29

Measuring Multivariate Selection

We have found that there are fundamental differences between the surviving birds and those eliminated, and we conclude that the birds which survived survived because they possessed certain structural characters, and that the birds which perished perished not through accident, but because they did not possess certain structural characters which would have enabled them to withstand the severity of the test imposed by nature; they were eliminated because they were unfit.
— Bumpus (1899)

Draft version 20 September 2010

While we have previously hinted at some features of selection on multiple traits (Chapters 10, 16), we now start our formal discussion of selection on a vector of traits. As in Chapter 28, we distinguish between *selection*, the within-generation change in the (fitness-weighted) distribution of trait values from the *response to selection*, how the distribution of trait values change across generations. As mentioned in Chapters 10 and 16, *phenotypic* correlations between traits within an individual influence within-generation changes, while *genetic* correlations (correlations among the breeding values for different traits within an individual) influence the between-generation response (Chapters 30-33). In particular, when a suite of traits are phenotypically-correlated, simply observing the within-generation change in their means or variances is *not* sufficient to determine which trait(s) selection was acting upon. One must remove the effects of phenotypic correlations to seperate **direct** from **correlated** selection. This is done using partial regressions (linear for changes in the means, quadratic for changes in variances and covariances), and much of our focus here is on such fitness-trait regressions. Essentially all the measures of phenotypic selection discussed here are multivariate extensions of measures introduced in Chapter 28. As these expressions are in terms of matrices and vectors, we rely rather heavily in places on matrix machinery (such as eigenvalues/vectors and canonical decompositions and, in a few places, matrix calculus). Appendix 4 discusses the mathematics of treating matrices as geometric objects, while Appendix 5 provides a brief refresher of important concepts and tools from the calculus of matrices (such as multidimensional derivatives and Taylor series). The reader may wish to review these before venturing into a few of the more technical sections covered in this (and subsequent) chapters.

The structure of this chapter is as follows: We first introduce the multivariate versions of differentials and gradients and their properties. Next, the geometry of quadratic regressions is examined in some detail, followed by discussions of multivariate nonparametric regressions and the strength of selection in natural populations. We conclude with some comments on unmeasured characters and the use of path analysis as an alternative (and often complementary) approach to the analysis of selection.

SELECTION ON MULTIVARIATE PHENOTYPES:
DIFFERENTIALS AND GRADIENTS

Chapter 28 described a variety of measures of univariate selection, with an emphasis on approximating the individual fitness function. In order to extend these methods to a vector of characters, we need to account for phenotypic correlations. To do so, we follow the multiple

regression approach of Lande and Arnold (1983), initially suggested by Pearson (1903). Similar approaches based on path analysis (LW Appendix 2) have also been suggested, and these offer a complementary method of analysis (as discussed at the end of this chapter).

The phenotype of an individual is now a vector $\mathbf{z} = (z_1, z_2, \cdots, z_n)^T$ of n character values. Denote the mean vector and covariance matrix of $\mathbf{z}$ before selection by $\boldsymbol{\mu}$ and $\mathbf{P}$, and by $\boldsymbol{\mu}^*$ and $\mathbf{P}^*$ after selection (but before reproduction). To avoid additional complications, we examine only a single episode of selection. Partitions over multiple episodes follow as fairly straightforward extensions of the univariate partitions discussed in Chapter 28.

Changes in the Mean Vector: The Directional Selection Differential S

The multivariate extension of the **directional selection differential** is the vector

$$\mathbf{S} = \boldsymbol{\mu}^* - \boldsymbol{\mu}$$

whose ith element is $S_i = \mu_i^* - \mu_i$, the differential for character z_i. (The fastidious reader might object to the nonstandard use of a capital bold letter, as opposed to the more standard bold lowercase letter, for a vector, but univariate selection theory uses S for the selection differential and we keep that notation here.) As with the univariate case, the Robertson-Price identity (Equation 10.7) holds, so that the elements of $\mathbf{S}$ represent the covariance between character value and relative fitness, $S_i = \sigma(z_i, w)$. This immediately implies (Equation 28.6) that the opportunity for selection I (the population variance in relative fitness) bounds the range of S_i,

$$\frac{|S_i|}{\sigma_{z_i}} \leq \sqrt{I} \tag{29.1}$$
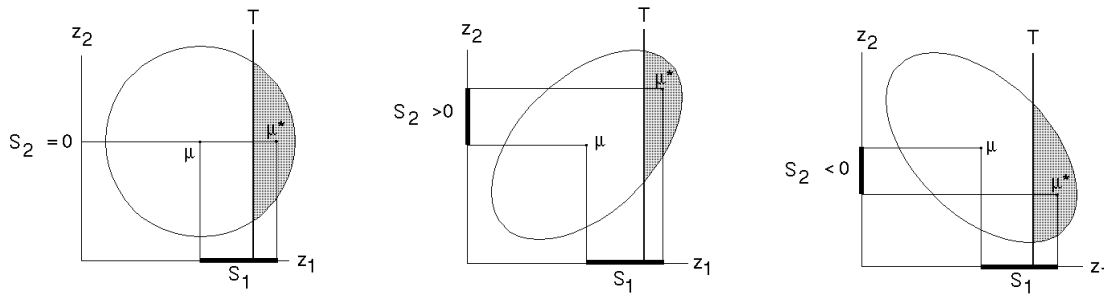

Figure 29.1.   Selection on a character can result in a within-generation change in the mean of other phenotypically correlated characters not themselves under direct selection. Assume that character 1 is under simple truncation selection so that only individuals with $z_1 > T$ reproduce. **Left:** When z_1 and z_2 are uncorrelated, $S_2 = 0$. **Center**: When z_1 and z_2 are positively correlated, $S_2 > 0$. **Right**: When z_1 and z_2 are negatively correlated, $S_2 < 0$.

As is illustrated in Figure 29.1, $\mathbf{S}$ confounds the *direct* effects of selection on a focal trait with the *indirect* effects due to selection on phenotypically correlated characters. Suppose character 1 is under direct selection to increase in value while character 2 is not directly selected. As Figure 29.1 shows, if z_1 and z_2 are uncorrelated, there is no within-generation change in μ_2 (the mean of z_2). However, if z_1 and z_2 are positively correlated, individuals with large values of z_1 also tend to have large values of z_2, resulting in a within-generation increase in μ_2. Conversely, if z_1 and z_2 are negatively correlated, selection to increase z_1 results in a within-generation decrease in μ_2. Hence, a character not under selection can

still experience a within-generation change due to selection on a phenotypically correlated character (**indirect** or **correlated** selection). Fortunately, the **directional selection gradient** $\boldsymbol{\beta} = \mathbf{P}^{-1}\mathbf{S}$ accounts for indirect selection due to phenotypic correlations, providing a less biased picture of the nature of directional selection acting on the component traits comprising **z**.

The Directional Selection Gradient $\boldsymbol{\beta}$

As was discussed (briefly) in Chapters 10 and 16, $\boldsymbol{\beta}$ removes the effects of phenotypic correlations because it is a vector of partial regression coefficients. From multiple regression theory (LW Chapter 8), the vector of partial regression coefficients for predicting the value of y given a vector of observations $\mathbf{z}$ is $\mathbf{P}^{-1}\boldsymbol{\sigma}(\mathbf{z}, y)$, where $\mathbf{P}$ is the covariance matrix of $\mathbf{z}$ and $\boldsymbol{\sigma}(\mathbf{z}, y)$ is the vector of covariances between the elements of $\mathbf{z}$ and the variable y. Since $\mathbf{S} = \boldsymbol{\sigma}(\mathbf{z}, w)$, it immediately follows that

$$\mathbf{P}^{-1}\boldsymbol{\sigma}(\mathbf{z}, w) = \mathbf{P}^{-1}\mathbf{S} = \boldsymbol{\beta} \tag{29.2}$$

is the vector of partial regression for the best linear regression of *relative* fitness w on phenotypic value $\mathbf{z}$, viz.,

$$w(\mathbf{z}) = 1 + \sum_{j=1}^{n} \beta_j(z_j - \mu_j) + e = 1 + \boldsymbol{\beta}^T(\mathbf{z} - \boldsymbol{\mu}) + e \tag{29.3}$$

Note that we have use the fact (LW Chapters 3, 8) that the regression must pass through the means of both w (1) and $\mathbf{z}$ ($\boldsymbol{\mu}$) to remove the intercept constant in the regression. We could equally write Equation 29.3 as $w = a + \boldsymbol{\beta}^T\mathbf{z} + e$.

Since β_j is a partial regression coefficient, it represents the change generated in relative fitness by changing z_j while holding all other character values in $\mathbf{z}$ constant — a one unit increase in z_j (holding all other characters constant) increases the expected relative fitness by β_j. When we can exclude the possibility that fitness is influenced by unmeasured characters phenotypically correlated with $\mathbf{z}$, a character under no directional selection has $\beta_j = 0$. When all other characters are held constant, the best linear regression predicts no change in expected fitness as we change the value of z_j. Thus, $\boldsymbol{\beta}$ accounts for the effects of phenotypic correlations only among the **measured** characters. Since $\mathbf{S} = \mathbf{P}\boldsymbol{\beta}$, we have

$$S_i = \sum_{j=1}^{n} \beta_j\, P_{ij} = \beta_i\, P_{ii} + \sum_{j \neq i}^{n} \beta_j\, P_{ij} \tag{29.4}$$

illustrating that the directional selection differential confounds direct selection on that character (β_i) with indirect contributions due to selection on phenotypically correlated characters ($\beta_j P_{ij}$). Equation 29.4 implies that if two characters are phenotypically uncorrelated ($P_{ij} = 0$), selection on one has no within-generation effect on the phenotypic mean of the other.

Operationally, fitness regressions are usually computed using **standardized variables**,

$$z_i' = \frac{z_i - \mu_i}{\sigma_i}$$

so that z_i' has mean zero and unit variance. The selection differential of the standardized variable is just the selection intensity $\overline{\imath}$ (Equation 10.6a), as

$$S_i' = \frac{S_i}{\sigma_i} = \overline{\imath}_i$$

Likewise the phenotypic covariance matrix $\mathbf{P}'$ for the standardized variables is simply the matrix of all pairwise correlations. Selection gradients using standardized variables are generally indicated by β'_i. With standardized variables, Equation 29.3 becomes

$$w(\mathbf{z}) = 1 + \left(\boldsymbol{\beta}'\right)^T \mathbf{z}' + e, \quad \text{where} \quad \boldsymbol{\beta}' = \left(\mathbf{P}'\right)^{-1} \mathbf{S}'$$

Example 29.1. The original application of the Lande-Arnold regression was on a population of pentatomid bugs (*Euschistus variolarius*), collected along the shore of Lake Michigan after a storm (Lande and Arnold 1983). Of the 94 individuals that were collected (legend has it that they were deposited into an open adult beverage container, a truly selfless act of science), 39 were alive. All individuals were measured for four characters: head (Hd) and thorax (Tx) width, and scutellum (Sc) and forewing (Fw) length. The data were then logarithmically transformed to more closely approximate normality. The transformed variables were then standarized. Selection differentials were calculated as the mean in each (transformed and standardized) trait among those bugs that survived minus the mean for all bugs (dead or alive). Since these are presented in terms of standard deviations, these are the selection intensities ($\bar{\imath}$). The resulting vector of standardized selection differentials and correlation matrix were as follows:

$$\mathbf{S}' = \begin{pmatrix} \bar{\imath}_{Hd} \\ \bar{\imath}_{Tx} \\ \bar{\imath}_{Sc} \\ \bar{\imath}_{Fw} \end{pmatrix} = \begin{pmatrix} -0.11 \\ -0.06 \\ -0.28^* \\ -0.43^{**} \end{pmatrix}, \quad \mathbf{P}' = \begin{pmatrix} 1.00 & 0.72 & 0.50 & 0.60 \\ 0.72 & 1.00 & 0.59 & 0.71 \\ 0.50 & 0.59 & 1.00 & 0.62 \\ 0.60 & 0.71 & 0.62 & 1.00 \end{pmatrix}$$

where $*$ and $**$ denote five and one percent significance for elements in $\mathbf{S}'$. Hence,

$$\boldsymbol{\beta}' = \left(\mathbf{P}'\right)^{-1} \mathbf{S}' = \begin{pmatrix} 1.00 & 0.72 & 0.50 & 0.60 \\ 0.72 & 1.00 & 0.59 & 0.71 \\ 0.50 & 0.59 & 1.00 & 0.62 \\ 0.60 & 0.71 & 0.62 & 1.00 \end{pmatrix}^{-1} \begin{pmatrix} -0.11 \\ -0.06 \\ -0.28 \\ -0.43 \end{pmatrix} = \begin{pmatrix} 0.02 \\ 0.53^{**} \\ -0.16 \\ -0.72^{**} \end{pmatrix}$$

The (standarized) selection differentials for scutellum and wing were both significantly different from zero, while the (standarized) gradients for thorax and wing were also significant. Thus, while the within-generation change in thorax length was negative (but not significantly different from zero), in reality there was strong direct selection to *increase* thorax length. A one standard-deviation increase in thorax length increases relative fitness by $\beta' = 0.53$, while a one standard deviation increase in wing length reduces fitness by 0.72.

Gradients can be standardized in other ways besides using the trait variance. Hereford et al. (2004) suggest that mean-standardized gradients,

$$\beta_\mu = \mu_z \beta$$

have certain advantages. For example, $\beta_\mu = 1$ when the trait is fitness, providing a very natural measure for comparing the relative strength of selection. A second technical advantage (discussed in detail in Chapter 36) is that mean-standardized gradients represent the increase in relative fitness for a *proportional* change in a trait, and thus are equivalent to the **elasticity** coefficients used in demography (van Tienderen 2000). Mean-standardization

can only occur for traits where the origin of the scale is not arbitrary. Thus, for traits with a biologically-motivated zero value, β_μ can be used. However, this restriction results in many interesting traits such as seasonal timing (the origin of the calendar is arbitrary) and composite traits (such as principal components) being excluded as candidates for mean-standardization (Kingsolver and Pfenning 2007).

Directional Gradients, Fitness Surface Geometry and Selection Response

When phenotypes are multivariate normal, $\boldsymbol{\beta}$ provides a convenient descriptor of the geometry of both the individual $w(z)$ and mean population fitness $\overline{W}$ surfaces. Consider a function $f(\mathbf{x})$ that returns a scalar, but whose argument is a vector $\mathbf{x} = (x_1, \cdots, x_n)^T$. Recall from vector calculus (Appendix 5) that the **gradient vector** $\nabla_{\mathbf{x}} f(\mathbf{x})$ of f with respect to $\mathbf{x}$ is defined as

$$\nabla_{\mathbf{x}} f(\mathbf{x}) = \begin{pmatrix} \partial f/\partial x_1 \\ \partial f/\partial x_2 \\ \vdots \\ \partial f/\partial x_n \end{pmatrix}$$

namely the vector of each partial with respect to f. Further recall that the gradient vector of a function points in the direction of change in the vector that will give the greatest (local) increase in f. Assuming $\mathbf{z}$ is MVN, we show (Example A5.3) that

$$\boldsymbol{\beta} = \nabla_{\boldsymbol{\mu}}[\ln \overline{W}(\boldsymbol{\mu})] = \overline{W}^{-1} \cdot \nabla_{\boldsymbol{\mu}}[\overline{W}(\boldsymbol{\mu})] \tag{29.5a}$$

which holds provided fitnesses are frequency-independent (Lande 1976, 1979). In this case $\boldsymbol{\beta}$ is the gradient of *mean population* fitness with respect to the mean vector $\boldsymbol{\mu}$. Since $\boldsymbol{\beta}$ gives the direction of steepest increase in the mean population fitness surface, $\overline{W}$ increases most rapidly when the change in the vector of means $\boldsymbol{\Delta\mu}$ is in the direction given by $\boldsymbol{\beta}$. If fitnesses are frequency-dependent (individual fitnesses change as the population mean changes, so that $\nabla_{\boldsymbol{\mu}}[w(\mathbf{z})] \neq \mathbf{0}$), then provided $\mathbf{z}$ is multivariate-normally distributed,

$$\boldsymbol{\beta} = \nabla_{\boldsymbol{\mu}}[\ln \overline{W}(\boldsymbol{\mu})] + \int \nabla_{\boldsymbol{\mu}}[w(\mathbf{z})]\, \varphi(\mathbf{z})\, \mathrm{d}\mathbf{z} \tag{29.5b}$$

where the second term accounts for the effects of frequency-dependence and φ is the MVN density function (Lande 1976). In this case $\boldsymbol{\beta}$ does not point in the direction of steepest increase in $\overline{W}$ unless the second integral is zero.

