Appendix 3

Further Topics in Matrix Algebra and Linear Models

Version 7 Jan 2021

This appendix builds on Chapters 9 and 10, presenting additional results from matrix algebra and linear model theory. We start by introducing two useful matrix operations, generalized inverses (for solving singular systems of equations) and the square root of a matrix (for obtaining a set of uncorrelated variables). These results are then used for a formal derivation of several properties of generalized least-squares (GLS) estimators. We next examine how linear model sums of squares can be written as quadratic forms and how these sums of squares are used in formal hypothesis testing and in the construction of confidence intervals. We then turn to equivalent linear mixed models (which allow calculations for one model to be performed on a potentially much simpler model), and the conclude with brief introductions to matrix derivatives and matrix decompositions.

GENERALIZED INVERSES AND SOLUTIONS TO SINGULAR SYSTEMS OF EQUATIONS

Linear systems of equations are ubiquitous in quantitative genetics and we have presented solutions for such systems by assuming that the appropriate matrices are nonsingular, and hence can be inverted. However, in the real world of large, complex, and unbalanced data sets, the existence of an inverse is by no means guaranteed. Consider the solution of the matrix equation $\mathbf{y} = \mathbf{A}\mathbf{x}$ for the unknown vector $\mathbf{x}$. If $\mathbf{A}$ is square and nonsingular, then $\mathbf{x} = \mathbf{A}^{-1}\mathbf{y}$ is the unique solution. However, what happens if $\mathbf{A}$ is singular or is nonsquare? In this case either: (i) the system has no solution and is said to be **inconsistent**, or (ii) there are an infinite number of solution (the solution set is a line, plane, or hyperplane). An example of an inconsistent system is

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

which cannot be satisfied by any $(x_1,\, x_2)$. Likewise, a system with an infinite number of solutions is

$$x_1 + x_2 = 1$$
$$2\,x_1 + 2\,x_2 = 2$$

which has a line of solutions of the form $x_2 = 1 - x_1$ for arbitrary x_1. While these two simple systems can be solved by inspection, a more systematic approach is required for arbitary systems. This is provided by using **generalized inverses**.

Generalized Inverses

Suppose a matrix $\mathbf{A}^-$ exists such that

$$\mathbf{A}\mathbf{A}^-\mathbf{A} = \mathbf{A} \tag{A3.1}$$

where $\mathbf{A}$ is $p \times q$ and $\mathbf{A}^-$ is $q \times p$ (note that $\mathbf{A}$ need not be square). Premultiplying both sides of the equation $\mathbf{A}\mathbf{x} = \mathbf{y}$ by $\mathbf{A}\mathbf{A}^-$ gives

$$\mathbf{A}\mathbf{A}^-\mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{A}^-\mathbf{y}$$

and hence

$$\mathbf{A}(\mathbf{x} - \mathbf{A}^{-}\mathbf{y}) = \mathbf{0}$$

implying that, if the system is consistent, a solution is

$$\mathbf{x} = \mathbf{A}^{-}\mathbf{y} \tag{A3.2}$$

Given the analogy with the inverse of a nonsingular square matrix, a matrix $\mathbf{A}^{-}$ satisfying Equation A3.1 is called a generalized inverse (**g-inverse**, **conditional inverse**, **pseudo-inverse**) of $\mathbf{A}$. While the notation of using $\mathbf{A}^{-}$ to denote the generalized inverse of $\mathbf{A}$ is widely used in the quantitative genetics, the notation $\mathbf{A}^{+}$ is also used in other literatures.

Unless $\mathbf{A}$ is nonsingular, Equation A3.1 does not define a unique matrix, so we refer to $\mathbf{A}^{-}$ as *a* generalized inverse instead of *the* generalized inverse. A unique generalized inverse, the **Moore-Penrose inverse**, can be obtained by imposing three additional conditions: (i) $\mathbf{A}^{-}\mathbf{A}\mathbf{A}^{-} = \mathbf{A}^{-}$; (ii) $(\mathbf{A}\mathbf{A}^{-})^{T} = \mathbf{A}\mathbf{A}^{-}$, and (iii) $(\mathbf{A}^{-}\mathbf{A})^{T} = \mathbf{A}^{-}\mathbf{A}$. However, for our purposes any $\mathbf{A}^{-}$ satisfying Equation A3.1 is sufficient. Equation A3.38 shows how generalized inverses can be computed using the singular value decomposition (Equation A3.36a), with other approaches given by Henderson (1984a). More detailed treatment of the properties of generalized inverses are given by Pringle and Rayner (1971), Rao and Mitra (1971), Dhrymes (1978), and Searle (1982), and we summarize some of these results below.

Consistency and Solutions to Consistent Systems

When dealing with linear models for complex designs, it is not immediately clear if the resulting OLS/GLS equations have solutions. Generalized inverses provide a check of consistency, and hence of whether a system of equations has any solutions. A linear system $\mathbf{A}_{m\times q}\,\mathbf{x}_{q\times 1} = \mathbf{y}_{m\times 1}$ is **consistent** if and only if

$$\mathbf{A}\mathbf{A}^{-}\mathbf{y} = \mathbf{y} \tag{A3.3}$$

Given a consistent system, all solutions have the form

$$\mathbf{x} = \mathbf{A}^{-}\mathbf{y} + (\mathbf{I} - \mathbf{A}^{-}\mathbf{A})\mathbf{c} \tag{A3.4}$$

where $\mathbf{c}$ is an arbitrary $q \times 1$ column vector. For example, taking $\mathbf{c} = \mathbf{0}$ recovers Equation A3.2, while if $\mathbf{A}^{-1}$ exists, then $\mathbf{I} - \mathbf{A}^{-1}\mathbf{A} = \mathbf{0}$ and the solution $\mathbf{x} = \mathbf{A}^{-1}\mathbf{y}$ is unique. To see that any expression of the form of Equation A3.4 is a solution, note that

$$\begin{aligned}\mathbf{A}\mathbf{x} &= \mathbf{A}(\mathbf{A}^{-}\mathbf{y} + [\mathbf{I} - \mathbf{A}^{-}\mathbf{A}]\mathbf{c}) \\ &= \mathbf{A}\mathbf{A}^{-}\mathbf{y} + (\mathbf{A} - \mathbf{A}\mathbf{A}^{-}\mathbf{A})\mathbf{c} = \mathbf{y} + (\mathbf{A} - \mathbf{A})\mathbf{c} \\ &= \mathbf{y}\end{aligned}$$

which follows from Equations A3.3 and A3.1, respectively.

Example A3.1. Consider the following system of equations

$$\begin{aligned}x_1 + 2x_2 + 3x_3 &= 5 \\ 2x_1 + x_2 + 2x_3 &= 6\end{aligned}$$

which can be written in matrix form as $\mathbf{A}\mathbf{x} = \mathbf{y}$, with

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 2 \end{pmatrix}, \qquad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \qquad \mathbf{y} = \begin{pmatrix} 5 \\ 6 \end{pmatrix}$$

The matrix

$$\mathbf{A}^- = \begin{pmatrix} -11/26 & 9/13 \\ 4/13 & -3/13 \\ 7/26 & -1/13 \end{pmatrix}$$

satisfies $\mathbf{A}\mathbf{A}^-\mathbf{A} = \mathbf{A}$ and thus is a generalized inverse of $\mathbf{A}$. Matrix multiplication shows that $\mathbf{A}\mathbf{A}^- = \mathbf{I}$, implying $\mathbf{A}\mathbf{A}^-\mathbf{y} = \mathbf{y}$. Thus, Equation A3.3 is satisfied and this system of equations is consistent for any $\mathbf{y}$. One solution is $\mathbf{x} = \mathbf{A}^-\mathbf{y}$, or

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} -11/26 & 9/13 \\ 4/13 & -3/13 \\ 7/26 & -1/13 \end{pmatrix} \begin{pmatrix} 5 \\ 6 \end{pmatrix} = \frac{1}{26} \begin{pmatrix} 53 \\ 4 \\ 23 \end{pmatrix}$$

More generally, since

$$\mathbf{I} - \mathbf{A}^-\mathbf{A} = \begin{pmatrix} 1/26 & 2/13 & -3/26 \\ 2/13 & 8/13 & -6/13 \\ -3/26 & -6/13 & 9/26 \end{pmatrix}$$

then from Equation A3.4, any solution to this system of equations has the form

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \frac{1}{26} \begin{pmatrix} 53 \\ 4 \\ 23 \end{pmatrix} + \begin{pmatrix} 1/26 & 2/13 & -3/26 \\ 2/13 & 8/13 & -6/13 \\ -3/26 & -6/13 & 9/26 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix}$$

which reduces to

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \frac{1}{26} \begin{pmatrix} 53 \\ 4 \\ 23 \end{pmatrix} + c \cdot \begin{pmatrix} 1 \\ 4 \\ -3 \end{pmatrix}$$

where c is an arbitrary constant.

Although an infinite number of solutions exists when $\mathbf{A}$ is singular, particular linear combinations (or **contrasts**) of the elements of $\mathbf{x}$ may have *unique values*. For example, consider the system $x_1 + x_2 = 1$. Here there are an infinite number of solutions for (x_1, x_2), but only a single solution, 1, for the contrast $x_1 + x_2$.

Consider some linear combination $\mathbf{b}^T\mathbf{x} = \sum b_i x_i$. If the vector of constants $\mathbf{b}$ satisfies

$$\mathbf{b}^T\mathbf{A}^-\mathbf{A} = \mathbf{b}^T \tag{A3.5a}$$

then $\mathbf{b}^T\mathbf{x}$ has a unique solution given by

$$\mathbf{b}^T\mathbf{x} = \mathbf{b}^T\mathbf{A}^-\mathbf{y} \tag{A3.5b}$$

To see this, note that Equation A3.4 gives the general solution as

$$\begin{aligned} \mathbf{b}^T\mathbf{x} &= \mathbf{b}^T(\mathbf{A}^-\mathbf{y} + [\mathbf{I} - \mathbf{A}^-\mathbf{A}]\mathbf{c}) \\ &= \mathbf{b}^T\mathbf{A}^-\mathbf{y} + (\mathbf{b}^T\mathbf{I} - \mathbf{b}^T\mathbf{A}^-\mathbf{A})\mathbf{c} \\ &= \mathbf{b}^T\mathbf{A}^-\mathbf{y} + (\mathbf{b}^T - \mathbf{b}^T)\mathbf{c} \\ &= \mathbf{b}^T\mathbf{A}^-\mathbf{y} \end{aligned}$$

which is independent of the arbitrary vector $\mathbf{c}$. Likewise, a matrix of contrasts, $\mathbf{B}\mathbf{x}$, has a unique solution $\mathbf{B}\mathbf{A}^-\mathbf{y}$, provided $\mathbf{B}$ satisfies $\mathbf{B}\mathbf{A}^-\mathbf{A} = \mathbf{B}$

Example A3.2. Consider the system of equations from Example A3.1. Is there a unique solution for the two linear contrasts $c_1 = x_2 - 4x_1$ and $c_2 = x_3 + 3x_1$? In matrix form,

$$\begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} x_2 - 4x_1 \\ x_3 + 3x_1 \end{pmatrix} = \mathbf{Bx}$$

where

$$\mathbf{B} = \begin{pmatrix} -4 & 1 & 0 \\ 3 & 0 & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

Using the generalized inverse for **A** from Example A3.1, matrix multiplication shows that

$$\mathbf{BA^-A} = \begin{pmatrix} -4 & 1 & 0 \\ 3 & 0 & 1 \end{pmatrix} = \mathbf{B}$$

Hence, the matrix version of Equation A3.5b gives the unique solution for this vector of contrasts as

$$\begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \mathbf{BA^-y} = \begin{pmatrix} -4 & 1 & 0 \\ 3 & 0 & 1 \end{pmatrix} \begin{pmatrix} -11/26 & 9/13 \\ 4/13 & -3/13 \\ 7/26 & -1/13 \end{pmatrix} \begin{pmatrix} 5 \\ 6 \end{pmatrix} = \begin{pmatrix} -8 \\ 7 \end{pmatrix}$$

To see that this solution is indeed unique, note that we can rearrange the contrast equations to obtain $x_2 = c_1 + 4x_1$ and $x_3 = c_2 - 3x_1$. Substituting into the original set of equations (Example A3.1),

$$x_1 + 2x_2 + 3x_3 = x_1 + 2(c_1 + 4x_1) + 3(c_2 - 3x_1) = 2c_1 + 3c_2 = 5$$
$$2x_1 + x_2 + 2x_3 = 2x_1 + (c_1 + 4x_1) + 2(c_2 - 3x_1) = c_1 + 2c_2 = 6$$

so that the original set of two equations and three unknowns reduces to a two equation-two unknown system. In matrix form this is

$$\begin{pmatrix} 2 & 3 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} 5 \\ 6 \end{pmatrix}$$

Since the coefficient matrix is invertible (its determinate is $2 \cdot 2 - 3 \cdot 1 \neq 0$), there is a unique solution for this pair of contrasts ($c_1 = -8$ and $c_2 = 7$).

