{}^T(\mathbf{U}^T\mathbf{U})\mathbf{x}_2 = \mathbf{x}_1{}^T\mathbf{x}_2$$

A special case of this is that orthonormal matrices do not change the length of vectors, as $||\mathbf{y}_1|| = \mathbf{y}_1{}^T\mathbf{y}_1 = \mathbf{x}_1{}^T\mathbf{x}_1 = ||\mathbf{x}_1||$. If θ is the angle between vectors $\mathbf{x}_1$ and $\mathbf{x}_2$, then following transformation by an orthonormal matrix,

$$\cos(\theta \,|\, \mathbf{y}_1, \mathbf{y}_2) = \frac{\mathbf{y}_1{}^T\mathbf{y}_2}{\sqrt{||\mathbf{y}_1||\ ||\mathbf{y}_2||}} = \frac{\mathbf{x}_1{}^T\mathbf{x}_2}{\sqrt{||\mathbf{x}_1||\ ||\mathbf{x}_2||}} = \cos(\theta \,|\, \mathbf{x}_1, \mathbf{x}_2)$$

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 described by the eigenvectors (i.e., how the original axes are *rotated*).

Suppose that the vector $\mathbf{y}$ satisfies the matrix equation

$$\mathbf{A}\mathbf{y} = \lambda\mathbf{y} \tag{A4.6}$$

for some scalar value λ and the square matrix $\mathbf{A}$. Geometrically, this means that the new vector resulting from transformation of $\mathbf{y}$ by $\mathbf{A}$ points in the same direction (or is exactly reflected about the origin if $\lambda < 0$) as $\mathbf{y}$. For such vectors, the only action of the matrix transformation is to scale them by some amount λ. 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 $\mathbf{A}$. Vectors satisfying Equation A4.6 are referred to as **eigenvectors** and their associated scaling factors are **eigenvalues**. If $\mathbf{y}$ is an eigenvector, then $a\mathbf{y}$ is also an eigenvector as $\mathbf{A}(a\mathbf{y}) = a(\mathbf{A}\mathbf{y}) = \lambda(a\mathbf{y})$. Note, however, that the associated eigenvalue λ remains unchanged. Hence, we typically scale eigenvectors to be of unit length to give **unit** or **normalized** eigenvectors. In particular, if $\mathbf{u}_i$ is the eigenvector associated with the ith eigenvalue, then the associated normalized eigenvector is $\mathbf{e}_i = \mathbf{u}_i/||\mathbf{u}_i||$.

The eigenvalues of square matrix $\mathbf{A}$ of dimension n are solutions of Equation A4.6, which is usually expressed as the **characteristic equation** $|\mathbf{A} - \lambda\mathbf{I}| = 0$. This can be also be expressed using the **Laplace expansion** as

$$|\mathbf{A} - \lambda\mathbf{I}| = (-\lambda)^n + S_1(-\lambda)^{n-1} + \cdots + S_{n-1}(-\lambda)^1 + S_n = 0 \tag{A4.7}$$

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 LW Chapter 8). Finding the eigenvalues thus requires solving a polynominal equation of order n, implying that there are exactly n eigenvalues (some of which may be identical). 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 $S_n = |\mathbf{A}|$, the determinant of $\mathbf{A}$. S_1 is also related to an

important matrix quantity, the **trace**. This is denoted by $\mathrm{tr}(\mathbf{A})$ and is the sum of the diagonal elements of the matrix,

$$\mathrm{tr}(\mathbf{A}) = \sum_{i=1}^{n} A_{ii}$$

Observe that $S_1 = \mathrm{tr}(\mathbf{A})$ 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,

$$\mathrm{tr}(\mathbf{A}) = \sum_{i=1}^{n} \lambda_i \qquad \text{and} \qquad |\mathbf{A}| = \prod_{i=1}^{n} \lambda_i \tag{A4.8}$$

Hence $\mathbf{A}$ is singular ($|\mathbf{A}| = 0$) if and only if at least one eigenvalue is zero. As we will see, if $\mathbf{A}$ is a covariance matrix, then its trace (the sum of its eigenvalues) is its total amount of variance.

Let $\mathbf{e}_i$ be the (unit-length) eigenvector associated with eigenvalue λ_i. If the eigenvectors of $\mathbf{A}$ can be choosen to be mutually orthogonal, e.g., $\mathbf{e}_i^T \mathbf{e}_j = 0$ for $i \neq j$, then we can express $\mathbf{A}$ as

$$\mathbf{A} = \lambda_1 \mathbf{e}_1 \mathbf{e}_1^T + \lambda_2 \mathbf{e}_2 \mathbf{e}_2^T + \cdots + \lambda_n \mathbf{e}_n \mathbf{e}_n^T \tag{A4.9b}$$

This is the **spectral decomposition** of $\mathbf{A}$. Since $||\mathbf{e}_i|| = 1$, Equation A4.3b gives the projection of $\mathbf{x}$ on $\mathbf{e}_i$ as $(\mathbf{x}^T \mathbf{e}_i)\mathbf{e}_i$. Hence,

$$\begin{aligned} \mathbf{A}\mathbf{x} &= \lambda_1 \mathbf{e}_1 \mathbf{e}_1^T \mathbf{x} + \lambda_2 \mathbf{e}_2 \mathbf{e}_2^T \mathbf{x} + \cdots + \lambda_n \mathbf{e}_n \mathbf{e}_n^T \mathbf{x} \\ &= \lambda_1 \mathrm{Proj}(\mathbf{x} \text{ on } \mathbf{e}_1) + \lambda_2 \mathrm{Proj}(\mathbf{x} \text{ on } \mathbf{e}_2) + \cdots + \lambda_n \mathrm{Proj}(\mathbf{x} \text{ on } \mathbf{e}_n) \end{aligned} \tag{A4.9b}$$

The last equality follows since $\mathbf{e}_i(\mathbf{e}_i^T \mathbf{x}) = (\mathbf{e}_i^T \mathbf{x})\mathbf{e}_i = (\mathbf{x}^T \mathbf{e}_i)\mathbf{e}_i$ as $\mathbf{e}_i^T \mathbf{x}$ is a scalar so that $\mathbf{e}_i^T \mathbf{x} = (\mathbf{e}_i^T \mathbf{x})^T = \mathbf{x}^T \mathbf{e}_i$. Thus, one can view a matrix as a series of vectors that form the projection space, so when a vector is multipled by this matrix, the resulting vector is the sum of projections over all of the vectors (the $\mathbf{e}_i$) than span the space defined by the matrix.