If we instead consider the **individual fitness** surface, we can alternatively consider $\boldsymbol{\beta}$ as the gradient of **individual fitnesses** *averaged* over the population distribution of phenotypes,

$$\boldsymbol{\beta} = \int \nabla_{\mathbf{z}}[w(\mathbf{z})]\, \varphi(\mathbf{z})\, \mathrm{d}\mathbf{z} \tag{29.6}$$

which holds provided $\mathbf{z} \sim$ MVN (Lande and Arnold 1983). To see this, using integration by parts gives

$$\int_{\mathbf{a}}^{\mathbf{b}} \nabla_{\mathbf{z}}[w(\mathbf{z})]\, \varphi(\mathbf{z})\, \mathrm{d}\mathbf{z} = w(\mathbf{z})\, \varphi(\mathbf{z}) \Big|_{\mathbf{a}}^{\mathbf{b}} - \int_{\mathbf{a}}^{\mathbf{b}} \nabla_{\mathbf{z}}[\varphi(\mathbf{z})]\, w(\mathbf{z})\, \mathrm{d}\mathbf{z}$$

Taking the limit as $\mathbf{a} \to -\infty$ and $\mathbf{b} \to \infty$, the first term on the right-hand side vanishes as $\varphi(\mathbf{z}) \to 0$ when $\mathbf{z} \to \pm\infty$ (as the phenotypic distribution is assumed to be bounded). If $\mathbf{z} \sim$

$MVN(\boldsymbol{\mu}, \mathbf{P})$, then Equation A5.2a gives $\nabla_{\mathbf{z}}[\varphi(\mathbf{z})] = -\varphi(\mathbf{z})\mathbf{P}^{-1}(\mathbf{z} - \boldsymbol{\mu})$, implying

$$\int \nabla_{\mathbf{z}}[w(\mathbf{z})]\,\varphi(\mathbf{z})\,d\mathbf{z} = -\int \nabla_{\mathbf{z}}[\varphi(\mathbf{z})]\,w(\mathbf{z})\,d\mathbf{z} = \int w(\mathbf{z})\,\varphi(\mathbf{z})\,\mathbf{P}^{-1}(\mathbf{z} - \boldsymbol{\mu})\,d\mathbf{z}$$

$$= \mathbf{P}^{-1}\left(\int \mathbf{z}\,w(\mathbf{z})\,\varphi(\mathbf{z})\,d\mathbf{z} - \boldsymbol{\mu}\int w(\mathbf{z})\,\varphi(\mathbf{z})\,d\mathbf{z}\right)$$

$$= \mathbf{P}^{-1}(\boldsymbol{\mu}^* - \boldsymbol{\mu}) = \mathbf{P}^{-1}\mathbf{S} = \boldsymbol{\beta}$$

The first integral on line two is the mean trait value weighted by relative fitness (μ^*) and the second interval is just the average value of relative fitness (1). Note from this derivation that Equation 29.6 holds regardless of whether fitness is frequency dependent or frequency independent. This follows from the gradient being taken with respect to $\mathbf{x}$, rather than with respect to the vector of means $\boldsymbol{\mu}$.

Finally, while our focus has been on the role $\boldsymbol{\beta}$ plays in measuring phenotypic selection, it also plays an important role in the response to selection. If we can assume the breeders' equation holds, $\boldsymbol{\beta}$ is the only measure of phenotypic selection required to predict the response in the means, as the vector $\mathbf{R}$ of response (changes in means) is given by $\mathbf{R} = \mathbf{G}\boldsymbol{\beta}$. Cheverud (1984) makes the important point that although it is often assumed a set of phenotypically correlated traits responses to selection in a coordinated fashion, this is not necessarily the case. Since $\boldsymbol{\beta}$ removes the effects of phenotypic correlations, phenotypic characters will only respond as a group if they are all under direct selection or if they are *genetically* correlated, a point we discuss in detail in the next several chapters.

Changes in the Covariance Matrix: The Quadratic Selection Differential C

Motivated by the univariate case wherein $C = \sigma[w, (z-\mu)(z-\mu)]$, we define the multivariate **quadratic selection differential** to be a square ($n \times n$) matrix $\mathbf{C}$ whose elements are the covariances between all pairs of quadratic deviations $(z_i - \mu_i)(z_j - \mu_j)$ and relative fitness w, viz.,

$$C_{ij} = \sigma[w, (z_i - \mu_i)(z_j - \mu_j)] \tag{29.7a}$$

As is derived below (Example 29.2), Lande and Arnold (1983) showed that

$$\mathbf{C} = \boldsymbol{\sigma}[w, (\mathbf{z} - \boldsymbol{\mu})(\mathbf{z} - \boldsymbol{\mu})^T] = \mathbf{P}^* - \mathbf{P} + \mathbf{S}\mathbf{S}^T \tag{29.7b}$$

If no quadratic selection is acting, the covariance between each quadratic deviation and fitness is zero and $\mathbf{C} = \mathbf{0}$. In this case Equation 29.7b gives

$$P_{ij}^* - P_{ij} = -S_i S_j \tag{29.8}$$

demonstrating that the $S_i S_j$ term corrects C_{ij} for the change in covariance caused by directional selection alone.

Example 29.2. We wish to show $\mathbf{P}^* - \mathbf{P} = \boldsymbol{\sigma}[w, (\mathbf{z} - \boldsymbol{\mu})(\mathbf{z} - \boldsymbol{\mu})^T] - \mathbf{S}\mathbf{S}^T$, which implies Equation 29.7b. From the definition of the variance-covariance matrix,

$$\mathbf{P} = E\left[(\mathbf{z} - \boldsymbol{\mu})(\mathbf{z} - \boldsymbol{\mu})^T\right] = \int (\mathbf{z} - \boldsymbol{\mu})(\mathbf{z} - \boldsymbol{\mu})^T\,p(\mathbf{z})\,d\mathbf{z}$$

$$\mathbf{P}^* = E\left[(\mathbf{z}^* - \boldsymbol{\mu}^*)(\mathbf{z}^* - \boldsymbol{\mu}^*)^T\right] = \int (\mathbf{z} - \boldsymbol{\mu}^*)(\mathbf{z} - \boldsymbol{\mu}^*)^T\,p^*(\mathbf{z})\,d\mathbf{z}$$

where $p^*(\mathbf{z}) = w(\mathbf{z}) \, p(\mathbf{z})$ is the distribution of $\mathbf{z}$ after selection (but before reproduction). Noting that $\boldsymbol{\mu}^* = \boldsymbol{\mu} + \mathbf{S}$,

$$
\begin{aligned}
(\mathbf{z} - \boldsymbol{\mu}^*)(\mathbf{z} - \boldsymbol{\mu}^*)^T &= (\mathbf{z} - \boldsymbol{\mu} - \mathbf{S})(\mathbf{z} - \boldsymbol{\mu} - \mathbf{S})^T \\
&= (\mathbf{z} - \boldsymbol{\mu} - \mathbf{S})((\mathbf{z} - \boldsymbol{\mu})^T - \mathbf{S}^T) \\
&= (\mathbf{z} - \boldsymbol{\mu})(\mathbf{z} - \boldsymbol{\mu})^T - (\mathbf{z} - \boldsymbol{\mu})\mathbf{S}^T - \mathbf{S}(\mathbf{z} - \boldsymbol{\mu})^T + \mathbf{S}\,\mathbf{S}^T
\end{aligned}
$$

Since $\int \mathbf{z} \, p^*(\mathbf{z}) \, \mathrm{d}\mathbf{z} = \boldsymbol{\mu}^*$ and $\int p^*(\mathbf{z}) \, \mathrm{d}\mathbf{z} = 1$, we have that $\int \mathbf{S}\,\mathbf{S}^T \, p^*(\mathbf{z}) \, \mathrm{d}\mathbf{z} = \mathbf{S}\,\mathbf{S}^T$,

$$
\int (\mathbf{z} - \boldsymbol{\mu})\,\mathbf{S}^T \, p^*(\mathbf{z}) \, \mathrm{d}\mathbf{z} = (\boldsymbol{\mu}^* - \boldsymbol{\mu})\mathbf{S}^T = \mathbf{S}\mathbf{S}^T, \quad \text{and}
$$

$$
\int \mathbf{S}(\mathbf{z} - \boldsymbol{\mu})^T \, p^*(\mathbf{z}) \, \mathrm{d}\mathbf{z} = \mathbf{S}(\boldsymbol{\mu}^*)^T - \mathbf{S}\boldsymbol{\mu}^T = \mathbf{S}(\boldsymbol{\mu}^* - \boldsymbol{\mu})^T = \mathbf{S}\mathbf{S}^T
$$

Combining these results,

$$
\begin{aligned}
\mathbf{P}^* &= \int (\mathbf{z} - \boldsymbol{\mu})(\mathbf{z} - \boldsymbol{\mu})^T \, w(\mathbf{z}) \, p(\mathbf{z}) \, \mathrm{d}\mathbf{z} - \mathbf{S}\mathbf{S}^T - \mathbf{S}\mathbf{S}^T + \mathbf{S}\mathbf{S}^T \\
&= E\left[w(\mathbf{z}) \cdot (\mathbf{z} - \boldsymbol{\mu})(\mathbf{z} - \boldsymbol{\mu})^T \right] - \mathbf{S}\mathbf{S}^T
\end{aligned}
$$

Since $E[\,w(\mathbf{z})\,] = 1$, we can write $\mathbf{P} = E[\,w(\mathbf{z})\,] \cdot \mathbf{P}$. Using the definition of $\mathbf{P}$ then gives

$$
\begin{aligned}
\mathbf{P}^* - \mathbf{P} &= E\left[w(\mathbf{z}) \cdot (\mathbf{z} - \boldsymbol{\mu})(\mathbf{z} - \boldsymbol{\mu})^T \right] - \mathbf{S}\mathbf{S}^T - E[\,w(\mathbf{z})\,] \cdot E\left[(\mathbf{z} - \boldsymbol{\mu})(\mathbf{z} - \boldsymbol{\mu})^T \right] \\
&= \boldsymbol{\sigma}\left[w(\mathbf{z}), (\mathbf{z} - \boldsymbol{\mu})(\mathbf{z} - \boldsymbol{\mu})^T \right] - \mathbf{S}\mathbf{S}^T
\end{aligned}
$$

with the last equality following from $\sigma(x, y) = E(x \cdot y) - E(x)\,E(y)$.

As was the case for $\mathbf{S}$, the fact that C_{ij} is a covariance immediately allows us to bound its range using the opportunity for selection (Chapter 28). Since $\sigma^2(x, y) \leq \sigma^2(x)\,\sigma^2(y)$,

$$
C_{ij}^2 \leq I \, \sigma^2\left[(z_i - \mu_i)(z_j - \mu_j) \right] \tag{29.9a}
$$

When z_i and z_j are bivariate normal, then (Kendall and Stuart 1983),

$$
\sigma^2\left[(z_i - \mu_i)(z_j - \mu_j) \right] = P_{ij}^2 + P_{ii}\,P_{jj} = P_{ij}^2(1 + \rho_{ij}^{-2}) \tag{29.9b}
$$

where ρ_{ij} is the phenotypic covariance between z_i and z_j. Hence, for Gaussian-distributed phenotypes,

$$
\left| \frac{C_{ij}}{P_{ij}} \right| \leq \sqrt{I} \, \sqrt{1 + \rho_{ij}^{-2}} \tag{29.10}
$$

which is a variant of the original bound based on I suggested by Arnold (1986). Note when $i = j$, $\rho_{ii} = 1$ and we recover Equation 28.14b.

The Quadratic Selection Gradient γ

Like the directional selection differential $\mathbf{S}$, the quadratic selection differential $\mathbf{C}$ confounds the effects of direct selection with selection on phenotypically correlated characters. However, as was true for $\mathbf{S}$, these indirect effects can also be removed by a regression. Consider the best *quadratic* regression of relative fitness as a function of phenotypic value,

$$
w(\mathbf{z}) = a + \sum_{j=1}^{n} b_j \, z_j + \frac{1}{2} \sum_{j=1}^{n} \sum_{k=1}^{n} \gamma_{jk} \, (z_j - \mu_j)(z_k - \mu_k) \tag{29.11a}
$$

$$
= a + \mathbf{b}^T \mathbf{z} + \frac{1}{2}(\mathbf{z} - \boldsymbol{\mu})^T \boldsymbol{\gamma} \, (\mathbf{z} - \boldsymbol{\mu}) \tag{29.11b}
$$

Using multiple regression theory, Lande and Arnold (1983) showed that the matrix $\boldsymbol{\gamma}$ of quadratic partial regression coefficients is given by

$$\boldsymbol{\gamma} = \mathbf{P}^{-1}\,\boldsymbol{\sigma}[\,w, (\mathbf{z} - \boldsymbol{\mu})(\mathbf{z} - \boldsymbol{\mu})^T\,]\,\mathbf{P}^{-1} = \mathbf{P}^{-1}\,\mathbf{C}\,\mathbf{P}^{-1} \qquad (29.12)$$

This is the **quadratic selection gradient** and (like $\boldsymbol{\beta}$) removes the effects of phenotypic correlations (among the measured traits), providing a more accurate picture of how selection is operating on the multivariate phenotype.

The vector of linear coefficients $\mathbf{b}$ for the quadratic regression need not equal the vector of partial regression coefficients $\boldsymbol{\beta}$ obtained by assuming only a *linear* regression. Equation 28.25 shows (for the univariate case) that if the phenotypic distribution is skewed, the linear term in the quadratic regression b is a function of both S and C, while the linear term in a linear regression β is only a function of S. When phenotypes are multivariate normal, skew is absent and Lande and Arnold (1983) show that $\mathbf{b} = \boldsymbol{\beta}$, recovering the multivariate version of the Pearson-Lande-Arnold regression,

$$w(\mathbf{z}) = a + \boldsymbol{\beta}^T \mathbf{z} + \frac{1}{2}(\mathbf{z} - \boldsymbol{\mu})^T \boldsymbol{\gamma}\,(\mathbf{z} - \boldsymbol{\mu}) \qquad (29.13)$$

As with the linear regression, one typically standarizes variables first, in which case Equation 29.13 can be written more compactly as

$$w(\mathbf{z}) = a + \boldsymbol{\beta}^T \mathbf{z} + \frac{1}{2}\mathbf{z}^T \boldsymbol{\gamma}\mathbf{z}$$

Since the γ_{ij} are partial regression coefficients, they predict the change in expected fitness caused by changing the associated quadratic deviation while holding all other variables constant. Increasing $(z_j - \mu_j)(z_k - \mu_k)$ by one unit in such a way as to hold all other variables and all other pairwise combinations of characters constant, relative fitness is expected to change by γ_{jk} for $j \neq k$ and by $\gamma_{jj}/2$ if $j = k$ (the difference arises because $\gamma_{jk} = \gamma_{kj}$, so that γ_{jk} appears twice in the regression unless $j = k$). The coefficients of $\boldsymbol{\gamma}$ thus describe the nature of selection on *quadratic* deviations from the mean for both single characters and pairwise combinations of characters. $\gamma_{ii} < 0$ implies fitness is expected to decrease as z_i moves away (in either direction) from its mean. As was discussed in Chapter 28, this is a necessary, *but not sufficient*, condition for stabilizing selection on character i. As a result, the term **convex selection** or **convex fitness surface** is often used with the term *stabilizing selection* restricted for when the fitness surface is *both* convex *and* the population distribution is under a peak in the fitness surface. Similarly, $\gamma_{ii} > 0$ implies fitness is expected to increase as i moves away from its mean (**concave selection**), again a necessary, but not sufficient conditional for disruptive selection. Turning to combinations of characters, non-zero values of γ_{jk} ($j \neq k$) suggests the presence of **correlation selection** — $\gamma_{jk} > 0$ suggests selection for a positive phenotypic correlation between characters j and k, while $\gamma_{jk} < 0$ suggests selection for a negative phenotypic correlation.