Estimability of Fixed Factors

The above results have implications for the estimation of fixed factors in the general linear model, $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$. Recall that the OLS solution for a vector $\boldsymbol{\beta}$ of fixed effects is $\widehat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$ (Equation 10.9a). If the design matrix **X** has **full column rank** (all columns of **X** are independent), $(\mathbf{X}^T\mathbf{X})^{-1}$ exists and the OLS solution for $\boldsymbol{\beta}$ is unique. However, when $(\mathbf{X}^T\mathbf{X})$ is singular (and hence does not have a unique inverse), it is not possible to obtain unique OLS estimates for all the fixed factors in a model. For example, suppose β_1 indicates a sex effect (male vs. female) and β_2 indicates the effect of a particular diet. If the design is such that no females used this diet while all males did, we do not have separate information on both sex and diet effects and hence can only estimate $\beta_1 + \beta_2$ rather than being able to estimate both β_1 and β_2 separately.

A linear combination of factors $\mathbf{b}^T\boldsymbol{\beta}$ is said to be **estimable** for a given design matrix **X** if there exists some column vector **a** that satisfies

$$E[\mathbf{a}^T\mathbf{y}] = \mathbf{b}^T\boldsymbol{\beta} \tag{A3.6a}$$

Estimability thus implies that there is some linear combination $\mathbf{a}^T\mathbf{y}$ of the original data whose expected value equals the desired linear combination of factors. Since $E[\mathbf{y}] = \mathbf{X}\boldsymbol{\beta}$, this definition implies that $\mathbf{b}^T\boldsymbol{\beta}$ is estimable if there exists a column vector $\mathbf{a}$ that satisfies $E[\mathbf{a}^T\mathbf{y}] = \mathbf{a}^T E[\mathbf{y}] = \mathbf{a}^T\mathbf{X}\boldsymbol{\beta} = \mathbf{b}^T\boldsymbol{\beta}$, implying $(\mathbf{a}^T\mathbf{X} - \mathbf{b}^T)\boldsymbol{\beta} = \mathbf{0}$, or that

$$\mathbf{X}^T\mathbf{a} = \mathbf{b} \tag{A3.6b}$$

An alternative (and equivalent) condition is that $\mathbf{b}$ satisfies

$$\mathbf{b}^T(\mathbf{X}^T\mathbf{X})^-(\mathbf{X}^T\mathbf{X}) = \mathbf{b}^T \tag{A3.6c}$$

Henderson (1984a) gives other equivalent conditions. Equation A3.6c implies that if $\mathbf{X}^T\mathbf{X}$ is nonsingular, all linear combinations of $\boldsymbol{\beta}$ are estimable. Note that Equation A3.6c is identical to the condition given by Equation A3.5a (taking $\mathbf{A} = \mathbf{X}^T\mathbf{X}$), implying that these solutions are also unique estimates. If estimable, the OLS solution of the vector $\mathbf{b}^T\boldsymbol{\beta}$ given by

$$\text{OLS}(\mathbf{b}^T\boldsymbol{\beta}) = \mathbf{b}^T\left(\mathbf{X}^T\mathbf{X}\right)^-\mathbf{X}^T\mathbf{y} \tag{A3.6d}$$

is unique and is independent of which generalized inverse is actually used.

Example A3.3. Consider the linear model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta}$, where

$$\boldsymbol{\beta} = \begin{pmatrix}\beta_1\\ \beta_2\\ \beta_3\end{pmatrix} \quad \text{and} \quad \mathbf{X} = \begin{pmatrix}1 & 1 & 0\\ 1 & 1 & 0\\ 0 & 0 & 1\end{pmatrix}, \quad \text{giving} \quad \mathbf{X}^T\mathbf{X} = \begin{pmatrix}2 & 2 & 0\\ 2 & 2 & 0\\ 0 & 0 & 1\end{pmatrix}$$

Note that $\mathbf{X}^T\mathbf{X}$ is singular (it has rank 2 as its first two columns are identical), so we cannot obtain unique estimates of all three parameters. For this design matrix, are β_3, $\beta_1 + \beta_2$, and β_1 all estimable? These three combinations correspond to vectors of $\mathbf{b}^T = (0, 0, 1)$, $(1, 1, 0)$, and $(1, 0, 0)$, respectively. For the first two $\mathbf{b}$ vectors, we can find a vector $\mathbf{a}$ that satisfies $\mathbf{X}^T\mathbf{a} = \mathbf{b}$, namely,

$$\mathbf{X}^T\begin{pmatrix}0\\ 0\\ 1\end{pmatrix} = \begin{pmatrix}0\\ 0\\ 1\end{pmatrix} \qquad \text{and} \qquad \mathbf{X}^T\begin{pmatrix}1/2\\ 1/2\\ 0\end{pmatrix} = \begin{pmatrix}1\\ 1\\ 0\end{pmatrix}$$

so that, from Equation A3.6b, these two linear combinations, β_3 and $(\beta_1 + \beta_2)$, are estimable. However, because

$$\mathbf{X}^T\begin{pmatrix}a_1\\ a_2\\ a_3\end{pmatrix} = \begin{pmatrix}a_1 + a_2\\ a_1 + a_2\\ a_3\end{pmatrix} \neq \begin{pmatrix}1\\ 0\\ 0\end{pmatrix}$$

β_1 is not estimable as $a_1 + a_2$ cannot simultaneously equal zero and one, and hence there exists no vector $\mathbf{a}$ that satisfies $\mathbf{X}^T\mathbf{a} = \mathbf{b}$ for this particular combinations of $\mathbf{X}$ and $\mathbf{b}$.

THE SQUARE ROOT OF A MATRIX

The concept of the square root of a symmetric nonsingular matrix provides another useful matrix tool for the analysis of linear models. In particular, using the square root of the covariance matrix transforms a vector of correlated variables into a new vector of variables

with covariance matrix $\mathbf{I}$, implying that the transformed variables are uncorrelated with unit variance.

Consider a symmetric nonsingular matrix $\mathbf{V}$ and define $\mathbf{V}^{1/2}$ as the matrix satisfying

$$\mathbf{V}^{1/2}\mathbf{V}^{1/2} = \mathbf{V} \tag{A3.7a}$$

In effect, $\mathbf{V}^{1/2}$ is the square root of a matrix, in that, when squared, we recover $\mathbf{V}$. Denoting the inverse of $\mathbf{V}^{1/2}$ as $\mathbf{V}^{-1/2}$, we also have the following properties

$$\mathbf{V}^{-1/2}\mathbf{V}^{1/2} = \mathbf{I}, \quad \mathbf{V}^{-1/2}\mathbf{V}^{-1/2} = \mathbf{V}^{-1}, \quad \text{and} \quad \mathbf{V}^{-1/2}\mathbf{V} = \mathbf{V}^{1/2} \tag{A3.7b}$$

Likewise, both $\mathbf{V}^{1/2}$ and its inverse are symmetric. As shown below, the diagonalization decomposition of $\mathbf{V}$ leads to expressions for $\mathbf{V}^{1/2}$ (Equation A3.33b) and $\mathbf{V}^{-1/2}$ (Equation A3.33c). Finally, the square root matrix of $\mathbf{I}$ is simply $\mathbf{I}$, with

$$\mathbf{I}^{-1/2}\mathbf{I}^{1/2} = \mathbf{I}, \quad \mathbf{I}^{-1/2}\mathbf{I}^{-1/2} = \mathbf{I}, \quad \text{and} \quad \mathbf{I}^{-1/2}\mathbf{I} = \mathbf{I} \tag{A3.7c}$$

Suppose the random vector $\mathbf{y}$ has covariance matrix $\mathbf{V}$ and consider the new vector $\mathbf{z} = \mathbf{V}^{-1/2}\mathbf{y}$. Recalling Equation 9.21b, the resulting covariance matrix for $\mathbf{z}$ becomes

$$\text{Var}(\mathbf{z}) = \mathbf{V}^{-1/2}\text{Var}(\mathbf{y})\mathbf{V}^{-1/2} = \mathbf{V}^{-1/2}\mathbf{V}\mathbf{V}^{-1/2} = \mathbf{V}^{-1/2}\mathbf{V}^{1/2} = \mathbf{I} \tag{A3.8a}$$

Thus, the transformed variables have unit variance and are uncorrelated. Suppose $\mathbf{y}$ is an $n \times 1$ column vector with $\mathbf{y} \sim \text{MVN}(\boldsymbol{\mu}, \mathbf{V})$. It follows that

$$\mathbf{z} = \mathbf{V}^{-1/2}(\mathbf{y} - \boldsymbol{\mu}) \sim \text{MVN}(\mathbf{0}, \mathbf{I}) \tag{A3.8b}$$

so that $z_i \sim \text{N}(0, 1)$, and hence the transformed variables are independent unit normals. Thus,

$$\begin{aligned}(\mathbf{y} - \boldsymbol{\mu})^T\mathbf{V}^{-1}(\mathbf{y} - \boldsymbol{\mu}) &= (\mathbf{y} - \boldsymbol{\mu})^T\mathbf{V}^{-1/2}\mathbf{V}^{-1/2}(\mathbf{y} - \boldsymbol{\mu}) \\ &= \mathbf{z}^T\mathbf{z} \\ &= \sum_{i=1}^{n} z_i^2 \sim \chi_n^2\end{aligned} \tag{A3.9}$$

The last step follows by recalling that the sum of n squared (and independent) unit normal random variables follows a χ^2 distribution with n degrees of freedom (Appendix 5). Thus when $\mathbf{y}$ is multivariate normal, the quadratic form $(\mathbf{y} - \boldsymbol{\mu})^T\mathbf{V}^{-1}(\mathbf{y} - \boldsymbol{\mu})$ follows a χ^2 distribution. As we will see shortly, Equation A3.9 is the basis for goodness-of-fit tests of linear models. We will return to other matrix decompositions that can be used to remove correlations at the end of the appendix, where we also examine the relationship between the eigenstructures of $\mathbf{V}$, $\mathbf{V}^{1/2}$, and $\mathbf{V}^{-1/2}$.