Example A4.2. Determine the **eigenstructure** (the set of eigenvalues and their associated unit eigenvectors) of the genetic variance-covariance matrix $\mathbf{G}$ given in Example A4.1. Writing the characteristic equation,

$$\begin{aligned} |\mathbf{G} - \lambda\mathbf{I}| &= \left| \begin{pmatrix} 4 - \lambda & -2 \\ -2 & 2 - \lambda \end{pmatrix} \right| \\ &= (4 - \lambda)(2 - \lambda) - (-2)^2 = \lambda^2 - 6\lambda + 4 = 0 \end{aligned}$$

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

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

The associated unit eigenvectors are

$$\mathbf{e}_1 \simeq \begin{pmatrix} -0.851 \\ 0.526 \end{pmatrix} \qquad \mathbf{e}_2 \simeq \begin{pmatrix} 0.526 \\ 0.851 \end{pmatrix}$$

These are orthogonal as $\mathbf{e}_1{}^T\mathbf{e}_2 = 0$. From Example A4.1, $||\boldsymbol{\beta}|| = \sqrt{10}$, while $\mathbf{e}_1{}^T\boldsymbol{\beta} \simeq 0.727$ and $\mathbf{e}_2{}^T\boldsymbol{\beta} \simeq 3.079$. Since $||\mathbf{e}_1|| = ||\mathbf{e}_2|| = 1$,

$$\cos(\theta|\mathbf{e}_1,\boldsymbol{\beta}) \simeq \frac{0.727}{\sqrt{10}} \simeq 0.230 \qquad \text{and} \qquad \cos(\theta|\mathbf{e}_2,\boldsymbol{\beta}) \simeq \frac{3.079}{\sqrt{10}} \simeq 0.974$$

giving the angle between $\mathbf{e}_1$ and $\boldsymbol{\beta}$ as $\theta(\mathbf{e}_1,\boldsymbol{\beta}) \simeq 76.7°$, while $\theta(\mathbf{e}_2,\boldsymbol{\beta}) \simeq 13.2°$. The corresponding scaled projections of $\boldsymbol{\beta}$ on these eigenvectors are

$$\lambda_1 \mathrm{Proj}(\boldsymbol{\beta} \text{ on } \mathbf{e}_1) \simeq \begin{pmatrix} -3.236 \\ 2 \end{pmatrix} \qquad \text{and} \qquad \lambda_2 \mathrm{Proj}(\boldsymbol{\beta} \text{ on } \mathbf{e}_2) \simeq \begin{pmatrix} 1.236 \\ 2 \end{pmatrix}$$

From Equation A4.9b,

$$\mathbf{R} = \mathbf{G}\boldsymbol{\beta} = \lambda_1 \mathrm{Proj}(\boldsymbol{\beta} \text{ on } \mathbf{e}_1) + \lambda_2 \mathrm{Proj}(\boldsymbol{\beta} \text{ on } \mathbf{e}_2)$$

As Figure A4.3 shows, the eigenstructure of $\mathbf{G}$ explains the unusual behavior of response seen in Figure A4.2. The eigenvector associated with the **leading eigenvalue** λ_1 accounts for most of the variation inherent in $\mathbf{G}$, and this eigenvector corresponds to a strong negative correlation between the additive genetic values of z_1 and z_2. Hence, even though $\boldsymbol{\beta}$ points in very much the same direction as $\mathbf{e}_2$, because $\lambda_1 >> \lambda_2$ the projection of $\boldsymbol{\beta}$ on $\mathbf{e}_1$ gives a vector of greater length than the projection on $\mathbf{e}_2$, and it is this projection vector that results in the decrease in μ_{z_1}.
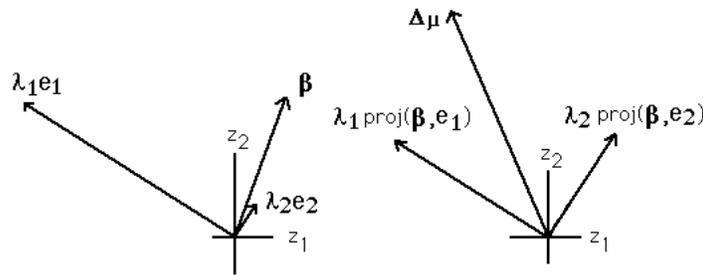

Figure A4.3. **Left**: The scaled eigenvectors associated with the variance-covariance matrix $\mathbf{G}$ from Example A4.1, plotted along with $\boldsymbol{\beta}$. Note that $\mathbf{e}_1$ and $\mathbf{e}_2$ are orthogonal and hence can be thought of as describing a new coordinate system. Since $\lambda_1 >> \lambda_2$, the leading eigenvector $\mathbf{e}_1$ largely dominates the transformation. **Right**: This is shown by taking the projections of $\boldsymbol{\beta}$ on each of these eigenvectors. Even though $\boldsymbol{\beta}$ is nearly parallel to $\mathbf{e}_2$, the projection of $\boldsymbol{\beta}$ on $\mathbf{e}_1$ yields a vector of greater length than the projection of $\boldsymbol{\beta}$ on $\mathbf{e}_2$. From Equation A4.16b, the response to selection $\mathbf{R}$ is the sum of these two projections.

PROPERTIES OF SYMMETRIC MATRICES

Many of the matrices encountered in quantitative genetics are **symmetric**, satisfying $\mathbf{A} = \mathbf{A}^T$. Examples include variance-covariance matrices and the $\boldsymbol{\gamma}$ matrix of quadratic coefficients in the Pearson-Lande-Arnold fitness regression. Here we review 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 $\mathbf{A}$ is symmetric, then if $\mathbf{A}^{-1}$ 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 $\mathbf{e}_i$ for $1 \leq i \leq n$ that satisfies

$$\mathbf{e}_i{}^T \mathbf{e}_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

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

4. A symmetric matrix $\mathbf{A}$ can be **diagonalized** as

$$\mathbf{A} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T \tag{A4.10a}$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix, and $\mathbf{U}$ is an orthonormal matrix ($\mathbf{U}^{-1} = \mathbf{U}^T$). If λ_i and $\mathbf{e}_i$ are the ith eigenvalue and its associated unit eigenvector of $\mathbf{A}$, then

$$\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n) = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & \lambda_n \end{pmatrix} \tag{A4.10b}$$

and

$$\mathbf{U} = (\,\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_n\,) \tag{A4.10c}$$

Geometrically, $\mathbf{U}$ is a unity matrix and thus describes a rigid rotation of the original coordinate system to a new coordinate system given by the eigenvectors of $\mathbf{A}$, while $\boldsymbol{\Lambda}$ is the amount by which unit lengths in the original coordinate system are scaled in the transformed system. Using Equation A4.10a, it is easy to show that

$$\mathbf{A}^{-1} = \mathbf{U}\boldsymbol{\Lambda}^{-1}\mathbf{U}^T \tag{A4.11a}$$

$$\mathbf{A}^{1/2} = \mathbf{U}\boldsymbol{\Lambda}^{1/2}\mathbf{U}^T \tag{A4.11b}$$