Example 29.3.   Brodie (1992) examined one-year survivorship in the garter snake *Thamnophis ordinoides* in a population in Oregon. Over a three year period, 646 snakes were marked, 101 of which were eventually recaptured. Four morphological and behavior characters were measured — overall stripedness of the body color pattern (stripes), sprint speed, distance moved until an antipredator display performed, and number of reversals of direction during flight from predators (reversals). None of the β_i or γ_{ii} were significant. However, there was a significant quadratic association between striping pattern and number of reversals, with $\gamma_{ij} =$

-0.268 ± 0.097 (confidence intervals were generated using the delete-one jackknife program of Mitchell-Olds 1989). As is shown in Figure 29.2, the best-fitting quadratic regression of individual fitness has a saddle point. Brodie suggests a biological explanation for selection favoring a negative correlation between these two characters. The presence of body stripes makes it difficult for predators to judge the speed of the snake, so frequent reversals would be disadvantageous. Conversely, when the body pattern is banded, blotched or spotted, detection of movement by visual predators is enhanced. In such individuals, frequent reversals can disrupt a visual search.
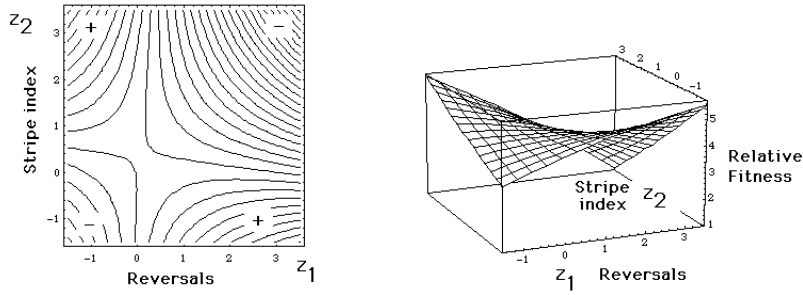

Figure 29.2.   The fitness surface (measured as one-year survivorship) for number of reversals (z_1) and body stripe index (z_2) in the garter snake *Thamnophis ordinoides*. There is a significant correlational gradient between these two characters, with all other directional and quadratic gradients being nonsignificant. **Left**: Plotting lines of equal fitness, with peaks represented by $+$ and valleys by $-$, shows the best-fitting quadratic fitness surface has a saddle point. **Right**: Three-dimensional representation of the best-fitting quadratic fitness surface. The eigenvalues of γ are 0.256 and -0.290, indicating roughly equal amounts of directional selection along one canonical axis (given by the index $0.77 \cdot z_1 - 0.64 \cdot z_2$) and convex selection along the other ($0.64 \cdot z_1 + 0.77 \cdot z_2$). Based on data from Brodie (1992).

Although it seems straightforward to infer the overall nature of selection by looking at these various pairwise combinations, this can give a very misleading picture about the geometry of the fitness surface (e.g., Figure 29.3). We discuss this problem and its solution shortly.

Finally, we can see the effects of phenotypic correlations in the quadratic selection differential. Solving for $\mathbf{C}$ by post- and pre-multiplying γ by $\mathbf{P}$ gives $\mathbf{C} = \mathbf{P}\,\gamma\,\mathbf{P}$, or

$$C_{ij} = \sum_{k=1}^{n} \sum_{\ell=1}^{n} \gamma_{k\ell}\, P_{ik}\, P_{\ell j} \tag{29.14}$$

showing that within-generation changes in phenotypic covariance, as measured by $\mathbf{C}$, are influenced by quadratic selection on phenotypically-correlated characters.

Quadratic Gradients, Fitness Surface Geometry and Selection Response

When phenotypes are multivariate-normally distributed, γ provides a measure of the average curvature of the **individual** fitness surface, as

$$\gamma = \int \mathbf{H_Z}[\,w(\mathbf{z})\,]\,\varphi(\mathbf{z})\,d\mathbf{z} \tag{29.15a}$$

where $\mathbf{H_Z}[\,f\,]$ denotes the hessian matrix of f with respect to $\mathbf{z}$ (the matrix of all second partial derivatives, $\mathbf{H}_{ij} = \partial^2/\partial z_i \partial z_j$) and is a multivariate measure of the quadratic (local) curvature of a function (Appendix A5). This result, due to Lande and Arnold (1983) can be obtained by an integration by parts argument similar to that used to obtain Equation 29.6, and holds for both frequency-dependent and frequency-independent fitnesses. When fitnesses are frequency-independent (again provided $\mathbf{z} \sim$MVN), $\boldsymbol{\gamma}$ provides a description of the curvature of the **log mean population** fitness surface, with

$$\mathbf{H_{\boldsymbol{\mu}}}[\,\ln \overline{W}(\boldsymbol{\mu})\,] = \boldsymbol{\gamma} - \boldsymbol{\beta}\boldsymbol{\beta}^T \tag{29.15b}$$

This result is due to Lande (cited in Phillips and Arnold 1989) and points out that there are two sources for curvature in the *mean* fitness surface: $-\boldsymbol{\beta}\boldsymbol{\beta}^T$ from directional selection and $\boldsymbol{\gamma}$ from quadratic selection.

Finally, when the breeders' equation holds, $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ are sufficient to describe how phenotypic selection alters the additive-genetic covariance matrix. As we show in Chapter 31, the additive-genetic covariance matrix $\mathbf{G}^*$ following selection (but before reproduction) is given by

$$\mathbf{G}^* - \mathbf{G} = \mathbf{G}\left(\boldsymbol{\gamma} - \boldsymbol{\beta}\boldsymbol{\beta}^T\right)\mathbf{G} \tag{29.16}$$

Fitness Surface Curvature and Within-generation Changes in Variances and Covariances

Equations 29.14 - 29.16 provide some insight into the connection between (within-generation) changes in phenotypic and genetic variances and the curvature of the fitness surface. To see these, we ignore the complications introduced by correlations. First, consider the within-generation change in the phenotypic variance. Equations 29.7b and 29.14 give

$$C_{ii} = \sigma^2(z_i^*) - \sigma^2(z_i) + S_i^2 = \gamma_{ii}P_{ii}^2 \tag{29.17a}$$

Hence, the within-generation change in the phenotypic variance for trait i is

$$\delta\sigma^2(z_i) = \gamma_{ii}P_{ii}^2 - S_i^2 \tag{29.17b}$$

Thus, convex selection ($\gamma_{ii} < 0$) reduces the phenotypic variance, while concave selection ($\gamma_{ii} > 0$) increases it. The net effect of directional selection is to always reduce the variance, so that undetected directional selection can mask the effects of convex selection and enhance the effects of concave selection. Likewise, from Equation 29.16 (assuming no genetic correlations), the within-generation change in the additive variance is

$$\delta\sigma^2(A_i) = \left(\gamma_{ii} - \beta_i^2\right)\sigma^4(A_i) \tag{29.17c}$$

As with the phenotypic variance, both convex and directional selection reduce the additive variance, while concave selection increases it. Note that Equation 29.17c (as well as Equation 29.16) is the *within-generation* change in the additive genetic variance. Recombination and segregation in these selected individuals change the additive variance present in their offspring by reducing the disequilibrium generated by selection and adding variation due to the segregation variance (Chapters 13, 24, and 31). Changes in the mean, on the other hand, are due to changes in the breeding value (not its variance), so that the within-generation change in mean breeding value equals the expected between-generation change in the mean (Chapter 10).

What about the effect of correlational selection ($\gamma_{ij} \neq 0$)? Again, assuming all correlations are (initially) zero Equation 29.14 gives

$$C_{ij} = \sigma(z_i^*, z_j^*) - \sigma(z_i, z_j) + S_i S_j = 2\gamma_{ij}P_{ii}P_{jj} \tag{29.18a}$$

and the within-generation change in the phenotypic covariance becomes

$$\delta\sigma(z_i, z_j) = 2\gamma_{ij}P_{ii}P_{jj} - S_iS_j \tag{29.18b}$$

Positive γ_{ij} increases the phenotypic correlation, while $\gamma_{ij} < 0$ reduces it. Note that directional selection does not have a uniform effect. If both i and j are selected in the same direction, this decreases the phenotypic correlation, while if they are selected in opposite directions, this increases the correlation. Assuming no (initial) genetic correlations, Equation 29.14 gives the (within-generation) change in the genetic covariance as

$$\delta\sigma(A_i, A_j) = \left(2\gamma_{ij} - \beta_i\beta_j\right)\sigma^2(A_i)\sigma^2(A_j) \tag{29.18b}$$

Under the infinitesimal model, this change is due entirely to disequilibrium, which (under no linkage) is reduced by half in the offspring (Chapters 13, 31).

Table 29.1 (on the next page) summarizes the major features of linear and quadratic differentials and gradients. Excellent overviews are also provided by Arnold et al. (2001) and Brodie et al. (1995).

MULTIDIMENSIONAL QUADRATIC FITNESS REGRESSIONS

As noted in Chapter 28, in many cases approximating the individual fitness function by a quadratic may give a very distorted view of the true fitness surface (e.g., when multiple fitness peaks are present). With this caveat in mind, quadratic fitness surfaces are still quite useful. One advantage is that a quadratic is the simplest surface allowing for curvature. Further, when phenotypes are gaussian distributed, the coefficients in the quadratic regression also appear as the coefficients of equations for predicting evolutionary change (Table 29.1). We briefly review some statistical issues of fitting such regressions before examining the geometry of multivariate quadratic regressions, which can get rather involved.

Estimation, Hypothesis Testing and Confidence Intervals

Even if we can assume that a best-fit quadratic is a reasonable approximation of the individual fitness surface, we are still faced with a number of statistical issues. Unless we test for, and confirm, multivariate normality (Appendix 4 reviews such tests), $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ must be estimated from separate regressions — $\boldsymbol{\beta}$ from the best linear regression, $\boldsymbol{\gamma}$ from the best quadratic regression. In either case, there are a large number of parameters to estimate — $\boldsymbol{\gamma}$ has $k(k+1)/2$ terms and $\boldsymbol{\beta}$ has k, for a total $k(k + 3)/2$. With 5, 10, and 25 characters, this corresponds to 20, 65 and 350 parameters. The number of observations n should greatly exceed $k(k + 3)/2$ in order estimate these parameters with any precision.

A second problem is **multicollinearity** — if many of the characters being measured are highly correlated, the phenotypic covariance matrix $\mathbf{P}$ can be nearly singular, so that small errors in estimating $\mathbf{P}$ result in large differences in $\mathbf{P}^{-1}$. This, in turn, gives a very large sampling variance for the estimate of β and γ. A quick check for multicollinearity is to regress each trait on all others. Subtracting the resulting model R^2 from one determines the **tolerance**, with very high R^2 and/or low tolerances indicating multicollinearity is likely an issue. One possibility solution for dealing with multicollinearity is to use principal components (Appendix 4) to extract a subset of the characters (measured as PCs, linear combinations of the characters) that explains most of the phenotypic variance of $\mathbf{P}$, and fitness regressions using the first few PCs as the characters can then be computed (Lande and Arnold 1983). This approach also reduces the problem of the number of parameters to estimate, but risks the real possibility of removing the most important characters. Further, PCs are often difficult to interpret biologically. While the first PC for morphological characters generally corresponds

Table 29.1.   Analogous features of directional and quadratic differentials and gradients.

Changes in Means	**Changes in Covariances**
(Directional Selection)	(Quadratic Selection)

Differentials measure the covariance between relative fitness and phenotype

$$S_i = \sigma\left[\, w, z_i \,\right] \qquad\qquad C_{ij} = \sigma\left[\, w, (z_i - \mu_i)(z_j - \mu_j) \,\right]$$

The opportunity for selection bounds the differential

$$\frac{|S_i|}{\sigma(z_i)} \leq \sqrt{I} \qquad\qquad \left|\frac{C_{ij}}{P_{ij}}\right| \leq \sqrt{I}\,\sqrt{1 + \rho_{ij}^{-2}}$$

for any distribution of **z**   provided **z** $\sim$ MVN

Differentials confound direct and indirect selection

$$\mathbf{S} = \boldsymbol{\mu}^* - \boldsymbol{\mu} = \mathbf{P}\boldsymbol{\beta} \qquad\qquad \mathbf{C} = \mathbf{P}^* - \mathbf{P} + \mathbf{S}\mathbf{S}^T = \mathbf{P}\,\boldsymbol{\gamma}\,\mathbf{P}$$

Gradients measure the amount of direct selection

$$\boldsymbol{\beta} = \mathbf{P}^{-1}\mathbf{S} \qquad\qquad \boldsymbol{\gamma} = \mathbf{P}^{-1}\mathbf{C}\mathbf{P}^{-1}$$

Gradients describe the slope and curvature of the mean population fitness
  surface, provided $z \sim$ **MVN and fitnesses are frequency-independent**

$$\beta_i = \frac{\partial\,\ln\overline{W}(\boldsymbol{\mu})}{\partial\mu_i} \qquad\qquad \gamma_{ij} = \frac{\partial^2\,\ln\overline{W}(\boldsymbol{\mu})}{\partial\mu_i\,\partial\mu_j} + \beta_i\beta_j$$

Gradients describe the average slope and average curvature of the individual fitness surface, provided z $\sim$ MVN

$$\beta_i = \int \frac{\partial\,w(\mathbf{z})}{\partial z_i}\,\varphi(\mathbf{z})\,\mathrm{d}\mathbf{z} \qquad\qquad \gamma_{ij} = \int \frac{\partial^2\,w(\mathbf{z})}{\partial z_i\,\partial z_j}\,\varphi(\mathbf{z})\,\mathrm{d}\mathbf{z}$$

Gradients appear as coefficients in fitness regressions

$$w(\mathbf{z}) = a + \sum \beta_j\,(z_j - \mu_j) \qquad w(\mathbf{z}) = a + \sum b_j\,(z_j - \mu_j) + \tfrac{1}{2}\sum_{j,i} \gamma_{jk}\,(z_j - \mu_j)(z_k - \mu_k)$$

$$w(\mathbf{z}) = a + \boldsymbol{\beta}^T(\mathbf{z} - \boldsymbol{\mu}) \qquad w(\mathbf{z}) = a + \mathbf{b}^T(\mathbf{z} - \boldsymbol{\mu}) + \tfrac{1}{2}(\mathbf{z} - \boldsymbol{\mu})^T\boldsymbol{\gamma}(\mathbf{z} - \boldsymbol{\mu})$$

$\boldsymbol{\beta}$ = slope of best linear fit $\boldsymbol{\gamma}$ = the quadratic coefficient of the best
                                                          quadratic fit. $\mathbf{b} = \boldsymbol{\beta}$ when $\mathbf{z} \sim$ MVN

Gradients appear as coefficients in evolutionary equations when $(\mathbf{z}, \mathbf{g}) \sim$ **MVN**

$$\Delta\boldsymbol{\mu} = \mathbf{G}\boldsymbol{\beta} \qquad\qquad \mathbf{G}^* - \mathbf{G} = \mathbf{G}\left(\boldsymbol{\gamma} - \boldsymbol{\beta}\boldsymbol{\beta}^T\right)\mathbf{G}$$

to a general measure of size (but see Somers 1986), the others are typically much more problematic. Finally, using PCs can spread the effects of selection on one character over several PCs, further complicating interpretation. While using the PCs of **P** can be problematic, as we will see, using the PCs associated with $\boldsymbol{\gamma}$ can provide considerable insight into the nature of selection.