DERIVATION OF THE GLS ESTIMATORS

One important application of the square root of a matrix is that it allows us to obtain generalized least-squares (GLS) estimators from ordinary least-squares (OLS) estimators. Suppose the linear model is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \qquad \text{with } \mathbf{e} \sim (0, \sigma_e^2\,\mathbf{R})$$

Premultiplying both sides by $\mathbf{R}^{-1/2}$ gives

$$\mathbf{z} = \mathbf{Z}\boldsymbol{\beta} + \mathbf{f} \qquad \text{with } \mathbf{f} \sim (0, \sigma_e^2\,\mathbf{I})$$

where

$$\mathbf{z} = \mathbf{R}^{-1/2}\mathbf{y}, \qquad \mathbf{Z} = \mathbf{R}^{-1/2}\mathbf{X}, \qquad \mathbf{f} = \mathbf{R}^{-1/2}\mathbf{e}$$

OLS can be applied to this model since the transformed residuals are uncorrelated and homoscedastic. Thus, GLS estimates are obtained from the OLS solution by substituting

$$\mathbf{z} = \mathbf{R}^{-1/2}\mathbf{y} \text{ for } \mathbf{y}, \qquad \mathbf{Z} = \mathbf{R}^{-1/2}\mathbf{X} \text{ for } \mathbf{X}, \qquad \mathbf{f} = \mathbf{R}^{-1/2}\mathbf{e} \text{ for } \mathbf{e} \tag{A3.10}$$

Substituting into the OLS solutions (Equation 10.9a) gives the GLS estimate of $\boldsymbol{\beta}$ as

$$\begin{aligned}\widehat{\boldsymbol{\beta}} &= \left((\mathbf{X}^T\mathbf{R}^{-1/2})(\mathbf{R}^{-1/2}\mathbf{X}) \right)^{-1} (\mathbf{X}^T\mathbf{R}^{-1/2})(\mathbf{R}^{-1/2}\mathbf{y}) \\ &= \left(\mathbf{X}^T\mathbf{R}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{R}^{-1}\mathbf{y}\end{aligned}$$

Likewise, substituting into the OLS covariance expression (Equation 10.11a) gives the resulting covariance matrix for the GLS estimates as

$$\text{Var}(\widehat{\boldsymbol{\beta}}) = \sigma_e^2 \left(\mathbf{X}^T\mathbf{R}^{-1}\mathbf{X}\right)^{-1}$$

If the residuals follow a multivariate normal distribution, $\mathbf{e} \sim \text{MVN}(\mathbf{0}, \mathbf{V})$, and $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$ is indeed the correct model, then for $\widehat{\mathbf{y}} = \mathbf{X}\boldsymbol{\beta}$, $\mathbf{y} - \widehat{\mathbf{y}} \sim \text{MVN}(\mathbf{0}, \mathbf{V})$ and it follows from Equation A3.9 that

$$(\mathbf{y} - \widehat{\mathbf{y}})^T \mathbf{V}^{-1} (\mathbf{y} - \widehat{\mathbf{y}}) \sim \chi^2_{n-p} \tag{A3.11a}$$

The degrees of freedom for the χ^2 distribution equals the number of observations minus the number of estimated parameters. Equation A3.11a provides a χ^2 test for the goodness-of-fit of a particular linear model. If $\mathbf{V}$ is a diagonal matrix, then

$$(\mathbf{y} - \widehat{\mathbf{y}})^T \mathbf{V}^{-1} (\mathbf{y} - \widehat{\mathbf{y}}) = \sum_{i=1}^{n} \frac{(y_i - \widehat{y}_i)^2}{V_{ii}} \sim \chi^2_{n-p} \tag{A3.11b}$$

QUADRATIC FORMS AND SUMS OF SQUARES

The analysis of linear models relies very heavily on **sums of squares** (Chapter 22), which can be expressed in matrix notation as quadratic forms (Equation 9.17a). To introduce the reader to the machinery used to work with sums of squares, we first present expressions for the mean and variance of a quadratic form, and then express linear model sums of squares as quadratic forms.

Moments of Quadratic Forms

When $\mathbf{x}$ is a vector of random variables, the quadratic form $\mathbf{x}^T\mathbf{A}\mathbf{x}$ is a scalar (1×1) random variable. If $\mathbf{x}$ has mean $\boldsymbol{\mu}$ and (nonsingular) covariance matrix $\mathbf{V}$, Equation 9.22 gives the expected value of this quadratic form as

$$E(\mathbf{x}^T\mathbf{A}\mathbf{x}) = \text{tr}(\mathbf{A}\mathbf{V}) + \boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu} \tag{A3.12a}$$

where the **trace** of a square matrix, $\text{tr}(\mathbf{M}) = \sum M_{ii}$, is the sum of its diagonal elements. Further, if $\mathbf{x} \sim \text{MVN}(\boldsymbol{\mu}, \mathbf{V})$, then as shown in Searle (1971), the variance of the quadratic form is given by

$$\sigma^2(\mathbf{x}^T\mathbf{A}\mathbf{x}) = 2\,\text{tr}\,(\mathbf{A}\mathbf{V}\mathbf{A}\mathbf{V}) + 4\boldsymbol{\mu}^T\mathbf{A}\mathbf{V}\mathbf{A}\boldsymbol{\mu} \tag{A3.12b}$$

The Sample Variance Expressed as a Quadratic Form

As an introduction to expressing sums of squares as quadratic forms, consider the sample variance for n observations,

$$\text{Var}(x) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

Define the **unit matrix** $\mathbf{J}_{n\times k}$ as an $n \times k$ matrix in which every element is unity, e.g.,

$$\mathbf{J}_{n\times 1} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}\Big\}n\,, \qquad \mathbf{J}_{2\times 3} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \qquad \text{etc.}$$

Likewise, define the matrix

$$\mathbf{N} = \frac{1}{n-1}\left(\mathbf{I} - \frac{1}{n}\mathbf{J}\right) = \frac{1}{n-1}\begin{pmatrix} 1-1/n & -1/n & \cdots & -1/n \\ -1/n & 1-1/n & \cdots & -1/n \\ \vdots & \vdots & \ddots & \vdots \\ -1/n & -1/n & \cdots & 1-1/n \end{pmatrix} \tag{A3.13a}$$

where here $\mathbf{J}$ is $n \times n$. Noting that

$$\mathbf{N}\mathbf{x} = \frac{1}{n-1}\left(\mathbf{x} - \frac{1}{n}\mathbf{J}\mathbf{x}\right) = \frac{1}{n-1}\begin{pmatrix} x_1 - \overline{x} \\ \vdots \\ x_n - \overline{x} \end{pmatrix} \tag{A3.13b}$$

it follows that

$$\mathbf{x}^T\mathbf{N}\mathbf{x} = \text{Var}(x) \tag{A3.14a}$$

To see this, observe that

$$\begin{aligned} \mathbf{x}^T\mathbf{N}\mathbf{x} &= \frac{1}{n-1}\begin{pmatrix} x_1 & \cdots & x_n \end{pmatrix}\begin{pmatrix} x_1 - \overline{x} \\ \vdots \\ x_n - \overline{x} \end{pmatrix} \\ &= \frac{1}{n-1}\sum_{i=1}^{n} x_i(\,x_i - \overline{x}\,) = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - \overline{x}\sum_{i=1}^{n} x_i\right) \\ &= \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x}\,)^2 = \text{Var}(x) \end{aligned} \tag{A3.14b}$$

The last step follows by noting that

$$-\overline{x}\sum_{i=1}^{n} x_i = -2\,\overline{x}\sum_{i=1}^{n} x_i + n\,\overline{x}^2$$

Example A3.4. Since we have expressed Var(x) as a quadratic form, we can use Equation A3.12a to compute its expected value and Equation A3.12b to compute its sampling variance. If $\mathbf{x} \sim (\boldsymbol{\mu}, \mathbf{V})$, the expected value of Var(x) is

$$E[\,\text{Var}(x)\,] = E(\mathbf{x}^T\mathbf{N}\mathbf{x}) = \text{tr}(\mathbf{N}\mathbf{V}) + \boldsymbol{\mu}^T\mathbf{N}\boldsymbol{\mu}$$

To compute this expression, first note from Equation A3.14b that

$$\boldsymbol{\mu}^T \mathbf{N} \boldsymbol{\mu} = \frac{1}{n-1} \sum_{i=1}^{n} (\, \mu_i - \overline{\mu}\,)^2$$

Likewise, from Equation A3.13b

$$\mathbf{NV} = \frac{\mathbf{V}}{n-1} - \frac{\mathbf{JV}}{n(n-1)}$$

which has diagonal elements

$$(\mathbf{NV})_{ii} = \frac{1}{n-1}\left(\sigma^2(z_i) - \frac{\sum_j \sigma(z_i, z_j)}{n}\right)$$

After some simplification, we have

$$\text{tr}(\mathbf{NV}) = \sum_{i=1}^{n} (\mathbf{NV})_{ii} = \frac{1}{n}\sum_{i=1}^{n} \sigma^2(z_i) - \frac{2}{n(n-1)} \sum_{i<j} \sigma(z_i, z_j)$$

Putting these results together gives

$$E[\text{Var}(x)] = \frac{1}{n}\sum_{i=1}^{n} \sigma^2(z_i) - \frac{2}{n(n-1)} \sum_{i<j} \sigma(z_i, z_j) + \frac{1}{n-1} \sum_{i=1}^{n} (\, \mu_i - \overline{\mu}\,)^2$$

where $\overline{\mu} = \sum \mu_i / n$. In the simple situation where all observations have the same mean and variance ($\mu_i = \mu$, $\sigma^2(z_i) = \sigma^2$) and are uncorrelated, this reduces to

$$E[\,\text{Var}(x)\,] = \sigma^2$$

Turning now to the sample variance of Var(x), if we are willing to assume that $\mathbf{x}$ is multivariate normal, then from Equation A3.12b,

$$\sigma^2[\text{Var}(x)] = \sigma^2(\mathbf{x}^T \mathbf{N} \mathbf{x}) = 2\,\text{tr}\,[\mathbf{NVNV}] + 4\boldsymbol{\mu}^T \mathbf{NVN}\boldsymbol{\mu}$$

If, for example, $\mathbf{V} = \sigma^2\,\mathbf{I}$ (the x_i are uncorrelated with common variance), then

$$\begin{aligned}\mathbf{NVNV} = \sigma^4\,\mathbf{NN} &= \frac{\sigma^4}{(n-1)^2}\left(\mathbf{I} - \frac{1}{n}\,\mathbf{J}_{n\times n}\right)\left(\mathbf{I} - \frac{1}{n}\,\mathbf{J}_{n\times n}\right) \\ &= \frac{\sigma^4}{(n-1)^2}\left(\mathbf{I} - \frac{2}{n}\,\mathbf{J}_{n\times n} + n^{-2}\,\mathbf{J}_{n\times n}\,\mathbf{J}_{n\times n}\right)\end{aligned}$$

The *ij*th element in $\mathbf{J}_{n\times n}\,\mathbf{J}_{n\times n}$ is n, giving $\mathbf{J}^2_{n\times n} = n\,\mathbf{J}_{n\times n}$. Hence, the *i*th diagonal element of $\mathbf{NVNV}$ is

$$\frac{\sigma^4}{(n-1)^2}\left(1 - \frac{2}{n} + n^{-2} n\right) = \frac{\sigma^4}{n(n-1)}$$

giving $\text{tr}(\mathbf{NVNV}) = \sigma^4/(n-1)$. When all of the means are equal, it follows that $\mathbf{N}\boldsymbol{\mu} = \mathbf{0}$ and the second term in Equation A3.12b vanishes, yielding

$$\sigma^2[\,\text{Var}(x)\,] = \frac{2\sigma^4}{n-1}$$

Sums of Squares Expressed as Quadratic Forms

In the same fashion that we decomposed the total variance into genetic and phenotypic components (Chapters 4–7), we can decompose the total variance of a response vector $\mathbf{y}$ into the variance accounted for by the linear model and the remaining (error or residual) variance. This is typically done by considering three sums of squares, with the **total sum of squares** (SS_T) being the sum of two components, the **error** (or **residual**) **sum of squares** (SS_E) and the **model sum of squares** (SS_M),

$$SS_T = SS_M + SS_E$$

The total sum of squares measures the total variability in the data, while the model sum of squares measures the amount of variation accounted for by the linear model. As noted in our discussions of univariate regression in Chapter 3, the fraction of total variance explained by a linear model is given by the **coefficient of determination**,

$$r^2 = \frac{SS_M}{SS_T} = 1 - \frac{SS_E}{SS_T} \tag{A3.15}$$

As summarized in Table A3.1, the sums of squares have different forms under OLS and GLS. Under OLS, the residuals are assumed to be independent with common variance σ_e^2. In this case, each observation/residual is weighted equally, and the total sum of squares is simply

$$SS_T = \sum_{i=1}^{n} (\, y_i - \overline{y} \,)^2$$

Sums of squares can be expressed as a quadratic form of the vector of observations $\mathbf{y}$, allowing the use of Equations 3A.12a and 3A.12b to obtain their expectations and variances. Recalling Equation A3.14b and A3.13a,

$$SS_T = \mathbf{y}^T \left(\mathbf{I} - \frac{1}{n} \mathbf{J} \right) \mathbf{y} \tag{A3.16a}$$

where $\mathbf{J}$ is $n \times n$.