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 $\boldsymbol{\Lambda}$ is diagonal, the ith diagonal elements of $\boldsymbol{\Lambda}^{-1}$ and $\boldsymbol{\Lambda}^{1/2}$ are λ_i^{-1} and $\lambda_i^{1/2}$ respectively, implying that if λ_i is an eigenvalue of $\mathbf{A}$, then λ_i^{-1} and $\sqrt{\lambda_i}$ are eigenvalues of $\mathbf{A}^{-1}$ and $\mathbf{A}^{1/2}$. Note that Equations A4.11a and b imply that $\mathbf{A}$, $\mathbf{A}^{-1}$ and $\mathbf{A}^{1/2}$ all have the same eigenvectors. Finally, using Equation A4.10a 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 $\mathbf{A}$,

$$\mathbf{U}^T\mathbf{A}\mathbf{U} = \mathbf{U}^T(\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T)\mathbf{U} = (\mathbf{U}^T\mathbf{U})\boldsymbol{\Lambda}(\mathbf{U}^T\mathbf{U})$$
$$= \boldsymbol{\Lambda} \tag{A4.12}$$

Example A4.3.   Consider the additive genetic variance-covariance matrix $\mathbf{G}$ from Examples A4.1 and A4.2. Recalling the results from Example A4.2 and using Equation A4.10a, we can express $\mathbf{G}$ as $\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T$, where

$$\boldsymbol{\Lambda} = \begin{pmatrix} 5.241 & 0 \\ 0 & 0.765 \end{pmatrix}$$

and

$$\mathbf{U} = (\, \mathbf{e}_1 \quad \mathbf{e}_2 \,) = \left(\begin{pmatrix} -0.851 \\ 0.526 \end{pmatrix} \quad \begin{pmatrix} 0.526 \\ 0.851 \end{pmatrix} \right)$$

From Equation A4.11a, the eigenvalues of $\mathbf{A}^{-1}$ are $(5.241)^{-1} \simeq 0.191$ and $(0.765)^{-1} \simeq 1.307$, while from Equation A4.11b, the eigenvalues of $\mathbf{A}^{1/2}$ are $\sqrt{5.241} \simeq 2.289$ and $\sqrt{0.765} \simeq 0.875$.

---

5. The **Rayleigh-Ritz** theorem gives useful bounds on quadratic products associated with the symmetric matrix $\mathbf{A}$: if the eigenvalues of $\mathbf{A}$ are ordered as $\lambda_{max} = \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n = \lambda_{min}$, then for any vector of constants $\mathbf{c}$,

$$\lambda_1 \, ||\mathbf{c}|| \geq \mathbf{c}^T \mathbf{A} \mathbf{c} \geq \lambda_n \, ||\mathbf{c}|| \tag{A4.13a}$$

Alternatively, if $\mathbf{c}$ is of unit length

$$\max_{||\mathbf{c}||=1} \mathbf{c}^T \mathbf{A} \mathbf{c} = \lambda_1 \tag{A4.13b}$$

$$\min_{||\mathbf{c}||=1} \mathbf{c}^T \mathbf{A} \mathbf{c} = \lambda_n \tag{A4.13c}$$

Where the maximum and minimum occur when $\mathbf{c} = \mathbf{e}_1$ and $\mathbf{c} = \mathbf{e}_n$, the eigenvectors associated with λ_1 and λ_n. This is an especially useful result for bounding variances. Consider a univariate random variable $y = \mathbf{c}^T \mathbf{x}$ formed by a linear combination of the elements of a random vector $\mathbf{x}$. Recall from LW Equation 8.19 that the variance of a sum $\sigma^2(y) = \mathbf{c}^T \boldsymbol{\Sigma}_{\mathbf{X}} \mathbf{c}$, where $\boldsymbol{\Sigma}_{\mathbf{X}}$ is the covariance matrix for $\mathbf{x}$. Applying Equation A4.13a,

$$\lambda_1 ||\mathbf{c}||^2 \geq \sigma^2(y) \geq \lambda_n ||\mathbf{c}||^2 \tag{A4.14}$$

where λ_1 is the largest (leading or **dominant**) eigenvalue and λ_n the smallest eigenvalue of $\boldsymbol{\Sigma}_{\mathbf{X}}$.

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 $\boldsymbol{\Sigma}_{\mathbf{X}}$ is nonsingular and of dimension n. Consider the transformation

$$\mathbf{y} = \mathbf{U}^T \mathbf{x} \tag{A4.15a}$$

where $\mathbf{U} = (\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_n)$ contains the normalized eigenvectors of $\boldsymbol{\Sigma}_{\mathbf{X}}$. Since $\mathbf{U}$ 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 $(e_1, \cdots, e_n)$. The value of y_i is just the lenght of the projection of $\mathbf{x}$ onto $\mathbf{e}_i$ Applying LW Equation 8.21b and Equation A4.12, the variance-covariance matrix for $\mathbf{y}$ is

$$\boldsymbol{\Sigma}_{\mathbf{y}} = \mathbf{U}^T \boldsymbol{\Sigma}_{\mathbf{X}} \mathbf{U} = \boldsymbol{\Lambda} \tag{A4.15b}$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix whose elements are the eigenvalues of $\boldsymbol{\Sigma}_{\mathbf{X}}$, 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 $\mathbf{U}$ thus creates a set of n uncorrelated variables, the ith of which is

$$y_i = \mathbf{e}_i{}^T \mathbf{x} \tag{A4.15c}$$

Since $\mathbf{e}_i$ are defined to be of unit length, from Equation A4.3b we have $y_i = \mathbf{e}_i{}^T \mathbf{x} = \text{Proj}(\mathbf{x} \text{ on } \mathbf{e}_i)$, so that this new variable is the projection of $\mathbf{x}$ onto the ith eigenvector of $\boldsymbol{\Sigma}_{\mathbf{X}}$, implying that the axes of new coordinate system are given by the orthogonal set of eigenvectors of $\boldsymbol{\Sigma}_{\mathbf{X}}$.