A variety of additional concerns were discussed in Chapter 28. Briefly, residuals of fitness regressions are expected by their nature to be poorly-behaved so that standard methods of confidence intervals on regression coefficients are often not appropriate. Mitchell-Olds and

Shaw (1987) suggest using the delete-one jackknife for approximating confidence intervals, and Mitchell-Olds (1989) has developed a program to jackknife multivariate quadratic regressions. Likewise, the discussions of randomization tests and cross-validation procedures in Chapter 28 extend to multivariate regressions in a straightforward manner. Multivariate tests of the presence of a single mode in the fitness surface are discussed by Mitchell-Olds and Shaw (1987), and we introduce the Box-Hunter confidence volume for a quadratic stationary point shortly (after introducing some required matrix machinery).

Regression Packages and Coefficients of γ

The coefficients of the elements of γ have a form that may be different from the output of a quadratic regression package. Suppose we have two variables with mean zero. Under a Lande-Arnold regression the quadratic contribution is

$$\frac{\gamma_{11}}{2} z_1 + \frac{\gamma_{22}}{2} z_2 + \gamma_{12} \cdot z_1 \cdot z_2$$

However, many regression packages output the quadratic coefficients as

$$b_{11} \cdot z_1 + b_{22} \cdot z_2 + b_{12} \cdot z_1 \cdot z_2$$

In such cases, $\gamma_{ii}/2 = b_{ii}$, or $\gamma_{ii} = 2b_{ii}$, while $\gamma_{ij} = b_{ij}$ for $i \neq j$. Failure to make this correction results in the reported γ_{ii} coefficients being only half their true value, *underestimating* the strength of quadratic selection on z_i (Lande and Arnold 1983, Stinchcombe et al. 2008). Just how widespread this mistake is in the literature remains unclear.

Geometric Aspects

In spite of their apparent simplicity, multivariate quadratic fitness regressions have a rather rich geometric structure. Scaling characters so that they have mean zero, the general quadratic fitness regression can be written as

$$w(\mathbf{z}) = a + \sum_{i=1}^n b_1 z_i + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \gamma_{ij} z_i z_j = a + \mathbf{b}^T \mathbf{z} + \frac{1}{2} \mathbf{z}^T \gamma \mathbf{z} \qquad (29.19)$$

If $\mathbf{z} \sim$ MVN, then $\mathbf{b} = \boldsymbol{\beta}$ (the vector of coefficients of the best *linear* fit). Note that if we regard Equation 29.19 as a second-order Taylor series approximation (Appendix 5) of $w(z)$, $\mathbf{b}$ and γ can be interpreted as the gradient and hessian of individual fitness evaluated at the population mean (here $\boldsymbol{\mu} = \mathbf{0}$ by construction). The nature of curvature of Equation 29.19 is determined by the matrix γ. Even though a quadratic is the simplest curved surface, its geometry can still be difficult to visualize (Phillips and Arnold 1989, Brodie et al. 1995).

We start our exploration of this geometry by considering the gradient of this best-fit quadratic fitness surface. Applying Equations A5.1b and A5.1c to Equation 29.19 gives

$$\nabla_{\mathbf{z}}[\, w(\mathbf{z})\,] = \mathbf{b} + \gamma \mathbf{z} \qquad (29.20a)$$

At the point $\mathbf{z}$, the direction of steepest ascent on the fitness surface (the direction in which to move in phenotype space to maximally increase individual fitness) is given by the vector $\mathbf{b} + \gamma \mathbf{z}$. If the true individual fitness surface is indeed a quadratic, then the average gradient of *individual* fitness taken over the distribution of phenotypes is

$$\int \nabla_{\mathbf{z}}[\, w(\mathbf{z})\,]\, p(\mathbf{z})\, \mathrm{d}\mathbf{z} = \mathbf{b} \int p(\mathbf{z})\, \mathrm{d}\mathbf{z} \, + \, \gamma \int \mathbf{z}\, p(\mathbf{z})\, \mathrm{d}\mathbf{z} = \mathbf{b} \qquad (29.20b)$$

as the last integral is $\boldsymbol{\mu}$ (which is zero by construction). Hence, if the true fitness function is quadratic, the average gradient of individual fitness is given by $\mathbf{b}$, independent of the distribution of $\mathbf{z}$.

Solving $\nabla_{\mathbf{z}}[w(\mathbf{z})] = \mathbf{0}$, a point $\mathbf{z}_o$ is a candidate for a local extrema (also called a **stationary point** as the gradient is zero) if $\boldsymbol{\gamma}\mathbf{z}_o = -\mathbf{b}$. When $\boldsymbol{\gamma}$ is nonsingular,

$$\mathbf{z}_o = -\boldsymbol{\gamma}^{-1}\mathbf{b} \tag{29.21a}$$

is the unique stationary point of this quadratic surface. Substituting into Equation 29.19, the expected individual fitness at this point is

$$w_o = a + \frac{1}{2}\mathbf{b}^T\mathbf{z}_0 \tag{29.21b}$$

as obtained by Phillips and Arnold (1989). Since $\partial^2 w(\mathbf{z})/\partial z_i\,\partial z_j = \gamma_{ij}$, the hessian of $w(\mathbf{z})$ is just $\boldsymbol{\gamma}$. Thus $\mathbf{z}_0$ is a local minimum if $\boldsymbol{\gamma}$ is **positive-definite** (all eigenvalues are positive), a local maximum if $\boldsymbol{\gamma}$ is **negative-definite** (all eigenvalues are negative), or a saddle point if the eigenvalues differ in sign (see Appendix 5). If $\boldsymbol{\gamma}$ is singular (has at least one zero eigenvalue) then there is no unique stationary point. An example of this is seen in Figure 29.3 where there is a ridge (rather than a single point) of phenotypic values having the highest fitness value. The consequence of a zero eigenvalue is that the fitness surface has no curvature along the axis defined by the associated eigenvector. If $\boldsymbol{\gamma}$ has k zero eigenvalues, then the fitness surface has no curvature along k dimensions. Ignoring fitness change along these dimensions, the remaining fitness space has a single stationary point, which is given by Equation 29.21a for γ and $\mathbf{b}$ reduced to the $n - k$ dimensions showing curvature.

A Brief Digression: Orthonormal and Diagonalized Matrices

We need some additional matrix machinery before furthering our discussion of the geometry of the quadratic fitness surface, which is defined by the geometry of the matrix $\boldsymbol{\gamma}$. Appendix 4 examines these issues of matrix geometry in greater detail.

Matrix transformations (multiplying a vector by a matrix) consist of two basic operations, rotations (changes in the direction of the 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

$$\mathbf{U}^T\,\mathbf{U} = \mathbf{U}\,\mathbf{U}^T = \mathbf{I} \tag{29.22}$$

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 coordinates are simply rotated by the same angle to create the new coordinate system (Appendix 4). Thus the angle between any two vectors remains unchanged following their transformation by the same orthonormal matrix. If the angle between $\mathbf{x}_1$ and $\mathbf{x}_2$ is θ, then the angle between $\mathbf{y}_1 = \mathbf{U}\mathbf{x}_1$ and $\mathbf{y}_2 = \mathbf{U}\mathbf{x}_2$ is also θ.

A symmetric matrix $\mathbf{A}$ (such as a variance-covariance matrix) can be **diagonalized** as

$$\mathbf{A} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T \tag{29.23}$$

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

$$\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n) = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & \lambda_n \end{pmatrix} \tag{29.24a}$$

and

$$\mathbf{U} = (\, \mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_n \,) \tag{29.24b}$$

Geometrically, $\mathbf{U}$ describes a rigid rotation of the original coordinate system while $\boldsymbol{\Lambda}$ is the amount by which unit lengths in the original coordinate system are scaled in the transformed system. Using Equation 29.23 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 $\mathbf{A}$,

$$\begin{aligned} \mathbf{U}^T \mathbf{A} \mathbf{U} = \mathbf{U}^T (\mathbf{U} \boldsymbol{\Lambda} \mathbf{U}^T) \mathbf{U} &= (\mathbf{U}^T \mathbf{U}) \boldsymbol{\Lambda} (\mathbf{U}^T \mathbf{U}) \\ &= \boldsymbol{\Lambda} \end{aligned} \tag{29.25}$$

As we will shortly see, the effect of using such a transformation is that (on this new scale) we remove all cross-product terms. Put another way, *on this new scale, there is no correlational selection*. A few very useful results immediately follow from Equation 29.25. For $\mathbf{A}^{1/2}$ and $\mathbf{A}^{-1}$, the $\mathbf{U}$ matrix is unchanged, while the diagonal elements in the associated $\boldsymbol{\Lambda}$ matrix are either given by the square root or inverse, respectively. Thus, $\mathbf{A}$, $\mathbf{A}^{1/2}$ and $\mathbf{A}^{-1}$ (provided it exists) all have the same eigenvectors and their eigenvalues are related as $\lambda_i, \lambda_i^{1/2}, \lambda_i^{-1}$.

Canonical Transformation of γ

While the (quadratic) fitness surface curvature is completely determined by γ, it is easy to be misled about the actual nature of the fitness surface if one simply tries to infer it by inspection of γ, as the following example illustrates.

Example 29.4.   Consider selection acting on two characters z_1 and z_2. Suppose we find that $\gamma_{11} = -2$ and $\gamma_{22} = -1$, suggesting that the individual fitness surface has negative curvature in both z_1 and z_2. At first glance the picture this evokes is stabilizing selection on both z_1 and z_2, with the stabilizing selection surface perhaps rotated due to selection for correlations between z_1 and z_2. The first caveat, as mentioned in Chapter 28, is that negative curvature (convexity), by itself, does not imply a local maximum. Even if γ is negative definite, the equilibrium point $\mathbf{z}_0$ may be *outside* of the observed range of population values and hence not currently applicable to the population being studied. A much more subtle point is that, as Figure 29.3 shows, the nature of the fitness surface is very dependent on the amount of selection for correlations between z_1 and z_2. Figure 29.3 considers the surfaces associated with three different values of γ_{12} under the assumption that $\mathbf{b} = 0$. Note that although in all three cases $\gamma_{12} > 0$ (i.e., selection favors increased correlations between the phenotypic values of z_1 and z_2), the fitness surfaces are qualitatively very different. When $\gamma_{12} = 0.25$, the individual fitness surface indeed shows stabilizing selection in both characters. However, when $\gamma_{12} = \sqrt{2} \simeq 1.42$, the fitness surface has a ridge in one direction, with stabilizing selection in the other. When $\gamma_{12} = 4$, the fitness surface is a saddle, with convex selection along one axis and concave selection along the other. An especially troubling point is that if the standard error of γ_{12} is sufficiently large we would not be able to distinguish between these very different types of selection even if we could show that $\gamma_{11}, \gamma_{22} < 0$, and $\gamma_{12} > 0$.
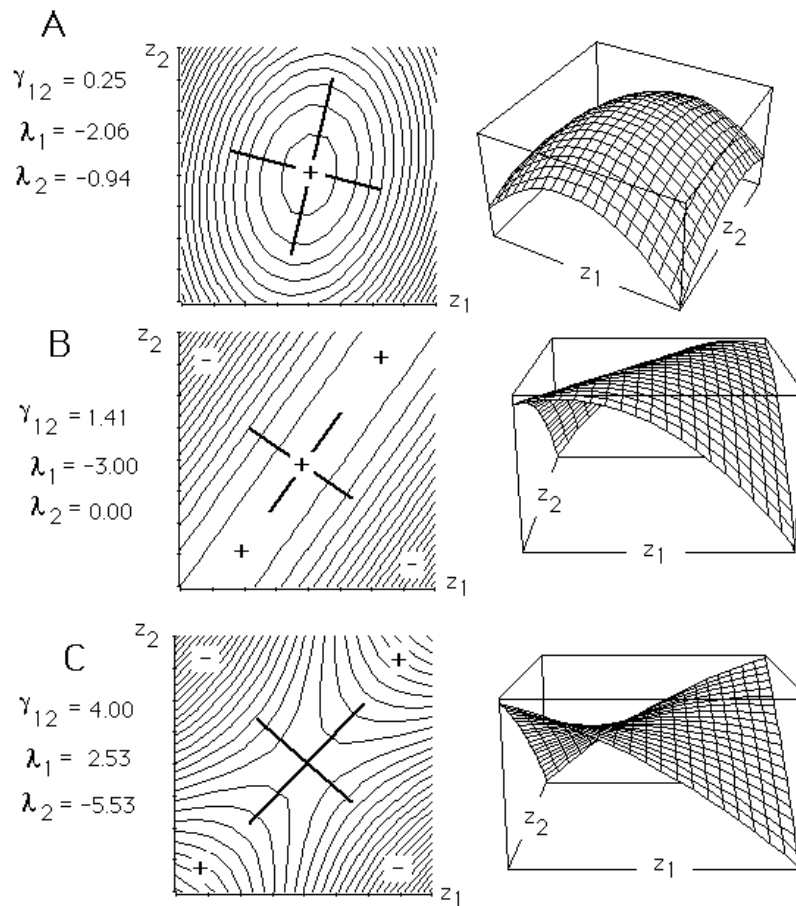
Figure 29.3. Three quadratic fitness surfaces, all of which have $\gamma_{11} = -2$ and $\gamma_{22} = -1$ and $\mathbf{b} = \mathbf{0}$. On the left are curves of equal fitness values, with peaks being represented by $+$, and valleys by $-$. **Axes of symmetry** of the surface (the **canonical** or **principal axes** of $\boldsymbol{\gamma}$) are denoted by the thick lines. These axes correspond to the eigenvectors of $\boldsymbol{\gamma}$. On the right are three dimensional plots of individual fitness as a function of the phenotypic values of the characters z_1 and z_2. **Top**: $\gamma_{12} = 0.25$. This corresponds to stabilizing selection on both characters, with fitness falling off more rapidly (as indicated by the shorter distance between contour lines) along the z_1 axis than along the z_2 axis. **Middle**: $\gamma_{12} = \sqrt{2} \simeq 1.41$, in which case $\boldsymbol{\gamma}$ is singular (as $\lambda_2 = 0$). The resulting fitness surface shows a ridge in one direction ($\mathbf{e}_2$, the eigenvector corresponding to $\lambda_2 = 0$) with strong stabilizing selection in the other ($\mathbf{e}_1$). **Bottom**: When $\gamma_{12} = 4$, the fitness surface shows a saddle point, with stabilizing selection along one of the canonical axes of the fitness surface ($\mathbf{e}_2$ corresponding to $\lambda_2 = -5.53$) and disruptive selection along the other ($\mathbf{e}_1$ corresponding to $\lambda_1 = 2.53$).

Thus, even for two characters, visualizing the individual fitness surface is not trivial and can easily be downright misleading. The problem is that the cross-product terms (γ_{ij} for $i \neq j$) make the quadratic form difficult to interpret geometrically. Removing these terms by a change of variables so that the axes of new variables coincide with the axes of symmetry of

the quadratic form (its canonical axes) greatly facilitates visualization of the fitness surface.