Now consider the error sum of squares,

$$SS_E = \sum_{i=1}^{n} (\, y_i - \widehat{y}_i \,)^2 = \sum_{i=1}^{n} \widehat{e}_i^2$$

Since $\widehat{\mathbf{e}} = \mathbf{y} - \widehat{\mathbf{y}}$ and $\widehat{\mathbf{y}} = \mathbf{X}\mathbf{b} = \mathbf{X} \left(\mathbf{X}^T \mathbf{X} \right)^{-1} \mathbf{X}^T \mathbf{y}$, we have

$$SS_E = \widehat{\mathbf{e}}^T \widehat{\mathbf{e}}, \quad \text{where} \quad \widehat{\mathbf{e}} = \left[\mathbf{I} - \mathbf{X} \left(\mathbf{X}^T \mathbf{X} \right)^{-1} \mathbf{X}^T \right] \mathbf{y} \tag{A3.16b}$$

Expanding this expression and noting that $\mathbf{X}^T \mathbf{X} \left(\mathbf{X}^T \mathbf{X} \right)^{-1} = \mathbf{I}$, this simplifies to

$$SS_E = \mathbf{y}^T \left[\mathbf{I} - \mathbf{X} \left(\mathbf{X}^T \mathbf{X} \right)^{-1} \mathbf{X}^T \right] \mathbf{y} \tag{A3.16c}$$

Finally, the model sum of squares is the difference between the total and error sums of squares,

$$SS_M = SS_T - SS_E = \mathbf{y}^T \left[\mathbf{X} \left(\mathbf{X}^T \mathbf{X} \right)^{-1} \mathbf{X}^T - \frac{1}{n} \mathbf{J} \right] \mathbf{y} \tag{A3.16d}$$

Table A3.1 Summary of useful results for the general linear model, $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$, under OLS and GLS assumptions for the distribution of residuals. The $\mathbf{Q}$ matrix for the listed sums of squares is the matrix in the quadratic form, $\mathbf{y}^T\mathbf{Q}\mathbf{y}$, while $\mathbf{J}$ is a matrix where each element is 1.

Ordinary Least Squares, OLS	**Generalized Least Squares, GLS**
Assumed distribution of residuals:	
$\mathbf{e} \sim (\mathbf{0}, \sigma_e^2\,\mathbf{I})$	$\mathbf{e} \sim (\mathbf{0}, \sigma_e^2\,\mathbf{R})$
Least-squares estimator of $\boldsymbol{\beta}$:	
$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$	$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{R}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{R}^{-1}\mathbf{y}$
Covariance matrix for $\widehat{\boldsymbol{\beta}}$:	
$(\mathbf{X}^T\mathbf{X})^{-1}\sigma_e^2$	$(\mathbf{X}^T\mathbf{R}^{-1}\mathbf{X})^{-1}\sigma_e^2$
Estimated residual variance, $\widehat{\sigma_e^2}$	
$\frac{(\mathbf{y}-\mathbf{X}\mathbf{b})^T(\mathbf{y}-\mathbf{X}\mathbf{b})}{n - \text{rank}(\mathbf{X})}$	$\frac{(\mathbf{y}-\mathbf{X}\mathbf{b})^T\mathbf{R}^{-1}(\mathbf{y}-\mathbf{X}\mathbf{b})}{n - \text{rank}(\mathbf{X})}$
Predicted values, $\widehat{\mathbf{y}} = \mathbf{X}\widehat{\boldsymbol{\beta}}$:	
$\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$	$\mathbf{X}(\mathbf{X}^T\mathbf{R}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{R}^{-1}\mathbf{y}$
Covariance matrix for predicted values, $\widehat{\mathbf{y}}$:	
$\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\sigma_e^2$	$\mathbf{X}(\mathbf{X}^T\mathbf{R}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\sigma_e^2$
SS_T, total sums of squares quadratic form matrix, $\mathbf{Q}$:	
$\mathbf{I} - \frac{1}{n}\,\mathbf{J}$	$\mathbf{R}^{-1} - \frac{1}{n}\,\mathbf{R}^{-1/2}\mathbf{J}\mathbf{R}^{-1/2}$
SS_M, model sums of squares quadratic form matrix, $\mathbf{Q}$:	
$\mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T - \frac{1}{n}\,\mathbf{J}$	$\mathbf{R}^{-1}\mathbf{X}\left(\mathbf{X}^T\mathbf{R}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{R}^{-1} - \frac{1}{n}\,\mathbf{R}^{-1/2}\mathbf{J}\mathbf{R}^{-1/2}$
SS_E, error sums of squares quadratic form matrix, $\mathbf{Q}$:	
$\mathbf{I} - \mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T$	$\mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{X}\left(\mathbf{X}^T\mathbf{R}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{R}^{-1}$
χ^2 goodness of fit statistic (assuming residuals are MVN):	
$\chi^2 = \frac{(\mathbf{y}-\widehat{\mathbf{y}})^T(\mathbf{y}-\widehat{\mathbf{y}})}{\sigma_e^2}$	$\chi^2 = \frac{(\mathbf{y}-\widehat{\mathbf{y}})^T\mathbf{R}^{-1}(\mathbf{y}-\widehat{\mathbf{y}})}{\sigma_e^2}$

Note that

$$\text{SS}_\text{M} = \sum_{i=1}^{n}(\,\widehat{y}_i - \overline{y}\,)^2$$

so that (for OLS) the model sum of squares is the sum of squared deviations of the predicted values from the overall mean.

The sums of squares under generalized least-squares (GLS) are slightly different, as we have to correct for heteroscedasticity and/or the lack of independence among the residuals. Assume that the residuals have covariance matrix $\sigma_e^2\,\mathbf{R}$. From Equation A3.10, $\mathbf{y}$ is replaced

by $\mathbf{R}^{-1/2}\mathbf{y}$ and $\mathbf{X}$ is replaced by $\mathbf{R}^{-1/2}\mathbf{X}$ in the above OLS expressions for sums of squares. Hence, the total sum of squares for GLS becomes

$$\begin{aligned} \mathrm{SS_T} &= \mathbf{y}^T\mathbf{R}^{-1/2}\left(\mathbf{I} - \frac{1}{n}\mathbf{J}\right)\mathbf{R}^{-1/2}\mathbf{y} \\ &= \mathbf{y}^T\left[\mathbf{R}^{-1} - \frac{1}{n}\mathbf{R}^{-1/2}\mathbf{J}\mathbf{R}^{-1/2}\right]\mathbf{y} \end{aligned} \tag{A3.17a}$$

Likewise, the error sum of squares becomes

$$\begin{aligned} \mathrm{SS_E} &= \widehat{\mathbf{e}}^T\mathbf{R}^{-1}\widehat{\mathbf{e}} \\ &= \mathbf{y}^T\left[\mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{X}\left(\mathbf{X}^T\mathbf{R}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{R}^{-1}\right]\mathbf{y} \end{aligned} \tag{A3.17b}$$

and the model sum of squares becomes

$$\mathrm{SS_M} = \mathbf{y}^T\left[\mathbf{R}^{-1}\mathbf{X}\left(\mathbf{X}^T\mathbf{R}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{R}^{-1} - \frac{1}{n}\mathbf{R}^{-1/2}\mathbf{J}\mathbf{R}^{-1/2}\right]\mathbf{y} \tag{A3.17c}$$

TESTING HYPOTHESES ABOUT LINEAR MODELS

Because sums of squares are very closely related to the variances accounted for by the various components of a particular linear model, it should not be surprising that these sums form the basis of hypothesis testing. Such tests can be quite involved, especially if we are evaluating the various components of a complex model. Here we consider the simplest case of the nested comparison of a full model with a reduced one.

If the residuals are multivariate-normally distributed with

$$\mathbf{e} \sim \mathrm{MVN}(\mathbf{0}, \sigma_e^2\,\mathbf{I}) \quad \text{for OLS;} \qquad \mathbf{e} \sim \mathrm{MVN}(\mathbf{0}, \sigma_e^2\,\mathbf{R}) \quad \text{for GLS}$$

then (recalling Equation A3.11a and A3.11b), $\mathrm{SS_E}/\sigma_e^2$ is the sum of squared unit normals and hence is χ^2-distributed. In particular, with n observations and p estimated parameters,

$$\frac{\mathrm{SS_E}}{\sigma_e^2} \sim \chi^2_{n-p} \tag{A3.18}$$

as a degree of freedom is lost for each estimated model parameter.

Suppose we have n observations and wish to compare two linear models: a **full model** fitting p parameters and a **reduced model** which uses only a subset ($q < p$) of the parameters in the full model. Do the additional $(p - q)$ fitted parameters provide a significant increase in the amount of variation accounted for by the model? Let $\mathrm{SS_{E_f}}$ and $\mathrm{SS_{E_r}}$ denote the appropriate (OLS or GLS) error sums of squares for the full and reduced models, respectively. Under the null hypothesis (that the full model provides the same fit as the reduced model), the difference in error sums of squares $(\mathrm{SS_{E_r}} - \mathrm{SS_{E_f}})$ is distributed as constant (σ_e^2) times a χ^2_{p-q}. Likewise, from Equation A3.18, $\mathrm{SS_{E_f}} \sim \sigma_e^2\,\chi^2_{n-p}$. Recalling the definition of the F distribution (the ratio of two scaled chi-square distributions; Appendix 5) it follows that

$$\frac{\left(\mathrm{SS_{E_r}} - \mathrm{SS_{E_f}}\right)/(p-q)}{\mathrm{SS_{E_f}}/(n-p)} = \left(\frac{n-p}{p-q}\right)\left(\frac{\mathrm{SS_{E_r}}}{\mathrm{SS_{E_f}}} - 1\right) \tag{A3.19}$$

is distributed as $F_{p-q,n-p}$ under the null hypothesis of no improved fit.

For example, we can ask if a particular linear model accounts for a significant fraction of the variation in y by considering that model versus the simplest reduced model, $y_i = \mu + e_i$

(namely, that none of the model factors have any effect). It is easily seen that the least-squares solution for μ is $\overline{y}$ for OLS and the weighted mean for GLS, giving $\mathrm{SS}_{E_r} = \mathrm{SS}_{\mathrm{T}}$. Since the number of parameters in the reduced model is $q = 1$, the test for whether a particular linear model accounts for a significant amount of the variation is

$$\left(\frac{n-p}{p-1}\right)\left(\frac{\mathrm{SS}_{\mathrm{T}}}{\mathrm{SS}_{\mathrm{E}_f}} - 1\right) = \left(\frac{n-p}{p-1}\right)\left(\frac{r^2}{1-r^2}\right) \tag{A3.20}$$

where r^2 is the coefficient of determination for the full model (Equation A3.15). This test statistic follows an $F_{p-1,n-p}$ distribution.