Example A4.4. Applying the change of variables suggested by Equation A4.15a to the vector $\mathbf{z}$ of characters with associated $\mathbf{G}$ matrix used in Example A4.1 and using the eigenvalues and vectors obtained in Example A4.2 gives

$$
\mathbf{y} = \mathbf{U}^T \mathbf{z} = \begin{pmatrix} \mathbf{e}_1{}^T \\ \mathbf{e}_2{}^T \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}
$$
$$
= \begin{pmatrix} -0.851 & 0.526 \\ 0.526 & 0.851 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}
$$
$$
= \begin{pmatrix} -0.851 z_1 + 0.526 z_2 \\ 0.526 z_1 + 0.851 z_2 \end{pmatrix}
$$

From Equation A4.15b, $\boldsymbol{\Sigma}_{\mathbf{y}} = \boldsymbol{\Lambda}$ as given in Example A4.3, 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 $y_1 = \mathbf{e}_1{}^T \mathbf{z} = -0.851 z_1 + 0.526 z_2$ and $y_2 = \mathbf{e}_2{}^T \mathbf{z} = 0.526 z_1 + 0.851 z_2$ we can transform the original coordinate system into a new system on which there are no additive genetic correlations between these new characters. Figure A4.4 shows that this transformation is just a rigid rotation of the axes.
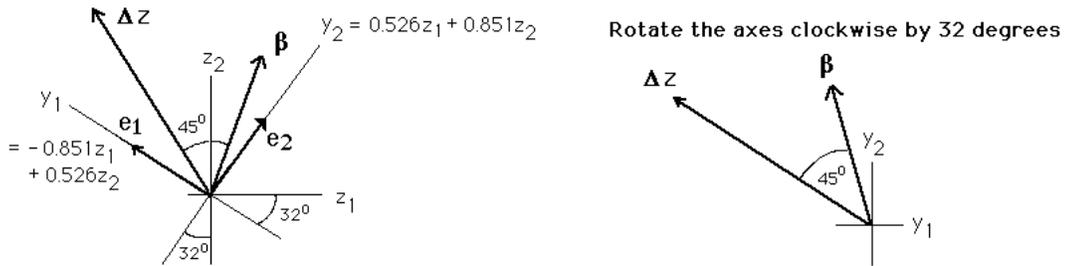

Figure A4.4. The transformation suggested from the diagonalization of $\mathbf{G}$ results in a rigid rotation of axes. The direction of the new axes are given by the eigenvectors $\mathbf{e}_1$ and $\mathbf{e}_2$. The angle between the new axis $\mathbf{e}_1$ and the original z_1 axis is given by taking the angle between $\mathbf{e}_1$ and $\mathbf{z}_1 = (1,0)^T$. Here $\| \mathbf{e}_1 \| = \| \mathbf{z}_1 \| = 1$ and $\mathbf{e}_1{}^T \mathbf{z}_1 = 0.851$, giving $\theta = \cos^{-1}(0.851) \simeq 32°$. The angle between $\mathbf{e}_2$ and the z_2 axis is also 32 degrees.

CANONICAL AXES OF QUADRATIC FORMS

The transformation $\mathbf{y} = \mathbf{U}^T\mathbf{x}$ given by Equation A4.15a 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 $\mathbf{x}^T\mathbf{Ax}$, as rotation of the original axes to align them with the eigenvalues of $\mathbf{A}$ removes all cross-product terms on this new coordinate system. Recall (Equation A4.5b) that $\mathbf{U}$ is a unitary matrix and hence $\mathbf{U}^T = \mathbf{U}^{-1}$. Hence,

$$\mathbf{Uy} = \mathbf{UU}^T\mathbf{x} = \mathbf{x}$$

Applying Equations A4.15a and A4.12 transforms the quadratic form to one in which the square matrix is diagonal,

$$
\begin{aligned}
\mathbf{x}^T\,\mathbf{A}\,\mathbf{x} &= (\mathbf{Uy})^T\mathbf{AUy} = \mathbf{y}^T(\mathbf{U}^T\mathbf{AU})\mathbf{y} \\
&= \mathbf{y}^T\,\mathbf{\Lambda}\,\mathbf{y} \\
&= \sum_{i=1}^{n} \lambda_i y_i^2
\end{aligned}
\tag{A4.16}
$$

where λ_i and $\mathbf{e}_i$ are the eigenvalues and associated (normalized) eigenvectors of $\mathbf{A}$ and $y_i = \mathbf{e}_i^T\mathbf{x}$. The new axes defined by $\mathbf{e}_i$ are the **canonical** (or **principal**) **axes**. Since the $y_i^2 \geq 0$, Equation A4.16 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 $\mathbf{A}$ is **positive definite**. If all eigenvalues are negative (all $\lambda_i < 0$), then $\mathbf{A}$ is **negative definite** as the quadratic form is always negative. If at least one eigenvalue is zero, then $\mathbf{A}$ is at most **semidefinite**, while if $\mathbf{A}$ has both positive and negative eigenvalues it is **indefinite**.

Equations of the form

$$\mathbf{x}^T\mathbf{Ax} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_i x_j = c^2 \tag{A4.17a}$$

arise fairly frequently in quantitative genetics. For example, they describe surfaces of constant variance (tracing out the surface created by vectors $\mathbf{b}$ such that $\mathbf{b}^T\mathbf{x}$ has constant variance c^2, see Figure A4.5) or surfaces of constant fitness in quadratic fitness regressions (those vectors of phenotypic values $\mathbf{z}$ such that $w(\mathbf{z}) = a + (\mathbf{z} - \boldsymbol{\mu})^T\boldsymbol{\gamma}(\mathbf{z} - \boldsymbol{\mu})$ is constant). Solutions to Equation A4.17a describe **quadratic surfaces** — for two dimensions these are the familiar conic sections (ellipses, parabolas, or hyperbolas). Equation A4.16 greatly simplifies the interpretation of these surfaces by removing all cross product terms, giving

$$\mathbf{x}^T\,\mathbf{A}\,\mathbf{x} = \sum_{i=1}^{n} \lambda_i y_i^2 = c^2 \tag{A4.17b}$$