Motivated by this, Phillips and Arnold (1989) suggest using two slightly different versions of the canonical transformation of γ to clarify the geometric structure of the best fitting quadratic fitness surface. Applying Equation 29.25, if we consider the matrix $\mathbf{U}$ whose columns are the eigenvalues of γ, the transformation $\mathbf{y} = \mathbf{U}^T \mathbf{z}$ (hence $\mathbf{z} = \mathbf{U}\mathbf{y}$, since $\mathbf{U}^{-1} = \mathbf{U}^T$ as $\mathbf{U}$ is orthonormal) removes all cross-product terms in the quadratic form,

$$
\begin{aligned}
w(\mathbf{z}) &= a + \mathbf{b}^T \mathbf{U}\, \mathbf{y} + \frac{1}{2}\, (\mathbf{U}\mathbf{y})^T \gamma (\mathbf{U}\mathbf{y}) \\
&= a + \mathbf{b}^T \mathbf{U}\, \mathbf{y} + \frac{1}{2}\, \mathbf{y}^T \left(\mathbf{U}^T \gamma \mathbf{U}\right) \mathbf{y} \\
&= a + \mathbf{b}^T \mathbf{U}\, \mathbf{y} + \frac{1}{2}\, \mathbf{y}^T \Lambda\, \mathbf{y} \\
&= a + \sum_{i=1}^{n} \theta_i\, y_i + \frac{1}{2} \sum_{i=1}^{n} \lambda_i\, y_i^2
\end{aligned}
\tag{29.26}
$$

where $\theta_i = \mathbf{e}_i^T \mathbf{b}$ and $y_i = \mathbf{e}_i^T \mathbf{z}$, with λ_i and $\mathbf{e}_i$ are the eigenvalues and associated unit eigenvectors of γ. Alternatively, if a stationary point $\mathbf{z}_0$ exists (e.g., γ is nonsingular (all $\lambda_i \neq 0$), the change of variables $\mathbf{y} = \mathbf{U}^T (\mathbf{z} - \mathbf{z}_0)$ further removes all linear terms (Box and Draper 1987), so that

$$
w(\mathbf{z}) = w_o + \frac{1}{2}\, \mathbf{y}^T \Lambda\, \mathbf{y} = w_o + \frac{1}{2} \sum_{i=1}^{n} \lambda_i\, y_i^2
\tag{29.27}
$$

where $y_i = \mathbf{e}_i^T (\mathbf{z} - \mathbf{z}_0)$ and w_o is given by Equation 29.21b. Equation 29.26 is called the **A canonical form** and Equation 29.27 the **B canonical form** (Box and Draper 1987). Both forms represent a rotation of the original axis to the new set of axes (the canonical axes of γ) that align them with axes of symmetry of the quadratic surface. The B canonical form further shifts the origin to the stationary point $\mathbf{z}_o$. Since the contribution to individual fitness from $\mathbf{b}^T \mathbf{z}$ is a hyperplane, its effect is to tilt the fitness surface. The B canonical form "levels" this tilting, allowing us to focus entirely on the curvature (quadratic) aspects of the fitness surface.

The orientation (the **principal** or **major axes**) of the quadratic surface is determined by the eigenvectors $(\mathbf{e}_1, \cdots, \mathbf{e}_n)$ of γ, while the eigenvalues $(\lambda_1, \cdots, \lambda_n)$ of γ determine the nature and amount of curvature of the surface along each canonical axis. Along the axis defined by $y_i = \mathbf{e}_i^T \mathbf{z}$, the individual fitness function has positive curvature (is concave) if $\lambda_i > 0$, has negative curvature (convex) if $\lambda_i < 0$, or has no curvature (a plane) if $\lambda_i = 0$. The amount of curvature is indicated by the magnitude of λ_i, the larger $|\lambda_i|$, the more extreme the curvature.

An alternative way to think about this canonical transformation is that the original vector $\mathbf{z}$ of n characters is transformed into a vector $\mathbf{y}$ of n independent selection indices (Simms 1990). Directional selection on the index y_i is measured by θ_i, while quadratic selection on y_i is measured by λ_i.

Returning to Figure 29.3, we see that the axes of symmetry of the quadratic surface are the canonical axes of γ. For $\gamma_{12} = 0.25$, $\lambda_1 = -2.06$ and $\lambda_2 = -0.94$ so that the fitness surface is convex along each canonical axis, with more extreme curvature along the y_1 axis. When $\gamma = \sqrt{2}$, one eigenvalue is zero while the other is -3, so that the surface shows no curvature along one axis (it is a plane), but is strongly convex along the other. Finally, when $\gamma_{12} = 4$, the two eigenvalues differ in sign, being -5.53 and 2.53. This generates a saddle point with the surface being concave along one axis and convex along with other, with the convex curvature being more extreme.

If $\lambda_i = 0$, the fitness surface along y_i has no curvature, so that the fitness surface is a ridge along this axis. If $\theta_i = \mathbf{e}_i^T \mathbf{z} > 0$ this is a rising ridge (fitness increases as y_i increases), it is a falling ridge (fitness decrease as y_i increase) if $\theta_i < 0$, and is flat if $\theta_i = 0$. Returning to Figure 29.3B, the effect of $\mathbf{b}$ is to tilt the fitness surface. Denoting values on the axis running along the ridge by y_1, if $\theta_1 > 0$ the ridge rises so that fitness increases as y_1 increases. Even if γ is not singular, it may be nearly so, with some of the eigenvalues being very close to zero. In this case, the fitness surface shows little curvature along the axes given by the eigenvectors associated with these near-zero eigenvalues. From Equation 29.26, the fitness change along a particular axis ($\mathbf{e}_i$) is $\theta_i\, y_i + (\lambda_i/2)\, y_i^2$. If $|\theta_i| >> |\lambda_i|$, the curvature of the fitness surface along this axis is dominated by the effects of linear (as opposed to quadratic) selection. Phillips and Arnold (1989) present a nice discussion of several other issues relating to the visualization of multivariate fitness surfaces, while Box and Draper (1987) review the statistical foundations of this approach.

Are Traits Based on Canonical Axes "Real"?

While there clearly are significant benefits from using the canonical rotation of γ to infer those trait combinations under the strongest quadratic selection, this approach has also sparked a lively debate in the literature. Blows (2007a, b) has championed it as providing considerable insight into the nature of selection, while Conner (2007) suggests that "these advantages are usually outweighed by the disadvantage that the results are not very biological inter-pretable", a point also echoed by Hunt et al. (2007). Basically, the concern these authors express is the same concern about the use of principal components: even though a specific combination of traits may account for most of the variation, its biological interpretation may be convoluted (at best). Discrete traits, on the other hand, are straightforward. This argument raises two questions: what are the true targets of selection and what is a trait?

The power of quantitative genetics is that *anything* we can measure can be regarded as a trait, no matter how strange or seemly biologically unreasonable it may be. Clearly, ecologists and evolutionary biologists working on specific traits (such as clutch size or body weight) bring a wealth of empirical knowledge about these traits to the table when considering selection. In this sense, field biologists regard many traits as "natural" objects. While most would agree that some are (e.g., clutch size), other traits like body shape is more problematic, as it can be defined and measured in a myriad of ways. At a deeper level, it is the nature of the question that usually determines whether a biologist regards a particular trait as a "natural" object. Even when considering the same general features in the same species, a developmental biologist's view of those traits that are natural may be quite different from an ecologist, and in turn both views may be different from an evolutionary biologist.

Thus, when a trait is not regarded as a "natural" object, but rather is some weighted combinations of values of "natural" objects, a biologist often feels that much of their intuitive and empirical knowledge about the individual components is diffused over some seemly arbitrary combination of their values. This is not an unreasonable view. However, selection is not reasonable in the sense that it does not care about how traits are define, it simply acts on particular multivariate phenotypes. When selection is acting on a complex structure in a medium or high dimensional space, simply examining the fitness of projections from this space onto univariate traits is incredibly misleading (Walsh 2007, Blows and Walsh 2008).

Thus, from the perspective of selection, the "natural" objects are the linear combination of trait values that comprise the canonical axes of γ. In this sense, there is a very real difference between concerns about interpretation of PCs for a phenotypic covariance matrix and the major axes of γ. The later define a real object, namely how selection views the traits under selection. One objection with using PCs from the phenotypic covariance matrix to define new traits is that this may diffuse a true target of selection, as the PCs are used simply to deal with phenotypic correlations. In contrast, the canonical axes of γ specifically *highlights* targets of

selection. As Sewall Wright (1935) insightfully noted: "It is the harmonious adjustment of all of the characertistics of the organism that is the object of selection, not the separate metrical 'characters'".

Strength of Selection: γ_{ii} Versus λ

We have seen that γ_{ii} can potentially give a very misleading picture of the nature of quadratic selection, while the eigenvalues λ of γ provide an exact description of the true nature of selection. Blows and Brooks (2003) stress this point, noting for an analysis of 19 studies that $|\gamma_{ii}|_{max} < |\lambda|_{max}$. A further point (mentioned above) is that some published studies report only half the true value for λ_{ii} due to incorrect translation of the coefficient of the quadratic regression. Thus, studies that report weak values for quadratic selection are potentially biased if they use γ_{ii} values, rather than the full geometry of γ, as described by the eigenvectors.

Example 29.5.   Brooks and Endler (2001) examined four color traits in male guppies associated with sexual selection. The estimated γ matrix was

$$\gamma = \begin{pmatrix} 0.016 & -0.016 & -0.028 & 0.103 \\ -0.016 & 0.00003 & 0.066 & -0.131 \\ -0.028 & 0.066 & -0.011 & -0.099 \\ 0.103 & -0.131 & -0.099 & 0.030 \end{pmatrix}$$

Just considering the diagonal elements suggests evidence for weak concave selection ($\gamma_{44} = 0.030$, $\gamma_{11} = 0.016$), and some evidence for very weak convex selection ($\gamma_{33} = -0.011$). However, the eigenvalues of γ are 0.132, 0.006, -0.038, and -0.064. Of these only the leading eigenvalue is significant, with the amount of concave selection being over four times that suggested from the largest γ_{ii} value.

Blows and Brooks (2003) note several advantages of focusing on estimating the λ_i versus the entire matrix of γ_{ij}, noting that there are n eigenvalues, and $n(n-1)/2$ elements in γ. Further, given that many eigenvalues may be close to zero, a **subspace** of γ (for example, the space spanned by the first few principal components of γ) may essentially capture all of the relevant information on the quadratic fitness surface. Following Simms (1990) and Simms and Rausher (1993), Blows and Brooks suggest that estimation and hypothesis testing can occur by first obtaining the eigenvectors of γ, and then use these to generate the transformed variables $\mathbf{y} = \mathbf{U}^T\mathbf{z}$ in the quadratic regression given by Equation 29.26. The quadratic terms correspond to eigenvalues, and confidence intervals, as well as significance, can be conducted within the standard GLM framework (LW Chapter 8). Bisgaard and Ankenman (1996) provide a formal statistical framework for this procedure.

Kruuk and Garant (2007), using a theorem (2.2) of Mercer and Mercer (2000), note that the largest eigenvalue of γ is always greater than any of the diagonal elements of γ. Thus, it is "algebraically inevitable" that at least one combination of traits will show stronger quadratic selection than any of the original traits (provided all $\gamma_{ii} \neq 0$). Likewise, there will be at least one combination with less selection and any of the single traits. Nevertheless, the biological issue here is whether λ_{ii} is *significantly* greater than γ_{ii} (as opposed to just being slightly larger).

Significance and Confidence Regions for a Stationary Point

Recall from Equation 29.21a that when γ is non-singular (contains no zero eigenvalues), then $\mathbf{z}_o = -\gamma^{-1}\mathbf{b}$ is the unique stationary point in the quadratic regression. If γ contains

k zero-valued eigenvalues, then there is no curvature along the trait combinations given by the k associated eigenvectors (the fitness surface is a k dimensional hyperplane along these directions), while there is curvature (and a unique stationary point) in the remaining fitness space. As mentioned, this stationary value is a maximum if all of the eigenvalues are negative (and hence a test for significant for a maximum is that all of the eigenvalues are significantly different from zero). Likewise, it is a local minimum if all the eigenvalues are negative (which can also be correspondingly tested). If γ contains both positive and negative values, then this is a saddle point. Thus, tests for maxima or minima are straightforward. What about the confidence region (or more correctly *volume*) for the stationary point?

A classical result for a quadratic regression is the **Box-Hunter confidence region**. Let $\mathbf{d_Z}$ denote the gradient vector of the best quadratic regression of w on $\mathbf{z}$, which (Equation 29.20a) is

$$\mathbf{d_Z} = \nabla_\mathbf{Z}[\,w(\mathbf{z})\,] = \mathbf{b} + \gamma\mathbf{z}$$

Suppose there are k traits and n observations, with the residuals to the quadratic regression being independent homoscedastic normal variables. Box and Hunter (1954) showed that a $100\,(1-\alpha)$ percent confidence volume is given by those vectors $\mathbf{z}$ that satisfy the quadratic inequality

$$\mathbf{d_Z}\mathbf{V}^{-1}\mathbf{d_Z} \leq k\,F_{1-\alpha,k,n-p} \tag{29.28}$$

where F is the $1-\alpha$ value for an F distribution with k and $n-p$ degrees of freedom ($p = k(k+3)/2$ is the total number of regression coefficients estimated) and $\mathbf{V}$ is an estimate of the covariance matrix for $\mathbf{d_Z}$. Values of $\mathbf{x}$ that satisfy Equation 29.28 are within the confidence volume for the stationary point. Also see Del Castillo and Cahya (2001) and Peterson et al. (2002) for further discussion and developments.

MULTIVARIATE NONPARAMETRIC FITNESS SURFACE ESTIMATION

As discussed in Chapter 28, using the best-fitting quadratic can result in a very misleading picture when there are multiple peaks or sharp thresholds (such as truncation selection) in the individual fitness surface. This obviously holds in multivariate space, and hence nonparameteric (free of any assumed functional form) regressions predicting fitness given a vector $\mathbf{z}$ of traits certainly have some advantages. There are, however, also significant disadvantages. We have already seen the difficulty in visualizing the multivariate fitness surface when only a simple quadratic is assumed. Fortunately, the eigenvalues of γ provide significant help in the interpretation of such quadratic surfaces. With general nonparametric methods, there is no general corresponding measure and thus one must resort to actual visualization of the fitness surface. This requires considering successive pairwise cross-sections of the fitness surface (projections onto two of the axes of the fitness function), giving a series of 3-D surfaces (one fitness and two trait axes, see Figure 29.4). Further, the results from a nonparametric regression do not immediately provide coefficients to predict the response to selection, while (under normality assumptions) these follow automatically with a quadratic regression. Given these complementary strengths and weaknesses, we strongly urge researchers to use *both* quadratic and nonparameteric regressions in the analysis of their selection data.

Projection Pursuit Regression

Schluter and Nychka (1994) extend Schluter's (1988) cubic spline univariate regressions (Chapter 28) to a vector of traits by using **projection pursuit regression** (Friedman and Stuetzle 1981). The basic idea behind PPR is to approximate some complex function $f(\mathbf{z})$ by

a series of **projection vectors** $\mathbf{a}_i$ and associated **ridge functions** f_i,

$$f(\mathbf{z}) \simeq f_1(\mathbf{a}_1^T \mathbf{z}) + f_2(\mathbf{a}_2^T \mathbf{z}) + \cdots + f_k(\mathbf{a}_k^T \mathbf{z})$$
$$= f_1(x_1) + f_x(x_2) + \cdots f_k(x_k) \tag{29.29}$$

Solutions require two numerically intensive steps: estimation of the best-fitting projection vectors $\mathbf{a}_1, \cdots, \mathbf{a}_k$ and then estimation of their associated ridge functions. Schluter and Nychka simply use cubic splines for the ridge functions. Thus, one chooses optimality and smoothing criteria and then obtains the first projection vector while fitting a cubic spline to the data along this projection. We then move onto the second projection vector and so on. The assumption is that most of the variation in the fitness surface residues on a rather low-dimesional space, so that the first few projection vectors are sufficient to approximate individual fitness. As with the eigenvectors of $\boldsymbol{\gamma}$ in a quadratic regression, the projection vectors $\mathbf{a}_i$ are those trait combinations (indexes of trait values) that experience the strongest nonlinear selection. The issue of dealing with visualization of complex spaces is largely accounted for by considering the pair-wise projections on the first few projection vectors (Example 29.6).