Construction of Confidence Intervals

When residuals are MVN, then OLS/GLM estimates are also MVN, with

$$\widehat{\boldsymbol{\beta}} \sim \mathrm{MNV}\left(\boldsymbol{\beta}, \mathbf{V_b}\right) \tag{A3.21}$$

where $\mathbf{V_b}$ is given by Equation 10.11a for OLS estimates and by Equation 10.13b for GLS estimates. Confidence intervals immediately follow from the MVN property that subsets are also MVN (Chapter 9). Hence, the standard 95% confidence normal interval holds for a given GLM estimate, with

$$\widehat{\beta}_i \pm 1.96\sqrt{V_{ii}} \tag{A3.22a}$$

where V_{ii} is the ith diagonal element in $\mathbf{V_b}$. Similarly, the P value for a two-sided test of the hypothesis that $\beta_i = \beta_{i,o}$ is given by

$$\Pr\left(|U| \geq \left|\frac{\widehat{\beta}_i - \beta_{i,o}}{\sqrt{V_{ii}}}\right|\right) \tag{A3.22b}$$

where U is a unit normal random variable.

Multivariate joint confidence intervals are in the form of ellipses (for two dimensions) and ellipsoids for high dimensions. To motivate the form of these intervals, we first consider **Hotelling's $\mathbf{T}^2$** statistic. This is essentially the multivariate extension of the classic univariate t test, where for $\overline{z} \sim \mathrm{N}(\mu_0, \sigma^2/n)$,

$$t = \frac{(\overline{z} - \mu_0)^2}{S^2/n}$$

for the null hypothesis that the mean is μ_0 given that the sample variance of z is S^2 (and hence the sample variance for $\overline{z}$ is S^2/n). Squaring both sides, we can express this as

$$t^2 = (\overline{z} - \mu_0)(S^2/n)^{-1}(\overline{z} - \mu_0)$$

Hotelling's T^2 statistic generalizes this to multivariate form, where for $\overline{\mathbf{z}} \sim \mathrm{MNV}(\boldsymbol{\mu}_0, \mathbf{S}/n)$,

$$T^2 = (\overline{\mathbf{z}} - \boldsymbol{\mu}_0)^T(\mathbf{S}/n)^{-1}(\overline{\mathbf{z}} - \boldsymbol{\mu}_0) \tag{A3.23a}$$

Under the null hypothesis ($\boldsymbol{\mu} = \boldsymbol{\mu}_0$),

$$\left(\frac{n-p}{(n-1)p}\right) \cdot T^2 \sim F_{p,n-p} \tag{A3.23b}$$

Confidence ellipsoids follow from these expressions under MVN assumptions. Because $\widehat{\boldsymbol{\beta}} \sim \mathrm{MNV}(\boldsymbol{\beta}, \mathbf{V_b})$, the set of all $\boldsymbol{\beta}$ values that are consistent (at the $1-\alpha$ percent level) with the observed data satisfies

$$(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T\mathbf{V_b}^{-1}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \leq \left(\frac{(n-1)p}{n-p}\right) F_{p,n-p}(1-\alpha) \tag{A3.24}$$

The critical value $F_{p,n-p}(1-\alpha)$ corresponds to the value of an F distribution with p numerator and $n-p$ denominator degrees of freedom that satisfies

$$\Pr\left[F_{p,n-p} \leq F_{p,n-p}(1-\alpha)\right] = 1-\alpha$$

The quadratic form given by Equation A3.24 describes an ellipsoid centered at $\widehat{\boldsymbol{\beta}}$ whose axes are given by the eigenvectors of $\mathbf{V_b}$ (WL Appendix 5 examines the analysis of such quadratic forms).

Example A3.5. Suppose that an analysis of a GLM returns the following values

$$\widehat{\boldsymbol{\beta}} = \begin{pmatrix} 1 \\ -2 \end{pmatrix}, \quad \mathbf{V_b} = \begin{pmatrix} 0.11 & -0.02 \\ -0.02 & 0.05 \end{pmatrix}$$

The two univariate 95% confidence intervals become

$$\widehat{\beta_1} \pm 1.96\sqrt{0.11} \ (\pm 0.65), \quad \text{and} \quad \widehat{\beta_2} \pm 1.96\sqrt{0.05} \ (\pm 0.44)$$

or $(0.35, 1.65)$ for β_1 and $(-2.44, -1.66)$ for β_2.

EQUIVALENT LINEAR MIXED MODELS

Two mixed models are said to be **equivalent** if they have the same mean vector $E(\mathbf{y})$ and covariance matrix $\sigma(\mathbf{y}, \mathbf{y})$. The utility of equivalent models is that the parameters of one model can always be expressed as linear combinations of the parameters of any equivalent model. Hence, by choosing an appropriate equivalent model, one can often greatly simplify computations. An example of this approach is the reduced animal model of Quaas and Pollak (1980) discussed in Chapter 30. Likewise, Equation 30.23, for estimating the BLUP values of dominance effects as a function of estimated breeding values, also follows from using equivalent models. Additional examples from BLUP are given by Henderson (1985c). Our purpose here is to briefly introduce the use and construction of equivalent models.

Consider two different mixed linear models, both using the same vector $\mathbf{y}$ of observations but with different assumed vectors of fixed ($\boldsymbol{\beta}$ vs. $\boldsymbol{\beta}_*$) and random ($\mathbf{u}$ and $\mathbf{e}$ vs. $\mathbf{u}_*$ and $\mathbf{e}_*$) effects. Model 1 is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}, \quad \text{where} \quad \mathbf{u} \sim (\mathbf{0}, \mathbf{G}) \quad \text{and} \quad \mathbf{e} \sim (\mathbf{0}, \mathbf{R})$$

while model 2 is

$$\mathbf{y} = \mathbf{X}_*\boldsymbol{\beta}_* + \mathbf{Z}_*\mathbf{u}_* + \mathbf{e}_*, \quad \text{where} \quad \mathbf{u}_* \sim (\mathbf{0}, \mathbf{G}_*) \quad \text{and} \quad \mathbf{e}_* \sim (\mathbf{0}, \mathbf{R}_*)$$

Recalling our treatment of general mixed linear models (Chapter 10), Equation 10.19 implies that for model 1,

$$\mathbf{y} \sim (\mathbf{X}\boldsymbol{\beta}, \mathbf{V}), \quad \text{where} \quad \mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}^T + \mathbf{R}$$

while for model 2,

$$\mathbf{y} \sim (\mathbf{X}_*\boldsymbol{\beta}_*, \mathbf{V}_*), \quad \text{where} \quad \mathbf{V}_* = \mathbf{Z}_*\mathbf{G}_*\mathbf{Z}_*^T + \mathbf{R}_*$$

Thus, these two models are equivalent if

$$\mathbf{X}\boldsymbol{\beta} = \mathbf{X}_*\boldsymbol{\beta}_* \tag{A3.25a}$$

and $\mathbf{V} = \mathbf{V}_*$, or

$$\mathbf{ZGZ}^T + \mathbf{R} = \mathbf{Z}_*\mathbf{G}_*\mathbf{Z}_*^T + \mathbf{R}_* \tag{A3.25b}$$

Equations A3.25a and A3.25b provide the framework for constructing equivalent models, and hence obtaining models that are potentially easier to analyze. Consider the situation where our interest is in the prediction of random effects and we wish to obtain an equivalent model that considers the same fixed effects but uses a different vector of random effects. (For example, instead of considering a vector of both parental and offspring breeding values, we might simply consider the vector of parental breeding values, using the parental estimates to subsequently estimate the breeding values in their offspring.) If the original model is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Zu} + \mathbf{e}, \qquad \text{where} \quad \mathbf{u} \sim (0, \mathbf{G}), \quad \text{and} \quad \mathbf{e} \sim (0, \mathbf{R})$$

and we wish to construct an equivalent model using *any* vector of random effects $\mathbf{u}_* \sim (0, \mathbf{G}_*)$ with

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_*\mathbf{u}_* + \mathbf{e}_*, \qquad \text{where} \quad \mathbf{u}_* \sim (0, \mathbf{G}_*), \quad \text{and} \quad \mathbf{e}_* \sim (0, \mathbf{R}_*)$$

For these models to be equivalent requires that $\mathbf{V} = \mathbf{V}_*$, hence it immediately follows from Equation A3.25b that the covariance matrix for the vector of new residual values, $\mathbf{e}_*$, is given by

$$\mathbf{R}_* = \mathbf{R} + \mathbf{ZGZ}^T - \mathbf{Z}_*\mathbf{G}_*\mathbf{Z}_*^T \tag{A3.26}$$

Given an estimate of $\mathbf{u}_*$, an estimate of $\mathbf{u}$ can be directly obtained, as parameters of a linear model can always be expressed as linear combinations of the parameters of any equivalent model. In this case, given the BLUP estimate ($\widehat{\mathbf{u}}_*$) of $\mathbf{u}_*$, the BLUP estimate of $\mathbf{u}$ is given by

$$\widehat{\mathbf{u}} = \mathbf{C}\mathbf{G}^{-1}\widehat{\mathbf{u}}_* \tag{A3.27}$$

where $\mathbf{C}$ is the covariance matrix between $\mathbf{u}_*$ and $\mathbf{u}$, and $\mathbf{G}$ is the covariance matrix associated with $\mathbf{u}$ (Henderson 1977b). This is just the linear regression of $\mathbf{u}_*$ on $\mathbf{u}$ (see Equation 9.27). Note that the vectors $\mathbf{u}_*$ and $\mathbf{u}$ can have different dimensionality, so that if $\mathbf{u}_*$ is $r \times 1$ and $\mathbf{u}$ is $q \times 1$, then $\mathbf{C}$ is an $r \times q$ matrix with $C_{ij} = \sigma(u_{*i}, u_j)$.