Since $(y_i)^2$ and $(-y_i)^2$ have the same value, the canonical axes of $\mathbf{A}$ are also the **axes of symmetry** for the quadratic surface generated by quadratic forms involving $\mathbf{A}$. When all the eigenvalues of $\mathbf{A}$ are positive (as occurs with nonsingular variance-covariance and other positive definite matrices), Equation A4.17b describes an ellipsoid whose axes of symmetry are given by the eigenvectors of $\mathbf{A}$. The distance from the origin to the surface along the axis given by $\mathbf{e}_i$ is $\lambda_i y_i^2 = c^2$ or $y_i = c\lambda_i^{-1/2}$, as can been seen by setting all the y_k equal to zero except for y_i, giving $\mathbf{x}^T\,\mathbf{A}\,\mathbf{x} = \lambda_i y_i^2 = c^2$. Figure A4.5 shows an example of a two-dimensional constant-variance surface: if we plot the entire set of vectors $\mathbf{b}$ such that the variable $y = \mathbf{b}^T\mathbf{x}$ has variance $c^2 = \mathbf{b}^T\,\mathbf{\Sigma_X}\,\mathbf{b}$, the tips of these vectors sweep out an ellipse.
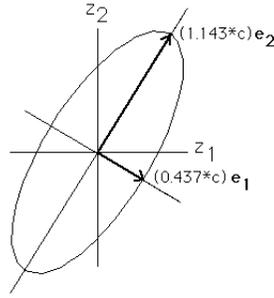
Figure A4.5. The general shape of surfaces of constant variance for the additive genetic variance-covariance matrix $\mathbf{G}$ given in Example A4.1. Defining a new composite character $y = az_1 + bz_2$, the rotated ellipse represents the set of (a, b) values that give y the same additive genetic variance c^2. The major axis of the ellipse is along $\mathbf{e}_2$, the eigenvector associated with the smallest eigenvalue of $\mathbf{G}$, where $\lambda_2 \simeq 0.765$, giving $\lambda_2^{-1/2} \simeq 1.143$. The minor axis of the ellipse is along $\mathbf{e}_1$, the eigenvector associated with the largest eigenvalue of $\mathbf{G}$, where $\lambda_1 \simeq 5.241$, giving $\lambda_1^{-1/2} \simeq 0.437$.
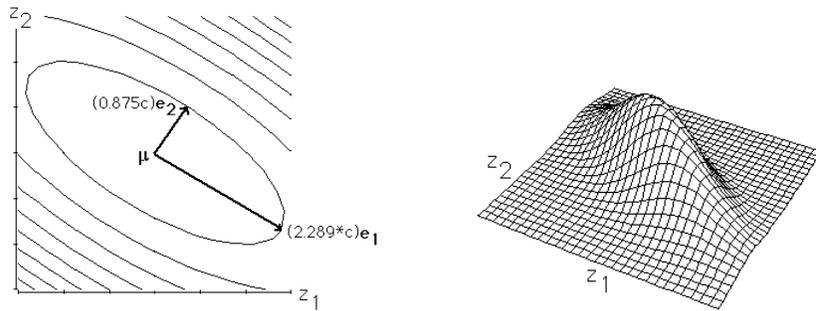

Figure A4.6. **Left**: Surfaces of equal probability assuming that the additive genetic values associated with the characters z_1 and z_2 in Example A4.1 are MVN$(\boldsymbol{\mu}, \mathbf{G})$. These surfaces are ellipses centered at μ, with the major axis of the ellipse along $\mathbf{e}_1$ and the minor axis along $\mathbf{e}_2$. **Right**: A plot of the associated probability density. Slicing along either the major or minor axis gives a normal curve. Since the variance in the major axis is greater, the curve is much broader along this axis. The covariance between the breeding values of z_1 and z_2 rotates the distribution so that the principal axes do not coincide with the (z_1, z_2) axes.

Implications for the Multivariate Normal Distribution

Recall the density of the multivariate normal distribution,

$$\phi(\mathbf{x}) = (2\pi)^{-n/2} \, |\boldsymbol{\Sigma}_{\mathbf{X}}|^{-1/2} \, \exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \, \boldsymbol{\Sigma}_{\mathbf{X}}^{\mathbf{-1}} \, (\mathbf{x} - \boldsymbol{\mu}) \right] \tag{A4.18a}$$

Thus surfaces of equal probability for MVN distributed vectors satisfy

$$(\mathbf{x} - \boldsymbol{\mu})^T \, \boldsymbol{\Sigma}_{\mathbf{X}}^{\mathbf{-1}} \, (\mathbf{x} - \boldsymbol{\mu}) = c^2 \tag{A4.18b}$$

From the discussion following Equation A4.17b, these surfaces are n-dimensional ellipsoids centered at $\boldsymbol{\mu}$ whose axes of symmetry are given by the principal components (the eigenvectors) of $\boldsymbol{\Sigma}_{\mathbf{X}}$. The length of the ellipsoid along the ith axis is $c\sqrt{\lambda_i}$ where λ_i is the eigenvalue

associated with the eigenvector $\mathbf{e}_i$ (Figure A4.6). A final point to note is that Equation A4.18b motivates the **Mahalanobis distance** measure

$$D = \sqrt{(\mathbf{x} - \boldsymbol{\mu})^T \, \boldsymbol{\Sigma}_{\mathbf{X}}^{-1} \, (\mathbf{x} - \boldsymbol{\mu})} \qquad (A4.19)$$

which provides a metric for outliers. As D increases, the probablity of that data vector becomes increasingly unlikely if it were indeed generated from a MVN with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}_{\mathbf{X}}$.

Applying the canonical transformation (Equation A4.15a), we can change coordinate systems by a rigid rotation to remove any correlations between the variables in $\mathbf{x}$. Taking $\mathbf{y} = \mathbf{U}^T(\mathbf{x} - \boldsymbol{\mu})$,

$$\mathbf{y} \sim \text{MVN}(\mathbf{0}, \boldsymbol{\Lambda}) \qquad (A4.20a)$$

where $\boldsymbol{\Lambda}$ and $\mathbf{U}$ are the matrices defined by Equations A4.10b/c for the diagonalization of $\boldsymbol{\Sigma}_{\mathbf{X}}$. In particular,

$$y_i = \mathbf{e}_i{}^T(\mathbf{x} - \boldsymbol{\mu}) \qquad \text{where} \quad y_i \sim \text{N}(0, \lambda_i) \qquad (A4.20b)$$

Note from Equation A4.20a that since the y_i are uncorrelated, they are independent as the joint probability density is the product of n individual univariate normal densities. We can further transform the original vector by taking

$$y_i = \frac{\mathbf{e}_i{}^T(\mathbf{x} - \boldsymbol{\mu})}{\sqrt{\lambda_i}} \qquad \text{giving} \qquad y_i \sim \text{N}(0, 1) \qquad (A4.20c)$$

Thus, the transformation

$$\mathbf{y} = \boldsymbol{\Lambda}^{-1/2}\mathbf{U}^T(\mathbf{x} - \boldsymbol{\mu}) \qquad (A4.20d)$$

implies that $\mathbf{y} \sim \text{MVN}(\mathbf{0}, \mathbf{I})$, the elements of $\mathbf{y}$ being n independent unit normal random variables.

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 $\mathbf{x}$ may account for most of the variance. The procedure of **principal component analysis** extracts these combinations by decomposing the variance of $\mathbf{x}$ into the contributions from 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 $\mathbf{x}$ has been accounted for.

Consider Figure A4.5. Since the set of points comprising the ellipse represents those linear combinations of the random variables of $\mathbf{z}$ 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 $\mathbf{e}_1$, while those furthest away lie along the axis define by $\mathbf{e}_2$. Here $\mathbf{e}_1$ and $\mathbf{e}_2$ are the principal components of $\mathbf{G}$, with the first principal component accounting for most of the variation of $\mathbf{G}$. In particular, the ratio of additive variances for the characters $y_1 = \mathbf{e}_1{}^T\mathbf{z}$ and $y_2 = \mathbf{e}_2{}^T\mathbf{z}$ is $\sigma^2(y_1)/\sigma^2(y_2) = \sigma^2(\mathbf{e}_1{}^T\mathbf{z})/\sigma^2(\mathbf{e}_2{}^T\mathbf{z}) = \mathbf{e}_1{}^T\mathbf{G}\,\mathbf{e}_1/\mathbf{e}_2{}^T\mathbf{G}\,\mathbf{e}_2 = \lambda_1/\lambda_2 \simeq 5.241/0.765 \simeq 6.85$, so that a character in the direction of $\mathbf{e}_1$ has almost seven times as much additive variance as a character lying in the direction of $\mathbf{e}_2$.