Thin-plate Splines

A closely related approach is **thin plate splines**, which are the two dimensional analogs of univariate cubic splines. Using a projection onto two trait axes, thin plate splines can be used to find the best fitting 3-D surface on this reduced set of axes. Again, this is another approach for visualization of complex surfaces.

Example 29.6.   Following up on the study of male color traits in guppies used in Example 29.5, Blows et al (2003) examined six traits (two morphological and four color) for their role in female attractiveness. Both quadratic and projection pursuit regressions were used to examine nonlinear selection on these traits. The trait loadings for the first two major axes of the quadaric regression (the eigenvectors $\mathbf{e}_1$ and $\mathbf{e}_2$ corresponding to the two leading eigenvectors of $\boldsymbol{\gamma}$) and the first two projection vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ for projection pursuit regression were as follows:

Trait	$\mathbf{e}_1$	$\mathbf{e}_2$	$\mathbf{a}_1$	$\mathbf{a}_2$
Body area	−0.065	0.157	0.185	0.551
Tail area	0.663	0.635	0.331	0.087
Black area	−0.245	0.645	0.430	−0.182
Fuzzy area	0.372	−0.173	0.101	−0.286
Iridescent area	0.436	−0.134	0.569	−0.757
Orange area	−0.411	0.333	0.581	−0.016

Note that these two sets of projection vectors for the different regressions are rather distinct. Thus, the major axes of the fitness surface are different if one uses a quadratic approximation as opposed to a projection pursuit regression approximation. As Figure 29.4 shows, the best fitting quadratic regression using $\mathbf{e}_1$ and $\mathbf{e}_2$ (so that $x_1 = \mathbf{e}_1^T \mathbf{z}$ and $x_2 = \mathbf{e}_2^T \mathbf{z}$ are the trait values on this fitness surface for an individual with trait vector $\mathbf{z}$) shows disruptive selection, with peaks on each of the four corners of the fitness surface. If one uses these axes but then fits the data using thin-plate splines, as Figure 29.4 shows the result is a rather different looking fitness functions, but one that also shows multiple peaks corresponding to different combinations of the traits that females find attractive. When projection pursuit regression is used, as Figure 29.4 shows, the fitness surface (in rough appearence) is similar to that for thin-plate splines using the quadratic axes. However, since the values for the two sets of axes (the trait loadings) are rather different, this superficial visual appearance must be mapped

into trait values. Again, the result is that the fitness surface has multiple peaks, with females preferring specific combinations of traits over others.
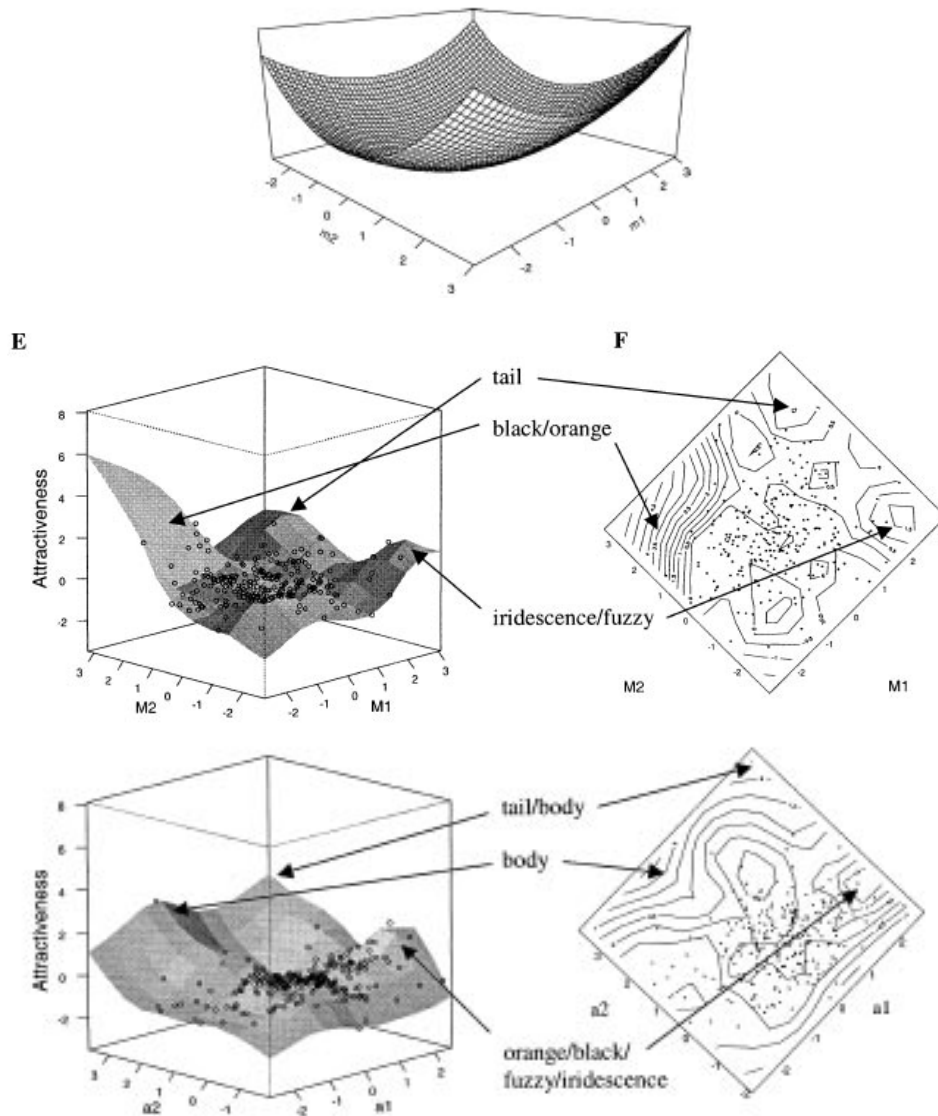

Figure 29.4. Visualization of the fitness surface for male attractiveness to female guppies as a function of color and size traits (details in Example 29.6). **Top**: The best-fitting quadratic regression. **Middle**: A thin-plate spline estimate of the fitness surface, using the axes given from the best quadratic regression. **Bottom**: The surface generated from projection pursuit regressions using the first two axes. While the thin plate spline and PPR fitness surfaces look visually similar, it is important to recall that the trait loadings on their axes are different (see Example 29.6). Contour plots on the right show the actual data (points) and the trait combinations that correspond to the different peaks.

STRENGTH OF SELECTION IN NATURAL POPULATIONS

Biologists often ask: just how strong is selection in natural populations? Phased this way, the question is ambiguous as it is not clear if the focus is on individual variance in fitness or direct selection on particular traits. One can have a very high variance in individual fitness, and hence much *potential* for phenotypic selection. However, if all of this variation in fitness is random (character-independent), then (in the extreme) there is no phenotypic selection. Focusing on the strength of selection in particular *traits*, Darwin (1859) felt that traits change very slowly, and hence selection on them was weak. Conversely, there are classic examples (such as the evolution of insecticide resistance) of rapid response in a trait, and hence presumably strong selection, but such cases are often due to sudden changes in the environment (such as the introduction of a new pesticide), and how representative such cases are remains unclear.

Attempts at measuring selection on quantitative traits in nature trace back to Bumpus (1899) and Weldon (1901). The classic book by Endler (1986) represents one of the first to attempt to summarize the average strength of selection, finding that strong selection "is not rare and may even be common", a conclusion that (at the time) was surprising to many. In 2001, Kingsolver and colleagues (Kingsolver et al. 2001, Hoekstra et al. 2001) harvested the rich literature of Lande-Arnold fitness estimates which had been accumulating over roughly twenty years, reaching the conclusion that the average strength of selection was modest. These papers generated much discussion, with some proponents claiming they supports weak to modest selection, while others claim they actually shows support for strong selection.

Kingsolver's Meta-analysis

Meta-analysis is the field of statistics that deals with summarizing the results from a large number of experiments (Hedges and Olkin 1985, Hunter and Schmidt 2005). Kingsolver and colleagues (Kingsolver et al. 2001, Hoekstra et al. 2001) performed such a meta-analysis on estimates of β and γ from 63 studies of natural populations published between 1984 and 1997. The frequency distributions of standardized values for both are given in Figure 29.5.
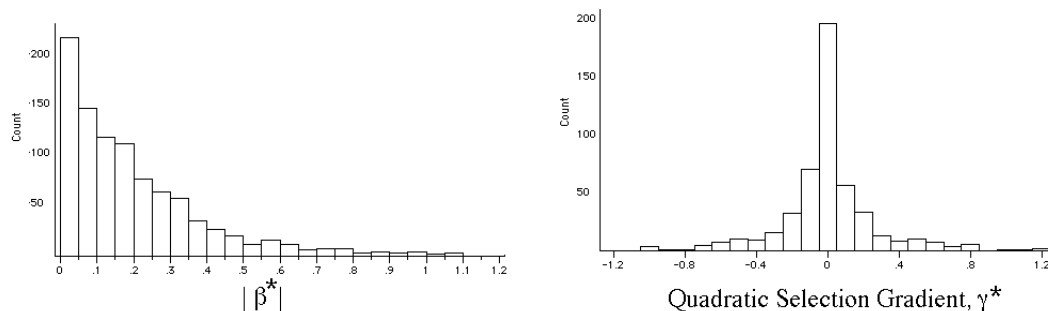

Figure 29.5.   Summary of estimates of directional and quadratic selection in the wild. All estimates are scaled in terms of phenotypic standard deviations ($\beta' = \beta/\sigma_z$, $\gamma' = \gamma/\sigma_z$). Data from the summary of Kingsolver et al. (2001). **Left:** Plot of roughly 1,000 estimated directional selection gradients in natural populations. The distribution of $|\beta'|$ was not significantly different from an exponential distribution. The medium value for $|\beta'|$ was 0.16. **Right:** Plot of roughly 470 estimates of γ' from natural populations. The medium value of $|\gamma'|$ was 0.10.

Kingsolver noted several trends from this data. First, the absolute values of (scaled) di-

rectional selection gradients essentially follows an exponential distribution with a medium (50% value) of 0.16. A β' of 0.16 implies that a one standard deviation change in the trait changes relative fitness by 16%. Thus, Kingsolver suggests that most directional selection in nature is fairly weak, although (due to the long tail of the exponential), there are a few large estimates of $|\beta'|$ (10% of the estimates in Kingsolver's summary exceeded 0.5). Further, Kingsolver suggests that the data likely contain two sources of bias (low power and **publication bias** — the failure to publish non-significant results), resulting in an *overestimate* of the average strength of β, further strengthening the claim of weak selection. Consistent with this suggestion, Kingsolver notes that most of the large estimates for $|\beta'|$ occur in studies with small sample sizes, with most estimates below 0.1 when the sample size was 1,000 or greater. Hence, it is possible that some of the large β' values are simply a consequence of sampling error due to small sample size (the Beavis effect discussed in LW Chapter 14 for QTL mapping).

Second, a wide-spread belief is that stabilizing selection is far more common than disruptive selection. However, Kingsolver observed an essentially symmetric distribution of γ' values, implying positively- and negatively-curved fitness surfaces are equally common. Further, the average strength of quadratic selection was weak, with the medium value of 0.10 for $|\gamma'|$.

Third, as shown in Figure 29.6, sexual selection (as measured by mating success) was stronger that natural selection (as measured by survival). However, Hoekstra et al. (2001) also noticed that viability selection (when measured over short intervals) was typically stronger than when measured over long intervals. However, the strength of sexual selection was largely time-scale invariant. Thus, sexual selection was stronger than viability selection over long time scales, but not over short time scales.
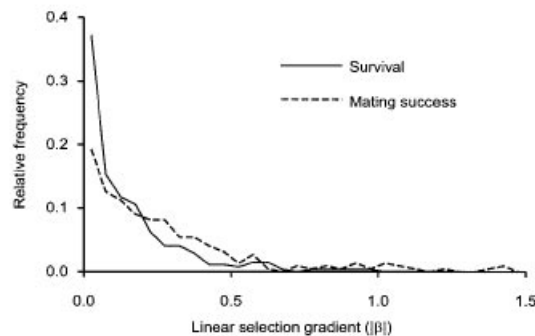

Figure 29.6.   The distribution of $|\beta'|$ for both sexual and natural selection in the Kingsolver dataset. These were measured by mating success and viability, respectively. The medium value for viability was 0.153, but 0.250 for mating success. After Hoekstra et al. (2001).

Directional Selection: Strong or Weak?

So, has the case for "weak" selection been made? As pointed out by Conner (2001), even "weak" selection can be very efficient. Consider the medium value of $\beta' = 0.16$, or $\beta = 0.16 \cdot \sigma_z$. From Equation 10.23b, the single-generation change in phenotypic standard deviations is $h^2 \cdot 0.16$. With a typically heritability of 0.4, only 16 generations of selection are required to shift the population mean by one standard deviation, and only 80 generations to shift the mean by five standard deviations.

However, if concerns about publication bias and low power are correct, then the true

medium value of $|\beta|$ may be substantially less than the "weak" medium value of 0.16 found by Kingsolver. Hersch and Phillips (2004) and Knapczyk and Conner (2007) both examined the validity of these concerns. Hersch and Phillips first considered power concerns. Recall that Figure 28.7 gives power curves as a function of $\rho = \bar{\imath}/\sqrt{I}$, which holds for univariate analyses of directional gradients. When multiple (potentially correlated) traits are considered, Hersch and Phillips show that the adjusted version of ρ to use in power calculations is

$$\rho^2 = \frac{\boldsymbol{\beta}^T \mathbf{S}}{I} \tag{29.29}$$

Hence, $\bar{\imath}^2$ is replaced by $\left(\mathbf{P}^{-1}\mathbf{S}\right)^T \mathbf{S}$ which accounts for the correlations among (measured) traits. Using Equation 29.29, Hersch and Phillips found that most of the studies summarized by Kingsolver for which power calculations can be performed (i.e., those that included estimates of I) were very underpowered, supporting Kingsolver's concern about under-reporting of small values and hence skewing estimates of $|\beta'|$ upwards.

Publication bias, also called the "**file-drawer effect**" (Rosenthal 1979) as non-significant results are not published, and hence simply "left in a file" (also see Rosenberg 2005) as also examined by Hersch and Phillips. They simulated cases where results were only published when one (or more) the gradients in a study were deemed significant. Although such a model does indeed allow for many non-significant values in the database, these are still biased by the conditioning on at least one of the results being significant. This scheme is certainly one reasonable model for publication bias. For each "publication", the simulations chose five "traits" by drawing five β values at random (Kingsolver's average number was 4.6) from a distribution with values similar to those seen by Kingsolver. If one (or more) of the five tests was significant, all values were retained, if not, all were discarded. As shown in Figure 29.7, this conditional sampling results in the medium value of β being substantially overestimated when sample size is small and/or when the amount of variation the effects explained was small (low R^2 values).
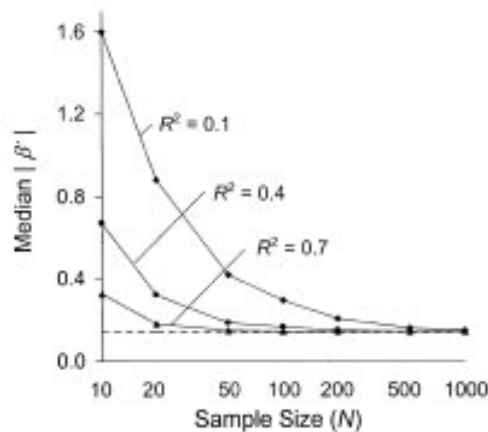

Figure 29.7.   Overestimation of selection gradients due to publication bias. Hersch and Phillips (2004) simulated a model in which five values of β were drawn from a distribution and only included if at least one of them was significant. This is certainly a reasonable model for how researchers might decide whether or not to publish their results. The model coefficient of determination R^2 is the total amount of variation in fitness examined by all of the traits (small R^2 implies low power). The dashed line represents the true medium value. Note the overestimation of the medium β values at small sample size and for small R^2 values.