DERIVATIVES OF VECTORS AND MATRICES

Our next-to-last topic in matrix algebra concerns the derivatives of vector- and matrix-valued functions, which we use rather extensively in Chapter 31. We present a few simple results here, and the reader is referred to Morrison (1976), Graham (1981), and Searle (1982) for more details. Consider first the simplest function of vector $\mathbf{x}$, namely the product of $\mathbf{x}$ and either a vector ($\mathbf{a}$) or matrix ($\mathbf{A}$) of constants. The derivatives of these functions with respect to the vector $\mathbf{x}$ become

$$\frac{\partial\, \mathbf{a}^T\mathbf{x}}{\partial\, \mathbf{x}} = \frac{\partial\, \mathbf{x}^T\mathbf{a}}{\partial\, \mathbf{x}} = \mathbf{a} \tag{A3.28a}$$

$$\frac{\partial\, \mathbf{Ax}}{\partial\, \mathbf{x}} = \mathbf{A}^T \tag{A3.28b}$$

Turning to quadratic forms, if $\mathbf{A}$ is symmetric, then

$$\frac{\partial\, \mathbf{x}^T\mathbf{Ax}}{\partial\, \mathbf{x}} = 2\mathbf{Ax} \tag{A3.29a}$$

Three useful identities involving quadratic forms follow

$$\frac{\partial\, (\mathbf{a} - \mathbf{x})^T\mathbf{A}(\mathbf{a} - \mathbf{x})}{\partial\, \mathbf{x}} = -2\mathbf{A}(\mathbf{a} - \mathbf{x}) \tag{A3.29b}$$

$$\frac{\partial\, (\mathbf{a} - \mathbf{Bx})^T(\mathbf{a} - \mathbf{Bx})}{\partial\, \mathbf{x}} = -2\mathbf{B}^T(\mathbf{a} - \mathbf{Bx}) \tag{A3.29c}$$

$$\frac{\partial\, (\mathbf{a} - \mathbf{Bx})^T\mathbf{A}(\mathbf{a} - \mathbf{Bx})}{\partial\, \mathbf{x}} = -2\mathbf{B}^T\mathbf{A}(\mathbf{a} - \mathbf{Bx}) \tag{A3.29d}$$

Two final useful identities follow from the chain rule of differentiation,

$$\frac{\partial \exp[\, f(\mathbf{x})\,]}{\partial \mathbf{x}} = \exp[\, f(\mathbf{x})\,] \cdot \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} \tag{A3.30a}$$

$$\frac{\partial \ln[\, f(\mathbf{x})\,]}{\partial \mathbf{x}} = \frac{1}{f(\mathbf{x})} \cdot \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} \tag{A3.30b}$$

Example A3.6. The OLS solution for a linear model is the value of $\boldsymbol{\beta}$ that minimizes the residual sum of squares given $\mathbf{y}$ and $\mathbf{X}$. In matrix form,

$$\sum_{i=1}^{n} e_i^2 = \mathbf{e}^T\mathbf{e} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{x}\boldsymbol{\beta})$$

Taking the derivative with respect to $\boldsymbol{\beta}$ and using Equation A3.29c (with $\mathbf{a} = \mathbf{y}$, $\mathbf{B} = \mathbf{X}$, and $\mathbf{x} = \boldsymbol{\beta}$) gives

$$\frac{\partial\, \mathbf{e}^T\mathbf{e}}{\partial\, \boldsymbol{\beta}} = \frac{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{x}\boldsymbol{\beta})}{\partial\, \boldsymbol{\beta}} = -2\mathbf{X}^T\,(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

Setting this equal to zero gives $\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} = \mathbf{X}^T\mathbf{y}$, which has solution

$$\boldsymbol{\beta} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}$$

If $\mathbf{X}^T\mathbf{X}$ is singular, a generalized inverse is used instead.

Example A3.7. Writing the MVN distribution as

$$\varphi(\mathbf{x}) = a \exp\left(-\frac{1}{2} \cdot (\mathbf{x} - \boldsymbol{\mu})^T\, \mathbf{V}^{-1}\, (\mathbf{x} - \boldsymbol{\mu})\right)$$

where $a = \pi^{-n/2}\, |\mathbf{V}|^{-1/2}$, then from Equation A3.30a,

$$\frac{\partial\, \varphi(\mathbf{x})}{\partial\, \mathbf{x}} = \varphi(\mathbf{x}) \cdot \frac{\partial\, \left[\, \left(-\frac{1}{2}\right) \cdot (\mathbf{x} - \boldsymbol{\mu})^T\, \mathbf{V}^{-1}\, (\mathbf{x} - \boldsymbol{\mu})\,\right]}{\partial\, \mathbf{x}}$$

Applying Equation A3.29b yields

$$\frac{\partial\, \varphi(\mathbf{x})}{\partial\, \mathbf{x}} = -\varphi(\mathbf{x}) \cdot \mathbf{V}^{-1}\, (\mathbf{x} - \boldsymbol{\mu}) \tag{A3.31a}$$

Note that $\varphi(\mathbf{x})$ is a scalar and hence its order of multiplication does not matter, while the order of the other variables (being matrices) is critical. Similarly, we can consider the MVN as a function of the mean vector $\boldsymbol{\mu}$, in which case Equation A3.30a implies

$$\frac{\partial\, \varphi(\mathbf{x}, \boldsymbol{\mu})}{\partial\, \boldsymbol{\mu}} = \varphi(\mathbf{x}, \boldsymbol{\mu}) \cdot \mathbf{V}^{-1}\, (\mathbf{x} - \boldsymbol{\mu}) \tag{A3.31b}$$

MATRIX DECOMPOSITIONS

Any matrix can be decomposed into a product of simpler matrices. This is the singular value decomposition (SVD), which is the foundation for AMMI analysis of genotype $\times$ environment interactions (Chapter 26), and also provides a method for obtaining a generalized inverse (Equation A3.38). With a square symmetric matrix, a special case of the SVD, diagonalization, provides significant insight on the eigenstructure of inverses, square roots, and other functions of a diagonalizable matrix. We consider diagonalization first before concluding by briefly examining the SVD.

Diagonalization of a Covariance Matrix

We previously introduced the square-root decomposition of a nonsingular symmetric matrix, $\mathbf{V} = \mathbf{V}^{1/2}\mathbf{V}^{1/2}$ (Equation A3.7a), and showed its utility in generating a new set of uncorrelated random variables (Equation A3.8a). This particular decomposition immediately follows from the more general diagonalization decomposition. In particular, a symmetric matrix $\mathbf{A}$ (such as a covariance matrix) can be **diagonalized** as

$$\mathbf{A} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T \tag{A3.32a}$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix and $\mathbf{U} = (\,\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n\,)$ is an **orthonormal matrix**. The columns of such matrices have unit length ($||\mathbf{u}_i|| = 1$) and are orthogonal ($\mathbf{u}_i^T\mathbf{u}_j = 0$ for $i \neq j$). These conditions also imply that $\mathbf{U}$ is a **unitary matrix**, with its inverse being given by its transpose, $\mathbf{U}^{-1} = \mathbf{U}^T$. Unitary matrices generate a rigid rotation (the angle between the vectors $\mathbf{x}$ and $\mathbf{y}$ is the same as the angle between the transformed vectors $\mathbf{Ux}$ and $\mathbf{Uy}$) of the original coordinate system to a new coordinate system given by $\mathbf{U}$ (WL Appendix 5).

If λ_i and $\mathbf{e}_i$ are the ith eigenvalue and its associated unit eigenvector of $\mathbf{A}$, then

$$\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n) = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & \lambda_n \end{pmatrix} \tag{A3.32b}$$

and

$$\mathbf{U} = (\,\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_n\,) \tag{A3.32c}$$

$\mathbf{U}$ describes a rotation of the original coordinate system to a new coordinate system given by the eigenvectors of $\mathbf{A}$, while the diagonal elements of $\boldsymbol{\Lambda}$ give the amount by which vectors of unit length in the original coordinate system are scaled in the transformed system. If we use the decomposition $\boldsymbol{\Lambda} = \sum_{i=1}^{n} \boldsymbol{\Lambda}_i$, where $\boldsymbol{\Lambda}_i$ is a diagonal matrix whose elements are all zero, except for λ_i, then Equation A3.32a becomes

$$\mathbf{A} = \mathbf{U}\left(\sum_{i=1}^{n} \boldsymbol{\Lambda}_i\right)\mathbf{U}^T = \sum_{i=1}^{n} \mathbf{U}\boldsymbol{\Lambda}_i\mathbf{U}^T = \sum_{i=1}^{n} \lambda_i \mathbf{e}_i\mathbf{e}_i^T \tag{A3.32d}$$

Note that $\mathbf{e}_i\mathbf{e}_i^T$, the outer product of $\mathbf{e}$ with itself (Equation 9.8b), is an $n \times n$ matrix. Equation A3.32d is called the **spectral decomposition**, **spectral factorization**, or **eigendecomposition** of $\mathbf{A}$ (WL Appendix 5).

Using Equation A3.32a, it is easy to show that

$$\mathbf{A}^{-1} = \mathbf{U}\boldsymbol{\Lambda}^{-1}\mathbf{U}^T \tag{A3.33a}$$

To see this, note that

$$\mathbf{A}^{-1}\mathbf{A} = \left(\mathbf{U}\boldsymbol{\Lambda}^{-1}\mathbf{U}^T\right)\left(\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T\right) = \mathbf{U}\boldsymbol{\Lambda}^{-1}\left(\mathbf{U}^T\mathbf{U}\right)\boldsymbol{\Lambda}\mathbf{U}^T = \mathbf{U}\boldsymbol{\Lambda}^{-1}\boldsymbol{\Lambda}\mathbf{U}^T = \mathbf{U}\mathbf{U}^T = \mathbf{I}$$

Similar logic yields

$$\mathbf{A}^{1/2} = \mathbf{U}\boldsymbol{\Lambda}^{1/2}\mathbf{U}^T \tag{A3.33b}$$
$$\mathbf{A}^{-1/2} = \mathbf{U}\boldsymbol{\Lambda}^{-1/2}\mathbf{U}^T \tag{A3.33c}$$
$$\mathbf{A}^k = \mathbf{U}\boldsymbol{\Lambda}^k\mathbf{U}^T \quad \text{for any integer } k \tag{A3.33d}$$

Further, the square root decomposition follows as

$$\begin{aligned}\mathbf{A}^{1/2}\mathbf{A}^{1/2} &= \left[\mathbf{U}\boldsymbol{\Lambda}^{1/2}\mathbf{U}^T\right]\left[\mathbf{U}\boldsymbol{\Lambda}^{1/2}\mathbf{U}^T\right] = \mathbf{U}\boldsymbol{\Lambda}^{1/2}\left[\mathbf{U}^T\mathbf{U}\right]\boldsymbol{\Lambda}^{1/2}\mathbf{U}^T \\ &= \mathbf{U}\boldsymbol{\Lambda}^{1/2}\mathbf{I}\boldsymbol{\Lambda}^{1/2}\mathbf{U}^T = \mathbf{U}\boldsymbol{\Lambda}^{1/2}\boldsymbol{\Lambda}^{1/2}\mathbf{U}^T \\ &= \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T = \mathbf{A}\end{aligned}$$

Likewise, using Equation A3.32a, we see that premultiplying **A** by $\mathbf{U}^T$ and then postmultiplying by **U** gives a diagonal matrix whose elements are the eigenvalues of **A**

$$\mathbf{U}^T\mathbf{A}\mathbf{U} = \mathbf{U}^T(\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T)\mathbf{U} = (\mathbf{U}^T\mathbf{U})\boldsymbol{\Lambda}(\mathbf{U}^T\mathbf{U}) = \boldsymbol{\Lambda} \tag{A3.34}$$

Finally, because $\boldsymbol{\Lambda}$ is diagonal, the ith diagonal elements of $\boldsymbol{\Lambda}^{-1}$, $\boldsymbol{\Lambda}^{1/2}$, $\boldsymbol{\Lambda}^{-1/2}$, and $\boldsymbol{\Lambda}^k$ are λ_i^{-1}, $\lambda_i^{1/2}$, $\lambda_i^{-1/2}$, and λ_i^k, respectively, implying that if λ_i is an eigenvalue of **A**, then λ_i^{-1}, $\lambda_i^{1/2}$, $\lambda_i^{-1/2}$, and λ_i^k, respectively, are eigenvalues of the matrices $\mathbf{A}^{-1}$, $\mathbf{A}^{1/2}$, $\mathbf{A}^{-1/2}$, and $\mathbf{A}^k$. Note that Equations A3.33a through A3.33d further imply that the matrices **A**, $\mathbf{A}^{-1}$, $\mathbf{A}^{1/2}$, $\mathbf{A}^{-1/2}$, and $\mathbf{A}^k$ all have the same eigenvectors, namely the columns of **U**.