In general, suppose we have an n-dimensional variance-covariance matrix $\boldsymbol{\Sigma}_{\mathbf{X}}$. Ordering the eigenvalues of $\boldsymbol{\Sigma}_{\mathbf{X}}$ as $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ then from Equation A4.13b the maximum variance for any linear combination $y = \mathbf{c}_1^T \mathbf{x}$ (subject to the constraint that $||\mathbf{c}_1|| = 1$) is

$$\max \ \sigma^2(y) = \max_{||\mathbf{c}_1||=1} \sigma^2(\mathbf{c}_1^T \mathbf{x}) = \mathbf{c}_1^T \boldsymbol{\Sigma}_{\mathbf{X}} \mathbf{c}_1 = \lambda_1$$

which occurs when $\mathbf{c}_1 = \mathbf{e}_1$. This vector the **first principal component** (often abbreviated as PC1). Excluding PC1, consider how the remaining variance can be explained. The vector $\mathbf{c}_2$ orthogonal to PC1 (e.g., $\mathbf{c}_2^T \mathbf{c}_1 = 0$) that maximizes the remaining variance, e.g., maximizes $\sigma^2(\mathbf{c}_2^T \mathbf{x})$, can be shown to be $\mathbf{e}_2$ and that the amount of the remaining variation it explains is λ_2 (e.g., Morrison 1976, Johnson and Wichern 1988). Proceeding in this fashion, we see that the ith PC is given by $\mathbf{e}_i$ and that the amount of variation it accounts for is

$$\lambda_i \bigg/ \sum_{k=1}^{n} \lambda_k = \frac{\lambda_i}{\text{tr}(\boldsymbol{\Sigma}_{\mathbf{X}})} \tag{A4.21}$$

Hence $\sum \lambda_i = \text{tr}(\boldsymbol{\Sigma}_{\mathbf{X}})$ is the total variance of the vector $\mathbf{x}$, while $\lambda_i/\text{tr}(\boldsymbol{\Sigma}_{\mathbf{X}})$ is the precent of that total variance exampled by the linear combination $\mathbf{x}^T \mathbf{e}_i$.

Example A4.5.   Again consider the additive genetic variance-covariance matrix $\mathbf{G}$ as given in Example A4.1. Since $\lambda_1 \simeq 5.241$, $\lambda_2 \simeq 0.765$ and $\text{tr}(\mathbf{G}) = 4 + 2 = 6$, the first PC explains $5.241/6 \simeq 0.8735$ or 87 percent of the variance in $\mathbf{G}$. Note, however, that although the first PC accounts for the majority of variation, the amount of variation explained by PC1 for any *particular* variable $y = \mathbf{b}^T \mathbf{x}$ depends on the projection of $\mathbf{b}$ onto PC1. For example, if $\mathbf{b} = \mathbf{e}_2$, then the projection of $\mathbf{b}$ onto PC1 has length zero and hence PC1 accounts for no variation of y.

Example A4.6.   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 maximun carapace width, and z_3 carapace height, the resulting sample variance-covariance matrix ($\mathbf{S}_{\mathbf{Z}}$, the sample estimate of $\boldsymbol{\Sigma}_{\mathbf{Z}}$) for these data is

$$\mathbf{S}_{\mathbf{Z}} = \begin{pmatrix} 138.77 & 79.15 & 37.38 \\ 79.15 & 50.04 & 21.65 \\ 37.38 & 21.65 & 11.26 \end{pmatrix}$$

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

$$\lambda_1 = 195.281, \qquad \lambda_2 = 3.687, \qquad \lambda_3 = 1.103$$

and the associated normalized eigenvectors are

$$\mathbf{e}_1 = \begin{pmatrix} 0.840 \\ 0.492 \\ 0.229 \end{pmatrix}, \qquad \mathbf{e}_2 = \begin{pmatrix} 0.488 \\ -0.870 \\ 0.079 \end{pmatrix}, \qquad \mathbf{e}_3 = \begin{pmatrix} 0.213. \\ 0.043 \\ -0.971 \end{pmatrix}$$

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

$$y_1 = \mathbf{e}_1^T \mathbf{z} = 0.840z_1 + 0.492z_2 + 0.229z_3$$

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

$$y_2 = \mathbf{e}_2{}^T\mathbf{z} = 0.488z_1 - 0.870z_2 + 0.079z_3$$
$$y_3 = \mathbf{e}_3{}^T\mathbf{z} = 0.213.z_1 + 0.043z_2 - 0.971z_3$$

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).

Example A4.6 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 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), Elewa (2004), Claude (2008) and especially Bookstein et al. (1985), Rohlf and Bookstein (1990), Reyment (1991), Bookstein (1997), Zelditch et al. (2005) and Slice (2005), for detailed treatments.

TESTING FOR MULTIVARIATE NORMALITY

While multivariate normality is often assumed, it is rarely tested. In LW Chapter 11 we briefly discussed two approaches for testing univariate normality, one graphical and the other based on if the observed skewness and/or kurtosis exceeds that expected for a Gaussian. Both of these can be extended to testing for multivariate normality. Additional methods are reviewed by Gnanadesikan (1977) and Seber (1984).

Graphical Tests: Chi-square Plots

A fairly simple graphical test can be developed by extending the notion of the normal probability plot used to check univariate normality (LW Chapter 11). Recall that in this case the observations are ranked and then plotted against their expected values under normality. Departures from linearity signify departures from normality.

We can apply this same approach to check for multivariate normality. From Equation A4.20d, if $\mathbf{z} \sim \text{MVN}(\boldsymbol{\mu}, \boldsymbol{\Sigma}_\mathbf{z})$, then each element of the vector

$$\mathbf{y} = \boldsymbol{\Lambda}^{-1/2}\mathbf{U}^T(\mathbf{z} - \boldsymbol{\mu})$$

is an independent unit normal ($\mathbf{y} \sim \text{MVN}(\mathbf{0}, \mathbf{I})$). Solving for $\mathbf{z}$ gives

$$(\mathbf{z} - \boldsymbol{\mu}) = \mathbf{U}\boldsymbol{\Lambda}^{1/2}\mathbf{y}$$