Knapczyk and Conner (2007) also consider sampling error and publication bias, but arrive at the very different conclusion that "our understanding of selection is not strongly biased by these commonly invoked sources of error". To examine publication bias, these authors bin the datasets examined by Kingsolver based on sample size. If there is no publication bias, the distribution of $|\beta'|$ values should have a similar form over different sample sizes (following the exponential observed for the entire set by Kingsolver). Knapczyk and Conner did indeed find departures from this pattern for studies with very small sample sizes (less than 38), but did not for studies with larger sample sizes. Indeed, using the exponential they actually found an excess of weak selection gradients in larger studies. Thus, they conclude that publication bias, except in the smallest studies, did not appreciably upwardly bias estimates of $|\beta'|$.

Further complicating matters is the paper by Hereford et al. (2004), who suggest that directional selection is indeed quite strong. These authors first point out concerns with standardizing selection gradients to express them in terms of phenotypic standard deviations. Instead, they suggest in cases where there is a natural zero point in the trait (such as age, size, or weight), that scaling the selection gradient using the trait mean (using $\beta_\mu = \mu_z\beta$ instead of $\beta' = \beta/\sigma_z$) provides a more natural standardization scheme. In particular, they note for fitness that $\beta_\mu = 1$, providing a benchmark for the strength of selection on a trait. Further, mean-scaled gradient values can be related to elasticity coefficients used in demography (van Tienderen 2000), a key point expanded upon in Chapter 36. Using a subset of the Kingsolver data (38 studies yielding 580 estimates) that provide the required information for this standardization, they find that the medium for $|\beta_\mu|$ of 0.54, or 54% as strong as selection on fitness. This is indeed strong selection. Hereford et al. develop a correction for bias, which reduces the medium value to 0.31, but this is still quite strong selection. The authors note that this very large value is not simply a function of using a biased subset of the Kingsolver data, as the medium value of variance-standardized gradients $|\beta'|$ in their sample was 0.09, less than the value of 0.16 obtained for the full dataset by Kingsolver. The observed strengths of selection were so large that the authors seemed almost troubled by them. They suggest that selection gradients are likely inflated by focusing on single episodes of selection, rather than life-time fitness. In essence, they suggest tradeoffs (Chapter 30) result in the actual gradient of lifetime fitness for a trait to be significantly less than the gradient based on a particular episode of selection. Using average elasticity values for avian populations, they find that typical values of selection on a trait, as expressed through trait effects on viability and fecundity, are about 40% and 6% as strong as selection on fitness.
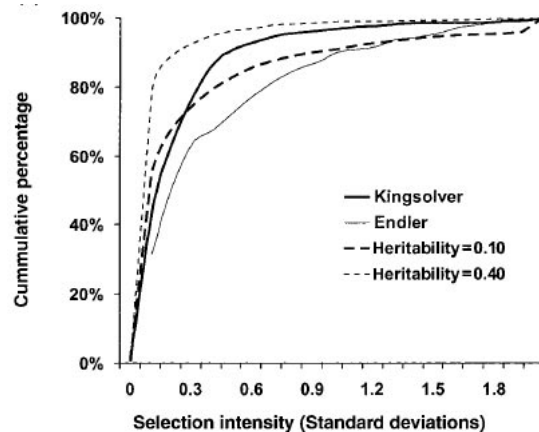

Figure 29.8. Four estimates of the distribution of selection intensities in nature. Two are

direct estimates of selection from studies in nature (Endler 1986, Kingsolver et al. 2001), while the other two are estimated using observed rates of microevolution and then estimating $\overline{\imath}$ using two different assumed heritability values. After Kinnison and Hendry (2001).

A final perspective on the strength of directional selection was offered by Hendry and Kinnison (1999) and Kinnison and Hendry (2001), who compiled data on over 2000 estimated rates of microevolution. Of course, such rates of divergence confound the strength of selection with the genetics of the trait. Strong selection may still result in a weak response if insufficient genetic variation is present. As we will see in Chapters 30 and 31, even if all of the component traits under selection have significant heritabilities, there may be little variation in the actual *direction* that selection is trying to move this vector of traits. However, a strenght of this approach is that such estimates focus on an average rate of change over time, smoothing out large values caused by brief episodes of strong selection. Such episodes are more likely to catch our attention and considering only them results in a very biased view of the average strength of selection on a trait over time. Figure 29.8 plots the cumulative selection intensities obtained by Endler (1986) and Kingsolver et al. (2001), as well as the inferred average values given the observed rates of microevolution, assuming heritabilities of 0.1 and 0.4. As can be seen, under either assumed heritability, there is an excess of weak selection (small $\overline{\imath}$) values relative to those seen by Kingsolver and Endler. While this does suggest that weak selection is the norm, an argument for strong selection can also be made. While strong selection can certainly occur over brief episodes, if it continues over significant time, genetic variation for response is quickly eroded and further response must wait for new variation (from mutation and immigration). Thus, with strong persistent selection, h^2 would likely become much smaller than 0.1. On the other hand, if selection is episodic, selection to reduce h^2 is smaller and hence genetic variation erodes less quickly. Thus, an apparent excess of small selection intensities could reflect true low selection intensities when averaged over many generations, smoothing an strong episodes. However, it could equally likely reflect strong persistent selection quickly eroding heritability and hence reducing response.

Quadratic Selection: Strong or Weak?

While the above concerns can result in biased overestimates of $|\gamma_{ii}|$, there are two countering factors that tend to underestimate $|\gamma|$. As mentioned earlier, many published reports likely have made errors in obtaining γ_{ii} values from the output of standard regression packages, resulting in only half their true value being reported (Stinchcombe et al. 2008). Second, as we have seen, Blows and Brooks (2003) point out that γ_{ii} likely significantly *underestimate* the strength of quadratic selection when multiple traits are considered. An analysis based on the eigenvalues λ of $\boldsymbol{\gamma}$ is much more informative as to the strength of quadratic selection in nature.
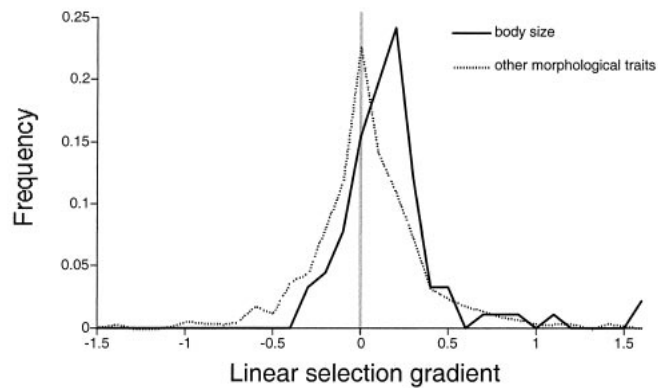
Figure 29.9. Distribution of standardized selection gradients for morphological traits and for size. While the distribution for morphological traits is largely symmetric around zero, the distribution for body size is skewed for positive values. After Kingsolver and Pfenning (2004).

Directional Selection on Body Size and Cope's Law

A very interesting analysis of the Kingsolver data set was by Kingsolver and Pfenning (2004), who examined the distribution of β' values for body size. An intriguing evolutionary observation is **Cope's law**, which is the tendency for the size of species within a lineage to increase over evolutionary time. As shown in Figure 29.9, Kingsolver and Pfenning found that while the distribution of β' values for morphological traits was symmetric about zero, the distribution for β' for body size was highly skewed towards positive values. Indeed, 79% of the estimated β' values for size were positive. Kingsolver and Pfenning thus suggest that individual selection for larger body size underlies Cope's law. This suggestion also presents a paradox: if true, why are most of the largest animals not present today? Kingsolver and Pfenning (2004, 2007) suggest that this is a consequence of differential extinction, pointing out that during the last widespread extinction of North American mammals that the largest were hit most heavily.

UNMEASURED CHARACTERS AND OTHER BIOLOGICAL CAVEATS

Even if we are willing to assume that the best-fitting quadratic regression is a reasonable approximation of the individual fitness surface, there are still a number of important biological caveats to keep in mind (Chapters 16 and 28). For example, the fitness surface can change in both time and space, often over short spatial/temporal scales (e.g., Kalisz 1986, Stewart and Schoen 1987, Scheiner 1989, Jordan 1991, Garant et al. 2007), so that one estimate of the fitness surface may be quite different from another for a different time and/or location (Figures 29.10 and 29.11). Hence, considerable care must be taken before pooling data from different times and/or sites to improve the precision of estimates. When the data are such that selection gradients can be estimated separately for different times or areas, space/time by gradient interactions can be tested in a straightforward fashion (e.g., Mitchell-Olds and Bergelson 1990). Further, the evolution of a trait may result in a change in the biotic environment, which in turn may change the nature of selection on that trait (Chapter 16).
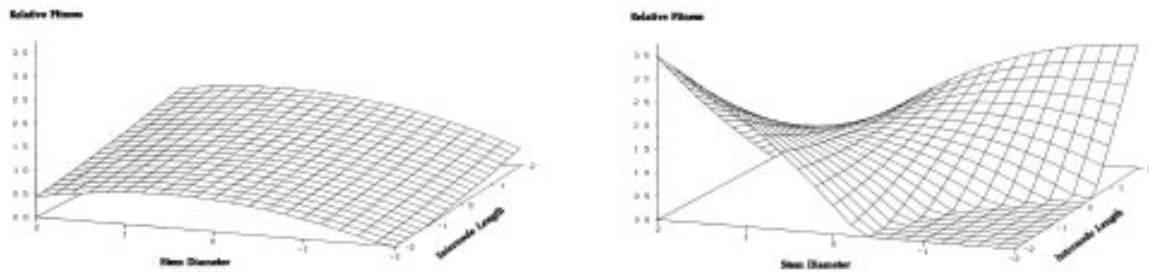
Figure 29.10.   The selection surface can change over time. These two surfaces are the best-fitting quadratic surface for relative fitness as a function of stem diameter and internode length for a population of the annual plant *Diodia teres* in North Carolina. **Left**. Estimate for June 1985 data. **Right**. Estimate for July 1985 data. From Jordon (1991).

Population structure can also influence fitness surface estimation. If the population being examined has overlapping generations, fitness data must be adjusted to reflect this (Chapters 23, 28, 36). Likewise, if members of the population differ in their amount of inbreeding, measured characters and fitness may show a spurious correlation if both are affected by inbreeding depression, which in turn can inflate the strenght of direction and convex selection (Willis 1996). Finally, immigration must be accounted for (e.g., Garant et al. 2005), especially if there are different selection regimes within the area that encompasses the study population (e.g., Figure 29.11).
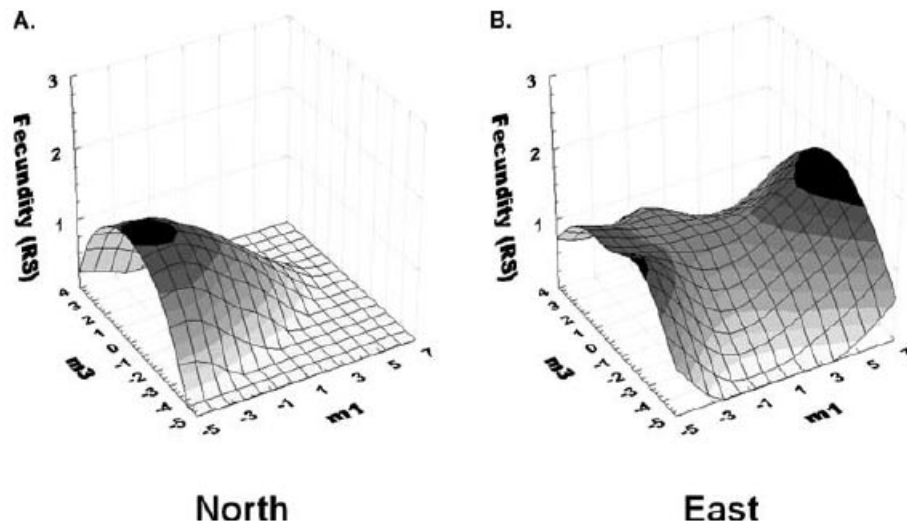

Figure 29.11.   The selection surface can change over space. The fecundity surface for great tits (*Parus major*) in two populations in Oxfordshire (U.K.) separated by less than four kilometers. From Garant et al. (2007).

Perhaps the most severe caveat for the regression approach of estimating $w(z)$ is un-measured characters — estimates of the amount of direct selection acting on a character are

biased if that character is phenotypically correlated with unmeasured characters also under selection (Lande and Arnold 1983, Mitchell-Olds and Shaw 1987). Adding one or more of these unmeasured characters to the regression can change initial estimates of β and γ. Conversely, selection acting on unmeasured characters that are phenotypically *uncorrelated* with those being measured has no effect on estimates of β and γ.

As detailed at some length in Chapter 16, environment-fitness correlations are a special (and potentially very important) example of an unmeasured "trait" (here an environmental factor) that is correlated with both the trait(s) under consideration as well as fitness. If one can obtain estimates of the genotypic values of individuals (either by using clones or sib, the later taken as mean family value), then genotype-fitness associations can be directly tested (Rausher 1992, Stinchcombe et al. 2002, Scheiner et al. 2002). While such as approach can be useful, when a pedigree is available, one can replace estimated genotypic values (which are very imprecise, especially when based on sib means) with estimated breeding values (EBVs) obtained from BLUP estimates (Chapter 16). Although powerful, there are some caveats with this approach (Postma 2006, Hadfield 2008). Recall that since BLUP estimation results in EBVs tending to be shrunk back towards the mean, the variance of EBVs is less than the additive genetic variance (Postma 2006). This results in directional selection gradients based on EBVs being inflated over their true values (Equation 16.21b). Likewise, this reduction in the variance of EBVs results in a bias towards convex selection on breeding values. Hence, the use of EBVs in quadratic regression is not recommended.

PATH ANALYSIS AND FITNESS ESTIMATION

Recall from LW Appendix 2, that path analysis and multivariate regressions are very closely connected, with path coefficients simply being standardized regression coefficients for the linear model suggested by the **path diagram**. This diagram in essence represents a *hypothesis* about the structure of causality. Regression methods and path analysis offer complementary approaches for examining associations between phenotypes and fitness. The purpose of a regression analysis is to predict fitness given character value, while path analysis provides a description of the nature of character covariances and how they interact with fitness. While regressions simply reply on the correlation among traits, path analysis examines the structure of these correlations, and a number of authors have applied this approach to the analysis of natural selection (e.g., Maddox and Antonovics 1983, Mitchell-Olds 1987, Crespi and Bookstein 1988, Crespi 1990, Kingsolver and Schemske 1991, Weis and Kapelinski 1994).