Correlations can be Removed by a Matrix Transformation

As we saw with the square-root transform (Equation A3.8a), a powerful use of diagonalization is that it allows one to extract a set of n uncorrelated variables for any $n \times n$ nonsingular covariance matrix, $\mathbf{V_X}$ (associated with the random vector, **x**). Consider the transformation

$$\mathbf{y} = \mathbf{U}^T\mathbf{x} \tag{A3.35a}$$

where $\mathbf{U} = (\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_n)$ contains the normalized eigenvectors of $\mathbf{V_X}$. Because **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 $(e_1, \cdots, e_n)$. Applying Equation 9.21b and Equation A3.34, respectively, the covariance matrix for **y** is

$$\mathbf{V_y} = \mathbf{U}^T\mathbf{V_X}\,\mathbf{U} = \boldsymbol{\Lambda} \tag{A3.35b}$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix whose elements are the eigenvalues of $\mathbf{V_X}$,

$$\sigma(y_i, y_j) = \begin{cases} \lambda_i & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

The rigid rotation introduced by **U** thus creates a set of n uncorrelated variables, the ith of which is given by

$$y_i = \mathbf{e}_i{}^T\mathbf{x} \tag{A3.35c}$$

This is the length of the projection of **x** onto the ith eigenvector of $\mathbf{V_X}$ (WL Appendix 5), so that the axes of the new coordinate system are given by the orthogonal set of eigenvectors of $\mathbf{V_X}$.

Defining the matrix **B** as

$$\mathbf{B} = \mathbf{U}\,\boldsymbol{\Lambda}^{-1/2} \tag{A3.35d}$$

the vector $\mathbf{y} = \mathbf{B}^T\mathbf{x}$ has a covariance matrix of $\mathbf{V_y} = \mathbf{I}$, which means that this transformation creates a set of uncorrelated variables, each with unit variance. To see this, note that

$$\begin{aligned}\mathbf{V_y} = \mathbf{B}^T\mathbf{V_X}\,\mathbf{B} &= \left(\mathbf{U}\boldsymbol{\Lambda}^{-1/2}\right)^T\left(\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T\right)\left(\mathbf{U}\boldsymbol{\Lambda}^{-1/2}\right)\\ &= \boldsymbol{\Lambda}^{-1/2}\left(\mathbf{U}^T\mathbf{U}\right)\boldsymbol{\Lambda}\left(\mathbf{U}^T\mathbf{U}\right)\boldsymbol{\Lambda}^{-1/2}\\ &= \boldsymbol{\Lambda}^{-1/2}\,\boldsymbol{\Lambda}\,\boldsymbol{\Lambda}^{-1/2} = \mathbf{I}\end{aligned} \tag{A3.35e}$$

Recall that earlier we used another transformation, $\mathbf{y} = \mathbf{V_X}^{-1/2}\mathbf{x}$, where from Equation A3.33c

$$\mathbf{V_X}^{-1/2} = \mathbf{U}\boldsymbol{\Lambda}^{-1/2}\mathbf{U}^T = \mathbf{B}\mathbf{U}^T$$

Because $\mathbf{U}$ is a unitary matrix, the transformations given by $\mathbf{V_X}^{-1/2}$ and $\mathbf{B}$ are simply rigid rotations of each other. An alternative to both these transformations is the **Cholesky decomposition**, $\mathbf{A} = \mathbf{C}^T\mathbf{C}$, of a square, symmetric matrix $\mathbf{A}$, where $\mathbf{C}$ is an **lower triangular matrix** (all elements above the diagonal are zero). If $\mathbf{C}$ is the Cholesky decomposition for $\mathbf{V_X}$, then $\mathbf{y} = \mathbf{C}^{-1}\mathbf{x}$ also returns a covariance matrix, $\mathbf{V_y}$, of $\mathbf{I}$.

The Singular-Value Decomposition (SVD)

Any $n \times p$ matrix $\mathbf{A}$ can be decomposed as the product of three matrices: an $n \times p$ diagonal matrix $\boldsymbol{\Lambda}$ and two unitary matrices, $\mathbf{U}$ which is $n \times n$ and $\mathbf{V}$ which is $p \times p$. The resulting **singular value decomposition** (**SVD**) is given by

$$\mathbf{A}_{n\times p} = \mathbf{U}_{n\times n}\boldsymbol{\Lambda}_{n\times p}\mathbf{V}^T_{p\times p} \tag{A3.36a}$$

We have indicated the dimensionality of each matrix to allow the reader to verify that each matrix multiplication conforms. The (nonzero) diagonal elements $\lambda_1, \cdots, \lambda_s$ of $\boldsymbol{\Lambda}$ correspond to the **singular values** of $\mathbf{A}$ and are ordered by decreasing magnitude. Returning to the unitary matrices $\mathbf{U}$ and $\mathbf{V}$, we can write each as a row vector of column vectors,

$$\mathbf{U} = (\mathbf{u}_1, \cdots, \mathbf{u}_i, \cdots \mathbf{u}_n), \qquad \mathbf{V} = (\mathbf{v}_1, \cdots, \mathbf{v}_i, \cdots \mathbf{v}_p) \tag{A3.36b}$$

where $\mathbf{u}_i$ and $\mathbf{v}_i$ are n- and p-dimensional column vectors (often called the **left** and **right singular vectors**, respectively). Since both $\mathbf{U}$ and $\mathbf{V}$ are unitary, each column vector has length one and are mutually orthogonal (i.e., if $i \neq j$, $\mathbf{u}_i\mathbf{u}_j^T = \mathbf{v}_i\mathbf{v}_j^T = 0$). Since $\boldsymbol{\Lambda}$ is diagonal, it immediately follows from matrix multiplication that we can write any element in $\mathbf{A}$ as

$$A_{ij} = \sum_{k=1}^{s} \lambda_k\, u_{ik}\, v_{kj} \tag{A3.36c}$$

where λ_k is the kth singular value and $s \leq \min(p, n)$ is the number of nonzero singular values (s is the rank of $\mathbf{A}$). Since only the first s diagonal elements of $\boldsymbol{\Lambda}$ are nonzero, the SVD can also be written as

$$\mathbf{A}_{n\times p} = \mathbf{U}_{n\times s}\boldsymbol{\Lambda}_{s\times s}\mathbf{V}^T_{s\times p}, \quad \text{with} \quad \mathbf{U} = (\mathbf{u}_1, \cdots, \mathbf{u}_s), \quad \mathbf{V} = (\mathbf{v}_1, \cdots, \mathbf{v}_s) \tag{A3.36d}$$

This is the **compact SVD**. Since $\boldsymbol{\Lambda}$ is diagonal, we can express Equation A3.36c as

$$A_{ij} = \mathbf{u}_i\boldsymbol{\Lambda}\mathbf{v}_j^T \tag{A3.36e}$$

A final useful identity is that the total variation of $\mathbf{A}$ (the sum of all its squared values) equals the sum of its squared singular values,

$$\sum_{ij} A_{ij}^2 = \sum_{k=1}^{s} \lambda_k^2 \tag{A3.36f}$$

A key feature of the singular value decomposition arises from the **Ekart-Young theorem** (1936), which relates the best approximation of a matrix by some lower-rank (say rank $m < s$) matrix. Define as our measure of goodness of fit between a matrix $\mathbf{A}$ and a lower rank approximation $\widehat{\mathbf{A}}$ as the sum of squared differences over all elements,

$$\sum_{ij}(A_{ij} - \hat{A}_{ij})^2$$

Eckart and Young showed that the best fitting approximation $\widehat{\mathbf{A}}$ of rank $m < s$ is given from the first m terms of the singular value decomposition (the **rank-m SVD**),

$$\hat{A}_{ij} = \sum_{k=1}^{m} \lambda_k \, u_{ik} \, v_{kj} \tag{A3.37a}$$

For example, the best rank-2 approximation is given by

$$A_{ij} \simeq \lambda_1 \, u_{i1} \, v_{j1} + \lambda_2 \, u_{i2} \, v_{j2} \tag{A3.37b}$$

where λ_i is the ith singular value of the $\mathbf{A}$ matrix, $\mathbf{u}$ and $\mathbf{v}$ are the associated singular vectors (see Example A3.8). The fraction of total variation of a matrix accounted for by taking the first m terms in its SVD is

$$\sum_{k=1}^{m} \lambda_k^2 / \sum_{ij} A_{ij}^2 = \frac{\lambda_1^2 + \cdots + \lambda_m^2}{\lambda_1^2 + \cdots + \lambda_s^2} \tag{A3.37c}$$

Finally, the SVD immediately allows one to compute a generalized inverse of $\mathbf{A}$. Using the compact SVD (Equation A3.36d), it follows that if $\mathbf{A}_{n\times p} = \mathbf{U}_{n\times s}\boldsymbol{\Lambda}_{s\times s}\mathbf{V}_{s\times p}^T$, then

$$\mathbf{A}^-_{p\times n} = \mathbf{V}_{p\times s}\boldsymbol{\Lambda}^{-1}_{s\times s}\mathbf{U}^T_{s\times n} \tag{A3.38}$$

To show this, we need to show that Equation A3.38 satisfies Equation A3.1, $\mathbf{A}\mathbf{A}^-\mathbf{A} = \mathbf{A}$. Rearranging and using the fact that $\mathbf{V}$ and $\mathbf{U}$ are unity matrices (their inverse is their transpose),

$$\begin{aligned}
\mathbf{A}\mathbf{A}^-\mathbf{A} &= \left[\mathbf{U}_{n\times s}\boldsymbol{\Lambda}_{s\times s}\mathbf{V}^T_{s\times p}\right]\left[\mathbf{V}_{p\times s}\boldsymbol{\Lambda}^{-1}_{s\times s}\mathbf{U}^T_{s\times n}\right]\left[\mathbf{U}_{n\times s}\boldsymbol{\Lambda}_{s\times s}\mathbf{V}^T_{s\times p}\right] \\
&= \mathbf{U}_{n\times s}\boldsymbol{\Lambda}_{s\times s}\left[\mathbf{V}^T_{s\times p}\mathbf{V}_{p\times s}\right]\boldsymbol{\Lambda}^{-1}_{s\times s}\left[\mathbf{U}^T_{s\times n}\mathbf{U}_{n\times s}\right]\boldsymbol{\Lambda}_{s\times s}\mathbf{V}^T_{s\times p} \\
&= \mathbf{U}_{n\times s}\boldsymbol{\Lambda}_{s\times s}\mathbf{I}_{s\times s}\boldsymbol{\Lambda}^{-1}_{s\times s}\mathbf{I}_{s\times s}\boldsymbol{\Lambda}_{s\times s}\mathbf{V}^T_{s\times p} \\
&= \mathbf{U}_{n\times s}\boldsymbol{\Lambda}_{s\times s}\boldsymbol{\Lambda}^{-1}_{s\times s}\boldsymbol{\Lambda}_{s\times s}\mathbf{V}^T_{s\times p} \\
&= \mathbf{U}_{n\times s}\boldsymbol{\Lambda}_{s\times s}\mathbf{V}^T_{s\times p} = \mathbf{A}
\end{aligned}$$

We can also use the SVD to solve much more general sets of linear equations. Until now, we have been assuming n equations and p unknowns, with $n = p$. More generally, we may wish to find solutions for the system

$$\mathbf{A}_{n\times p}\,\mathbf{x}_{p\times 1} = \mathbf{c}_{n\times 1} \tag{A3.39a}$$

When $n > p$, the system is said to be **overdetermined**, while it is said to be **underdetermed** when $n < p$. We can solve this system using the SVD as follows:

$$\begin{aligned}
\mathbf{A}_{n\times p}\,\mathbf{x}_{p\times 1} &= \mathbf{c}_{n\times 1} \\
\mathbf{U}_{n\times s}\boldsymbol{\Lambda}_{s\times s}\mathbf{V}^T_{s\times p}\,\mathbf{x}_{p\times 1} &= \mathbf{c}_{n\times 1} \\
\mathbf{V}_{p\times s}\boldsymbol{\Lambda}^{-1}_{s\times s}[\mathbf{U}^T_{s\times n}\mathbf{U}_{n\times s}]\boldsymbol{\Lambda}_{s\times s}\mathbf{V}^T_{s\times p}\,\mathbf{x}_{p\times 1} &= \mathbf{V}_{p\times s}\boldsymbol{\Lambda}^{-1}_{s\times s}\mathbf{U}^T_{s\times n}\mathbf{c}_{n\times 1} \\
\mathbf{V}_{p\times s}[\boldsymbol{\Lambda}^{-1}_{s\times s}\boldsymbol{\Lambda}_{s\times s}]\mathbf{V}^T_{s\times p}\,\mathbf{x}_{p\times 1} &= \mathbf{V}_{p\times s}\boldsymbol{\Lambda}^{-1}_{s\times s}\mathbf{U}^T_{s\times n}\mathbf{c}_{n\times 1} \\
[\mathbf{V}_{p\times s}\mathbf{V}^T_{s\times p}]\,\mathbf{x}_{p\times 1} &= \mathbf{V}_{p\times s}\boldsymbol{\Lambda}^{-1}_{s\times s}\mathbf{U}^T_{s\times n}\mathbf{c}_{n\times 1} \\
\mathbf{x}_{p\times 1} &= \mathbf{V}_{p\times s}\boldsymbol{\Lambda}^{-1}_{s\times s}\mathbf{U}^T_{s\times n}\mathbf{c}_{n\times 1}
\end{aligned} \tag{A3.39b}$$

where each term above in brackets is an identity matrix.