Using this and recalling Equation A4.11a,

$$(\mathbf{z} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}_{\mathbf{z}}^{-1} (\mathbf{z} - \boldsymbol{\mu}) = \left(\mathbf{U}\boldsymbol{\Lambda}^{1/2}\mathbf{y}\right)^T \left(\mathbf{U}\boldsymbol{\Lambda}^{-1}\mathbf{U}^T\right)\left(\mathbf{U}\boldsymbol{\Lambda}^{1/2}\mathbf{y}\right)$$

$$= \mathbf{y}^T \boldsymbol{\Lambda}^{1/2} \left(\mathbf{U}^T\mathbf{U}\right) \boldsymbol{\Lambda}^{-1} \left(\mathbf{U}^T\mathbf{U}\right) \boldsymbol{\Lambda}^{1/2}\mathbf{y}$$

$$= \mathbf{y}^T\mathbf{y} = \sum_{i=1}^{n} y_i^2 \qquad\qquad (A4.22)$$

Thus if $\mathbf{z} \sim$ MVN, the quadratic form given by Equation A4.22 is the sum of n independent squared unit normal random variables. By definition, this sum is a χ^2 random variable with n degrees of freedom (e.g., Morrison 1976), suggesting that one test for multivariate normality is to compare the goodness of fit of the scaled distances

$$d_i^2 = (\mathbf{z}_i - \overline{\mathbf{z}})^T \mathbf{S}_{\mathbf{z}}^{-1}(\mathbf{z}_i - \overline{\mathbf{z}}) \qquad\qquad (A4.23)$$

to a χ_n^2. Here $\mathbf{z}_i$ is the vector of observations from the ith individual, $\overline{\mathbf{z}}$ the vector of sample means, and $\mathbf{S}_{\mathbf{z}}^{-1}$ the inverse of the sample variance-covariance matrix. (We use the term distance because $\boldsymbol{\Sigma}_{\mathbf{y}} = \mathbf{I}$, giving the variance of any linear combination $\mathbf{c}^T\mathbf{y}$ as $\mathbf{c}^T\boldsymbol{\Sigma}_{\mathbf{y}}\mathbf{c} = \mathbf{c}^T\mathbf{I}\mathbf{c} = ||\mathbf{c}||^2$. Thus, regardless of orientation, any two $\mathbf{y}$ vectors having the same length also have the same variance, which equals their squared Euclidean distance.) We can order these distances as

$$d_{(1)}^2 \leq d_{(2)}^2 \leq \cdots \leq d_{(m)}^2$$

where m is the number of individuals sampled. Note that $d_{(i)}^2$ is the ith smallest distance, whereas d_i^2 is the distance associated with the vector of observations for the ith individual. Let $\chi_n^2(\alpha)$ correspond to the value of a chi-square random variable x with n degrees of freedom that satisfies $\text{Prob}(x \leq \chi_n^2(\alpha)) = \alpha$. Under multivariate normality, we expect the points

$$\left(d_{(i)}^2, \chi_n^2\left(\frac{i - 1/2}{m}\right) \right)$$

to fall along a line, as the ith ordered distance has i/m observations less than or equal to it (the factor of $1/2$ is added as a correction for continuity). As with normal probability plots, departures from multivariate normality are indicated by departures from linearity.

Example A4.7.   Consider again the data of Jolicoeur and Mosimann (1960) on carapace characters in 24 male turtles. Are the characters z_1 (carapace length) and z_2 (maximun carapace width) jointly bivariate normally distributed? Here $n = 2$ and $m = 24$ and

$$\overline{\mathbf{z}} = \begin{pmatrix} 113.13. \\ 88.29 \end{pmatrix}, \quad \mathbf{S}_{\mathbf{z}} = \begin{pmatrix} 138.77 & 79.15 \\ 79.15 & 50.04 \end{pmatrix}, \quad \mathbf{S}_{\mathbf{z}}^{-1} = \begin{pmatrix} 0.0737 & -0.1165 \\ -0.1165 & 0.2043 \end{pmatrix}$$

where $\mathbf{S}_{\mathbf{z}}$ is the sample covariance matrix. A partial list of the 24 vectors of observations are

$$\mathbf{z}_1 = \begin{pmatrix} 93 \\ 74 \end{pmatrix}, \quad \cdots, \quad \mathbf{z}_{11} = \begin{pmatrix} 113 \\ 88 \end{pmatrix}, \quad \cdots, \quad \mathbf{z}_{24} = \begin{pmatrix} 135 \\ 106 \end{pmatrix}$$

Applying Equation A4.23, these observations translate into the distances

$$d_1^2 = 4.45, \quad \cdots, \quad d_{11}^2 = 0.002, \quad \cdots, \quad d_{24}^2 = 9.277$$

After rank ordering, these correspond to $d^2_{(23)}$, $d^2_{(1)}$, and $d^2_{(24)}$, respectively. For $d^2_{(23)}$, the matching value when distances are chi-squared distributed is

$$\chi^2_2 \left(\frac{23 - 1/2}{24} \right) = \chi^2_2 \, (0.9375.)$$

From chi-square tables, we find $\text{Prob}(\chi^2_2 \leq 5.545) = 0.9375$. so that the data point generated from $\mathbf{z}_1$ is $(4.45, 5.545)$. Likewise, the chi-square values for $d^2_{(1)}$ and $d^2_{(24)}$ are 0.043 and 7.727, respectively. Proceeding similarly for the other values, we obtain the curve plotted in Figure A4.7. This curve departs somewhat from linearity. Further, under the assumption of multivariate normality, the line is expected to pass through the origin, while the best linear fit of these data departs from the origin. Transforming the data by taking logs gives a slightly better fit to a MVN (Figure A4.7).
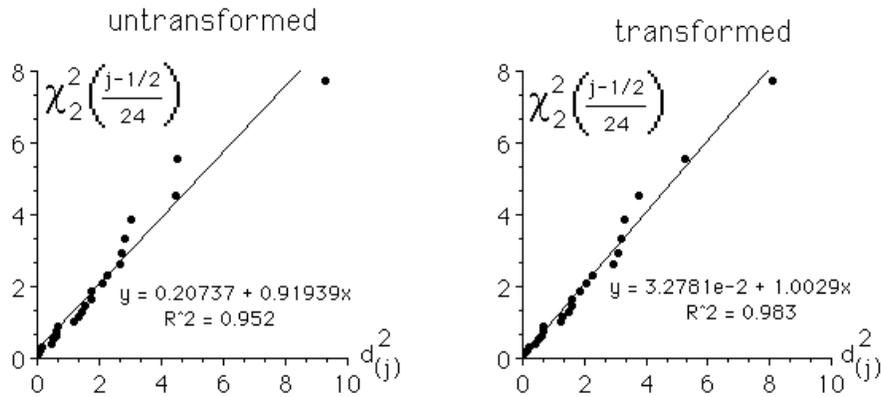

Figure A4.7.   Plots of ranked distance data ($d^2_{(j)}$ being the jth smallest distance) versus the expected corresponding χ^2 value for the data in Example 1. **Left**: The untransformed data does not appear to depart significantly from linearity. **Right**: Log-transforming the data gives a slightly better linear fit ($r^2 = 0.983$ versus $r^2 = 0.952$ for the untransformed data), with the best-fitting line passing through the origin as expected if the distance data follow a χ^2 distribution.

Mardia's Test: Multivariate Skewness and Kurtosis

As was the case for univariate normality, we can test for multivariate normality by examining the skewness and kurtosis of the sample. Mardia (1970) proposed multivariate extensions of skewness and kurtosis and suggested a large sample test based on the asymptotic distribution of these statistics. Let $\mathbf{z}_i$ be the i-th vector of observations, $\bar{\mathbf{z}}$ the vector of sample means, and $\mathbf{S_z}$ sample covariance matrix. If there are m vectors of observations (with each vector measuring n characters), then the multivariate skewness is estimated by

$$b_{1,n} = \frac{1}{m^2} \sum_{i=1}^{m} \sum_{j=1}^{m} \left((\mathbf{z}_i - \bar{\mathbf{z}})^T \, \mathbf{S_z}^{-1} \, (\mathbf{z}_j - \bar{\mathbf{z}}) \right)^3 \tag{A4.24a}$$

while the multivariate kurtosis is estimated by

$$b_{2,n} = \frac{1}{m} \sum_{i=1}^{m} \left((\mathbf{z}_i - \overline{\mathbf{z}})^T \mathbf{S}_{\mathbf{z}}^{-1} (\mathbf{z}_i - \overline{\mathbf{z}}) \right)^2 \tag{A4.24b}$$

If $\mathbf{z} \sim \text{MVN}$, then $b_{1,n}$ and $b_{2,n}$ have expected values 0 and $n(n+2)$. For large m, Mardia (1970) showed that the (scaled) multivariate skewness is asymptotically distributed as a chi-square random variable with f degrees of freedom, viz.,

$$\frac{m}{6} b_{1,n} \sim \chi_f^2, \qquad \text{where } f = \frac{n(n+1)(n+2)}{6} \tag{A4.25a}$$

Likewise for large m, the multivariate kurtosis (following appropriate scaling) is distributed as a unit-normal, viz.,

$$\frac{b_{2,n} - n(n+2)}{\sqrt{8n(n+2)/m}} \sim N(0,1) \tag{A4.25b}$$

If either Equation A4.25a or A4.25b is significant, then multivariate normality is rejected.

Example A4.8.   Again, let us examine the data of Jolicoeur and Mosimann (1960). Do the data considered in Example A4.7 display significant skewness or kurtosis? Here $n = 2$ and $m = 24$. Applying Equations A4.25a/b gives $b_{1,2} = 0.6792$, $b_{2,2} = 7.6043$. Considering skewness first,

$$\frac{m}{6} b_{1,2} = \frac{24}{6} 0.6792 = 2.717$$

is approximately chi-square distributed with $f = 2(2+1)(2+2)/6 = 4$ degrees of freedom. Since $\text{Prob}(\chi_4^2 \geq 2.717) \simeq 0.606$, this is not significant. Turning to kurtosis, Equation A4.25b gives

$$\frac{b_{2,n} - n(n+2)}{\sqrt{8n(n+2)/m}} = \frac{7.6043 - 8}{1.633} \simeq -0.2423$$

which is also not significant as $\text{Prob}(|N(0,1)| \geq 0.2423) \simeq 0.81$. Transforming the data by taking logs gives $b_{1,2} = 0.2767$ and $b_{2,2} = 7.1501$, improving the departure from skewness but increasing the departure from kurtosis. Applying Equations A4.25a/b gives 1.068 and -0.5204, again these are not significant. Reyment (1971) gives a number of other biological examples using Mardia's test.

Literature Cited

Bookstein, F., B. 1997. }si Morphometric Tools for Landmark Data: Geometry and Biology. Cambridge University Press. [A4]

Bookstein, F., B. Chernoff, R. Elder, J. Humphries, G. Smith, and R. Strauss. 1985. *Morphometrics in evolutionary biology*. Special publication 15, The Academy of Natural Sciences of Philadelphia. [A4]

Claude, J. 2008. *Morphometrics with R* . Springer [A4]

Dhrymes, P. J. 1978. *Mathematics for econometrics*. Springer-Verlag, New York. [A4]

Elewa, A. M. T. 2004. *Morphometrics: Applications in Biology and Paleontology.* Springer [A4]

Franklin, J. N. 1968. *Matrix theory*. Prentice-Hall, New Jersey. [A4]

Gantmacher, F. R. 1960. *The theory of matrices* (two volumes). Chelsea, New York. [A4]

Gnanadesikan, R. 1977. *Methods for statistical data analysis of multivariate observations*. John Wiley and Sons, New York. [A4]

Horn, R. A. and C. R. Johnson. 1985. *Matrix analysis*. Cambridge University Press, Cambridge. [A4]

Johnson, R. A. and D. W. Wichern. 1988. *Applied multivariate statistical analysis, 2nd ed.* Prentice-Hall, New Jersey. [A4]

Jolicoeur, P. and J. E. Mosimann. 1960. Size and shape variation in the painted turtle. A principal component analysis. *Growth* 24: 339–354. [A4]

Mardia, K. V. 1970. Measures of multivariate skewness and kurtosis with applications. *Biometrika* 57: 519–530. [A4]

Morrison, D. F. 1976. *Multivariate statistical methods*. McGraw-Hill, New York. [A4]

Morrison, D. F. 1983. *Applied linear statistical methods*. Prentice-Hall, New Jersey. [A4]

Pimentel, R. A. 1979. *Morphometrics. The multivariate analysis of biological data*. Kendall and Hunt, Dubuque, Iowa. [A4]

Reyment, R.A. 1971. Multivariate normality in morphometric analysis. *Mathematical Geology* 3: 357–368. [A4]

Reyment, R.A. 1991. *Multidimensional palaeobiology*. Pergamon Press, Oxford. [A4]

Reyment, R.A. , R.E. Blackith, and N. A. Campbell. 1984. *Multivariate morphometrics*. Academic Press, New York. [A4]

Rohlf, F. J. and F. L. Bookstein. 1990. *Proceeding of the Michigan morphometrics workshop*. Special Publication No. 2, The University of Michigan Museum of Zoology, Ann Arbor, Michigan. [A4]

Searle, S. S. 1982. *Matrix algebra useful for statistics*. John WIley and Sons, New York. [A4]

Seber, G. A. F. 1984. *Multivariate observations*. John WiIley and Sons, New York. [A4]

Slice, D. E. 2005. *Modern Morphometrics in Physical Anthropology.* Springer [A4]

Wilf, H. S. 1978. *Mathematics for the physical sciences*. Dover, New York. [A4]

Zelditch, M., D. Swiderski, D. H. Sheets, and W. Fink. 2005. *Geometric Morphometrics for Biologists*. Academic Press. [A4]