Finally, we can connect the SVD with diagonalization as follows. Consider an $n \times p$ data matrix, $\mathbf{X}$, where $\mathbf{X}$ is of rank $s \leq p < n$. The covariance matrix for this data has the form of $\mathbf{X}^T\mathbf{X}$. Using the SVD of $\mathbf{X}$, we have

$$\begin{aligned}\mathbf{X}^T\mathbf{X} &= [\mathbf{U}_{n\times s}\boldsymbol{\Lambda}_{s\times s}\mathbf{V}_{s\times p}^T]^T[\mathbf{U}_{n\times s}\boldsymbol{\Lambda}_{s\times s}\mathbf{V}_{s\times p}^T] \\ &= \mathbf{V}_{p\times s}\boldsymbol{\Lambda}_{s\times s}[\mathbf{U}_{s\times n}^T\mathbf{U}_{n\times s}]\boldsymbol{\Lambda}_{s\times s}\mathbf{V}_{s\times p}^T \\ &= \mathbf{V}_{p\times s}\boldsymbol{\Lambda}_{s\times s}\boldsymbol{\Lambda}_{s\times s}\mathbf{V}_{s\times p}^T \\ &= \mathbf{V}_{p\times s}\boldsymbol{\Lambda}_{s\times s}^2\mathbf{V}_{s\times p}^T\end{aligned}$$

where we have used $\boldsymbol{\Lambda}_{s\times s}^T = \boldsymbol{\Lambda}_{s\times s}$, as $\boldsymbol{\Lambda}_{s\times s}$ is diagonal. Comparison to Equation A3.23 shows that the elements of $\mathbf{V}$ correspond to the eigenvectors of the covariance matrix, while the eigenvalues are given by the squares of the singular values, $\lambda_i = s_i^2$.

Example A3.8. Consider the follow performances of three soybean lines in five different environments in New York from a larger dataset given by Gauch (1992). Each entry represents the average of four replications. For environments, the first letter denotes the testing location, while the two numbers denote the year of the test. For ease of presentation, we have rounded entry means and the associated row and column means.

	Genotypes				
Environment	EVAN	WILK	CHIP	row mean	$\widehat{E}$
A77	2725	2471	2333	2510	494
V79	1111	578	1278	989	-1027
R81	2038	1386	2350	1925	-91
I85	1736	1607	1588	1644	-372
G85	3258	2961	2813	3011	995
column mean	2174	1801	2072	$\vert\overline{2016}$	
$\widehat{G}$	158	-215	56		

Denote the entry means of genotype i in environment j by z_{ij}, the environment and genotype means by $z_{\cdot j}$ and $z_{i\cdot}$ and the grand mean by $z_{\cdot\cdot}$ (= 2016). Under an OLS fixed-effects model, the estimated main effects are given by $\widehat{\mu} = z_{\cdot\cdot}$, $\widehat{E}_j = z_{\cdot j} - z_{\cdot\cdot}$, and $\widehat{G}_i = z_{i\cdot} - z_{\cdot\cdot}$, and these are given in the above table. The interaction terms are estimated by

$$\widehat{GE}_{ij} = z_{ij} - (\widehat{\mu} + \widehat{G_i} + \widehat{E_j})$$

For example, for EVAN in A77, $\widehat{GE} = 2725 - (2016 + 158 + 494) = 57$. The resulting GE table of interactions becomes

	Genotypes		
Environment	EVAN	WILK	CHIP
A77	57	176	-233
V79	-36	-196	233
R81	-45	-324	369
I85	-66	178	-112
G85	89	165	-254

Now let's examine singular value decomposition (SVD) approximations for this table of GE values, which writen in matrix form (using columns equaling genotypes and rows environments) as

$$\mathbf{GE} = \begin{pmatrix} 57 & 176 & -233 \\ -36 & -196 & 233 \\ -45 & -324 & 369 \\ -66 & 178 & -112 \\ 89 & 165 & -254 \end{pmatrix}$$

In R, the compact SVD (Equation A3.36d) of a matrix X is given by the command `svd(X)`, returning the compact SVD of **GE** as

$$\begin{pmatrix} 0.40 & 0.21 & 0.18 \\ -0.41 & 0.00 & 0.91 \\ -0.66 & 0.12 & -0.30 \\ 0.26 & -0.83 & 0.11 \\ 0.41 & 0.50 & 0.19 \end{pmatrix} \begin{pmatrix} 746.10 & 0 & 0 \\ 0 & 131.36 & 0 \\ 0 & 0 & 0.53 \end{pmatrix} \begin{pmatrix} 0.12 & 0.64 & -0.76 \\ 0.81 & -0.51 & -0.30 \\ 0.58 & 0.58 & 0.58 \end{pmatrix}$$

The first singular value accounts for $746.10^2/(743.26^2 + 131.36^2 + 0.53^2)$ = 97.0% of the total variation of **GE**, while the second singular value accounts for 3.0%, so that together they account for essentially all of the total variation. The rank-1 SVD approximation of **GE** is given by setting all of the diagonal elements of $\boldsymbol{\Lambda}$ except the first entry to zero,

$$\mathbf{GE}_1 = \begin{pmatrix} 0.40 & 0.21 & 0.18 \\ -0.41 & 0.00 & 0.91 \\ -0.66 & 0.12 & -0.30 \\ 0.26 & -0.83 & 0.11 \\ 0.41 & 0.50 & 0.19 \end{pmatrix} \begin{pmatrix} 746.10 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0.12 & 0.64 & -0.76 \\ 0.81 & -0.51 & -0.30 \\ 0.58 & 0.58 & 0.58 \end{pmatrix}$$

Similarly, the rank-2 SVD is given by setting all but the first two singular values to zero,

$$\mathbf{GE}_2 = \begin{pmatrix} 0.40 & 0.21 & 0.18 \\ -0.41 & 0.00 & 0.91 \\ -0.66 & 0.12 & -0.30 \\ 0.26 & -0.83 & 0.11 \\ 0.41 & 0.50 & 0.19 \end{pmatrix} \begin{pmatrix} 746.10 & 0 & 0 \\ 0 & 131.36 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0.12 & 0.64 & -0.76 \\ 0.81 & -0.51 & -0.30 \\ 0.58 & 0.58 & 0.58 \end{pmatrix}$$

The resulting rank-1 and rank-2 approximations for **GE** are

$$\mathbf{GE}_1 = \begin{pmatrix} 34.57 & 189.94 & -224.67 \\ -35.78 & -196.59 & 232.54 \\ -57.51 & -316.03 & 373.81 \\ 22.33 & 122.71 & -145.15 \\ 36.04 & 198.02 & -234.23 \end{pmatrix}, \quad \mathbf{GE}_2 = \begin{pmatrix} 56.95 & 175.95 & -233.05 \\ -36.28 & -196.28 & 232.72 \\ -44.91 & -323.91 & 369.09 \\ -66.03 & 177.97 & -112.03 \\ 88.94 & 164.94 & -254.06 \end{pmatrix}$$

Since there are only three non-zero singular values, the rank-3 SVD returns the matrix **GE**, i.e., $\mathbf{GE}_3 = \mathbf{GE}$. For example, consider the G x E term for EVAN and A77. The rank-1 SVD value for this entry is 34.57, the rank-2 VSD value is 56.95, and the rank-3 (full rank) just the original value of 57.

Example A3.9. From Equation A3.38, a generalized inverse of **GE** is given by

$$\begin{pmatrix} 0.12 & 0.81 & 0.58 \\ 0.64 & -0.51 & 0.58 \\ -0.76 & -0.30 & 0.58 \end{pmatrix} \begin{pmatrix} \frac{1}{746.10} & 0 & 0 \\ 0 & \frac{1}{131.36} & 0 \\ 0 & 0 & \frac{1}{0.53} \end{pmatrix} \begin{pmatrix} 0.40 & -0.41 & -0.66 & 0.26 & 0.41 \\ 0.21 & 0.00 & 0.12 & -0.83 & 0.50 \\ 0.18 & 0.91 & -0.30 & 0.11 & 0.19 \end{pmatrix}$$

$$= \begin{pmatrix} 0.1989 & 1 & -0.3253 & 0.1170 & 0.2106 \\ 0.1971 & 1 & -0.3271 & 0.1256 & 0.2060 \\ 0.1965 & 1 & -0.3254 & 0.1237 & 0.2058 \end{pmatrix}$$

Denoting this matrix as $\mathbf{GE}^-$, matrix multiplication shows that Equation A3.1 is satisfied, namely that $\mathbf{GE}\,\mathbf{GE}^-\,\mathbf{GE} = \mathbf{GE}$, showing that $\mathbf{GE}^-$ is indeed a generalized inverse of **GE**.

Literature Cited

Dhrymes, P. J. 1978. *Mathematics for econometrics.* Springer-Verlag, New York, NY. [A3]

Ekart, C., and G. Young G. 1936. The approximation of one matrix by another of lower rank. *Psychometrika* 1: 211-218. [A3]

Gauch, H. G. Jr. 1992. *Statistical analysis of regional yield trials: AMMI analysis of factorial designs*. Elsevier. Amsterdam, NL [A3]

Graham, A. 1981. *Kronecker products and matrix calculus with applications*. Halsted Press, New York, NY. [A3]

Henderson, C. R. 1977b. Best linear unbiased prediction of breeding values not in the model for records. *J. Dairy Sci.* 60: 783–787. [A3]

Henderson, C. R. 1984a. *Applications of linear models in animal breeding*. Univ. Guelph, Guelph, ON. [A3]

Henderson, C. R. 1985c. Equivalent linear models to reduce computations. *J. Dairy Sci.* 68: 2367–2377. [A3]

Morrison, D. F. 1976. *Multivariate statistical methods.* McGraw-Hill, New York, NY. [A3]

Pringle, R. M., and A. A. Rayner. 1971. *Generalized inverse matrices with applications to statistics.* Griffin, London, UK. [A3]

Quaas, R. L., and E. J. Pollak. 1980. Mixed model methodology for farm and ranch beef cattle testing programs. *J. Anim. Sci.* 51: 1277–1287. [A3]

Rao, C. R., and S. K. Mitra. 1971. *Generalized inverse of matrices and its applications.* John Wiley & Sons, New York, NY. [A3]

Searle, S. R. 1971. *Linear models.* John Wiley & Sons, New York, NY. [A3]

Searle, S. R. 1982. *Matrix algebra useful for statistics.* John Wiley and Sons, New York, NY. [A3]