The key distinction between a regression and a path analysis of selection is that a Lande-Arnold regression attempts to account biologically for the covariance between fitness and the measured traits, while a path analysis further attempts to account for the covariance *among* the measured traits. To see an example of this, suppose traits are ordered in time, such as germination g -> vegetative growth rate r -> flowering time f. A regression analysis simply uses each pair-wise correlation without using the ordered structure inherent in the variables, while a path analysis explicitly uses this information (Figure 29.12, Example 29.7).
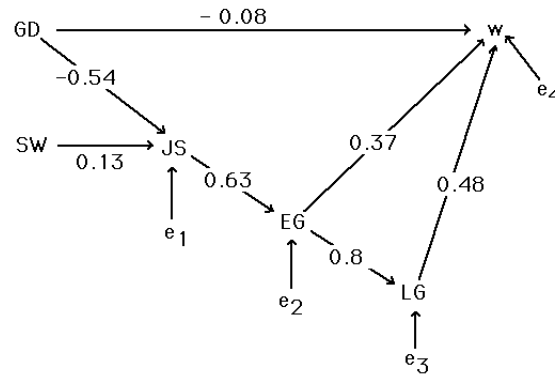
Figure 29.12.   An example of the use of path analysis in describing multivariate selection is Mitchell-Olds and Bergelson (1990), who examined fitness (measured by adult size) for the annual plant *Impatiens capensis*. Germination date (GD), seed weight (SW), June size (JS), early growth rate (EG), and late growth rate (LG) were the measured characters. Only significant paths are shown, with the standardized partial regression coefficient (or path coefficient) given above each path, where the square of these coefficients in the correlation between the connected paths. After Mitchell-Olds and Bergelson (1990).

Example 29.7.   Mitchell-Olds and Bergelson (1990) measured fitness (using adult size as a proxy) in the annual plant *Impatiens capensis* as a function of five characters: seed weight, germination date, June size, early growth rate, and late growth rate. Figure 29.12 displays the significant paths between these variables and fitness. Note that the path diagram provides a description of the actual nature of the correlations, in particular, the causal connections assumed between variables. From this path diagram, we can examine the contribution to relative fitness from the direct effect of a character and from its indirect effects through its effect on other characters. For example, the direct effect of early growth is 0.37 (the path $EG \rightarrow w$). Thus the correlation coefficient between the direct effect of EG and fitness is 0.37, so that this path accounts for 13 percent of the variation in fitness (as $0.37^2 \sim 0.13$). Early growth also influences fitness through an indirect path in that it influences late growth rate which in turns has a direct effect on fitness ($EG \rightarrow LG \rightarrow w$), which has product $0.8 \cdot 0.48 = 0.38$. Thus the total effect of early growth on fitness is the sum of the direct and indirect effects, or 0.75. Considering the other characters,

Character	Direct Effect	Indirect Effect
Seed weight	0.04	0.10
Germinate date	−0.08	−0.32
June Size	−0.02	0.52
Early growth rate	0.37	0.38
Late growth rate	0.48	0.00

showing that indirect effects are more important that direct effects for most characters.

Thus when traits are ordered in time (such as a developmental pathway or an ontogeny), a path analysis may provide additional insight into the nature of selection beyond a simple Lande-Arnold regression. Another situation where a path analysis may prove useful is when biotic factors, such as other species, are the agents of selection on our focal trait(s). While

we have framed multi-trait selection in the context of a number of traits in an individual and how these influence fitness, an equally common situation is when the "traits" are not properties of an individual, but rather biotic agents, such as the predators, competitors, or pollinators of the focal species. Path analysis allows one to examine the interactions among these multiple components and how they influence the fitness of the trait in the focal species.

Example 29.8.   Weis and Kapelinski (1994) examined the nature of selective forces acting on gall size of the Tephritid fly *Eurosta solidadginis*, whose larvae makes a protective gall on goldrenrods. Larvae residing in small galls are susceptible to the parasitoid wasp *Eurytoma gigantea*, whose ovipositor can reach the larvae through small, but not large, galls. Countering this, when galls become sufficiently large, they are preyed upon by insectivorous birds (mainly downy woodpeckers in the study area in Pennsylvania). Thus, there is selection pressure on both large and small galls, but through different biotic agents. Two other parasitoids, a second wasp and a beetle, also feed on larvae within galls, but size does not appear to be a factor. However, galls attacked by these two insects appear to be distasteful, with the resulting parasitoids not eaten by birds. Thus, this is a complex system with different agents, and the possibility of frequency-dependent search images, all acting to impart selection on the focal trait, gall size. The authors examined two different path diagrams for selection intensity on gall size, one representing a conventional model of selection, with gall density and mean gall size included in the model. A second (search-image based) model replaces these two measures with the density of small and density of large galls. Under the conventional model, there is a strong (but not significant) negative association (-0.43) between gall size and attacks by *Eurytoma* and a strong (and significant) association (0.37) between mean gall size and bird attacks. Conversely, under the frequency-dependent model, there was no association between small or large gall density and *Eurytoma* attacks, but a strong (and significant) association between density of small galls (-0.50) and density of large galls (0.46) and bird attacks.

A final use of path analysis is when there are unmeasured traits. One common concern is that apparent selection on a morphological trait(s) may simply be a reflection of selection on size (Figure 29.13). Crespi and Bookstein (1989) and Crespi (1990) suggest using the measured morphological traits to extract the first principal component (PC 1), which as typically a generalized measure of size. The analysis proceeds by using PC 1 as a measure of size and then considering the residual values for the traits when the effects of PC 1 are removed. Care must be taken, however, if the traits show any allometry other than **isometric growth** (shape is independent of size, see LW Chapter 11). In such cases, PC 1 can contain both size *and* shape information. Jolicoeur (1963) and Somers (1986) suggest in this case that one extract PC 1 from the *correlation* (as opposed to covariance) matrix for log-transformed traits. When isometric growth is present, the eigenvector corresponding to the first PC should have all traits weighted equally (Jolicoeur 1963), with (for n traits) each element in the vector being $1/\sqrt{n}$, which provides one gauge of whether the assumption of isometric growth is appropriate.
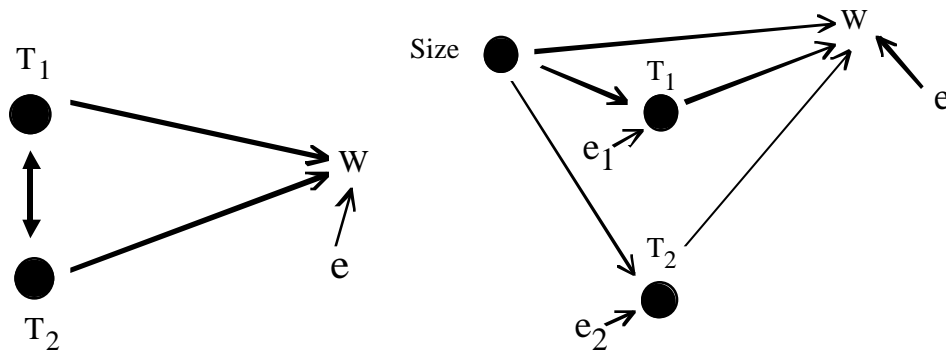
Figure 29.13.   Morphological traits T_1 and T_2 are being examined for their effects on fitness w. **Left:** Under a standard Lande-Arnold regression, we allow for direct effects of both traits on fitness w as well as for indirect effects from correlations between these two traits (double-headed arrow). Here e represents the additional variance in fitness not accounted for by the paths through T_1 and T_2. **Right:** Suppose that size influences both traits as well as fitness. When we ignore size, it is possible that all of the correlations between trait value and fitness arise because each trait is serving as a surrogate for size. Since this is a missing trait (size is not included in the analysis), this effect would not be picked up by a Lande-Arnold regression. By extracting a surrogate measure for size (such as PC 1) from the data, we can include this in the analysis. The path diagram allows for direct paths between size and fitness, and between both trait and fitness, after size effects have been removed.

LEVELS OF SELECTION

Finally, while our focus (thus far) has been on predicting fitness given the phenotype of an individual, it is also possible that selection is also acting at levels above the individual, such as **group-level effects** (Chapter 18). For example, Breden and Wade (1989) found that larval group size in the Imported Willow Leaf Beetle (*Plagiodera versicolora*) increases individual survivorship, which each additional group member increasing fitness by around seven percent. Here we briefly examine the analysis of selection at multiple levels. While one can use Price's equation to decompose the selection differential on individual value into within- and between-group components, as we show this can be misleading. Akin to selection differentials being unreliable indicators of targets of selection when traits are correlated, so to is the case when the traits are individual and group values. As above, the amount of direct selection on any trait can be determined by the appropriate multiple regression.

Contextual Analysis

Heisler and Damuth (1987) proposed that any effects of selection acting at some level above the individual can be estimated by using the method of **contextual analysis**, which is widely used in the social sciences (e.g., Boyd and Iversen 1979). This is a regression-based approach and amounts to simply adding group-level traits to a Lande-Arnold regression. Traits scored at the level of groups can be **aggregate characters**, simple functions of the individual phenotypes within a group (such as the group mean) or they can be **global** or **emergent characters** which can only be defined at the group level (such as number of group members and group density). As the following example illustrates, incorporating group–level traits into the Lande-Arnold fitness regression is very straightforward.

Example 29.9. Aspi et al. (2003) examined the effects of selection on the riparian plant *Silene tatarica*, a threatened species from Finland in the family Caryophlleaceae. Plants tend to grow in patches and the authors envisioned that plant density within a path might influence both pollinator visits and herbivory. The individual traits they measured were plant height (z_1) and number of stems (z_2), while two aggregate traits were measured (the mean of each trait for the patch) along with group-level trait of plant density d. The resulting regression for predicting the relative fitness of individual j in patch i is

$$w_{ij} = 1 + \beta_1 z_{1,ij} + \beta_2 z_{2,ij} + \beta_3 \overline{z}_{1,i} + \beta_4 \overline{z}_{2,i} + \beta_5 d_i + e_{ij}$$

The regression coefficients β_1 and β_2 correspond to direct selection on individual trait value, while β_3 and β_4 correspond to direct selection on the patch mean of each trait and β_5 to direct selection on the density within a patch. Aspi et al. standardized all variables, so that a one standard deviation change in the variable of interest results in an expected change of β in fitness. For 1999, the estimated regression coefficients over a sample of 922 individuals were

	Height	Mean height	Stem No.	Mean stem No.	Density
β	0.589***	0.653***	0.187**	−0.209***	0.631***

All β were significant, with ** denoting $p < 0.01$ and *** denoting $p < 0.001$. Note that (on a standardize scale) selection on group density is as strong as individual selection. Selection on height at the individual and group level was in the same direction, and the authors suggest this is due to pollinator attraction. However, selection on stems was in opposite directions, with individual selection to increase number of stems, but patch-level selection to decrease them. The authors suggest that patch-level selection may be due to herbivory by reindeer.

A contextual analysis need not stop at two levels, as the approach is easily extended to include additional hierarchical levels of population organization, and hence the potential for some selection at these levels. Indeed, in some settings the individual level is ignored, and only higher levels of organization are contrasted. One example of this is Banschbach and Herbers (1996), who compared selection on nest- versus colony-level traits in the forest ant *Myrmica punctiventris*, finding that fertility selection largely operates on the level of the nest as opposed to the level of multiple-nest colonies.

In many settings, group composition is fairly obvious, due to a patchy distribution of individuals in space. However, in some situations, appropriate set of individuals to include in a group can be very unclear. Indeed, for populations where indivdiuals appear to be continuously in space with little obvious breaks, group-level effects may be present when very close individuals are included, but largely disappear as the defined group becomes larger. In theory, one could assign different group sizes and use model-selection approaches to determine the group structure that gives the best fit of the fitness regression. However, care must be taken to distinguish a statistical fit from biological reality. A very reasonable biological definition of a group was offered by Uyenonyama and Feldman (1981), namely the set of all individuals that influence the fitness of a focal individual.

An alternative to using group means is to model neighbor interactions influencing fitness (e.g., Equation 18.59a). In such cases, the fitness of a focal individual i can be written (for a single trait) as

$$w_i = 1 + \beta_1 z_i + \beta_2 \left(\sum_{i \neq j} z_j \right) + e_i \tag{29.30a}$$

The difference with this approach is that group mean (which contains the value of the focal individual) is replaced by the interactions caused by the neighbors of the focal individual.

This model is easily connected to the standard selection on group means model since

$$\sum_{i \neq j} z_j = n\overline{z} - z_i$$

when i interacts with $n - 1$ neighbors. Hence, we can write Equation 29.30a as

$$w_i = 1 + \beta_1 z_i + \beta_2 \left(n\overline{z} - z_i \right) + e_i$$
$$= 1 + \left(\beta_1 - \beta_2 \right) z_i + \beta_2 \, n\overline{z} + e_i \qquad (29.30b)$$

The (subtle) distinction between the group mean and neighborhood fitness models is apparent in the case where only the neighbors influence the fitness of an individual. In this case, $\beta_1 = 0, \beta_2 \neq 0$, so that in the neighborhood fitness model there is no weight on individual value. However, if a group mean fitness model is fitted to this same data, there is a non-zero regression slope on individual fitness (slope = $-\beta_2$), reflecting the input of an individual's value to the group mean.

Selection Can Be Anatagonistic Across Levels

When selection is operating at both the individual and group levels, there is no *a priori* reason why the direction of selection should be the same (or different!) at the two levels. When the direction of selection is the same, and group-level effects are ignored, the effects of individual selection are overestimated. Conversely, the more interesting case is when individual and group level selection are **antagonistic**, working in opposite directions. Example 29.9 shows both cases: for height, selection at both levels was in the same direction, while for stem number, selection was antagonistic, being in opposite directions at the individual and group levels. Antagonistic selection is commonly seen in the limited number of studies looking at multilevel selection, although in part this might be due to a bias in choosing those traits that might be excepted to be under differential individual vs. group selection pressure.

    One of the first applications of contextual analysis to levels-of-selection was by Stevens et al. (1995), using the North American plant Jewelweed (*Impatiens capensis*). They found that an overall measure of size was selected to decrease at the group level but increase at the individual level. This pattern was observed for both survival rate to two years and seed production in cleistogamus (open-pollinated flowers). The individual and group partial regression slopes (variance standardized over trait values) for survival rate on overall size were $\beta_I = 1.74$ and $\beta_{group} = -3.03$. Hence, group selection was stronger (and of opposite sign) than individual selection. The partial regression slopes for cleistogamus seed production on size were $\beta_I = 0.51$ and $\beta_{group} = -0.52$. Similar observations were made by Weinig et al. (2007), who used a set of RILs in *Arabidopsis thaliana*, finding that two composite traits (size and elongation) were favored to increase under directional selection but to decrease under group selection. For both traits, individual was stronger than group selection (the β were roughly twice as large). A final example is Tsuji (1995), who worked with the Japanese ant *Pristomyrmex pungens*, an unusual species with no queens, with workers able to reproduce parthenogenetically. Larger size was favored at the individual level ($\beta_I = 0.07$), but selected against at colony level ($\beta_{group} = -0.11$).

Group Selection Is Likely Density-Dependent

A number of workers (Goodnight et al. 1992; Donohue 2003, 2004) suggested that at low density, group selection may be weak (or essentially absent), with individual selection dominating. However, as group density increases, so does competition, and hence so do group effects. An empirical observation from plant ecology, the **law of constant final yield** (Harper

1977, Weiner and Freckleton 2010), has been offered as support for this view (Goodnight et al. 1992). Plant ecologists noted that total yield of a group initially increases with low density, but after a certain critical density is reached, the total biomass of the group remains roughly constant. Even though more plants are now in the group, their individual contribution has decreased. Goodnight et al. suggest that this law results from a balance between group and individual selection. One consequence of potential density-dependence at the group level is that an analysis of individual selection that ignores a group selection component can be mistaken for frequency-dependent selection (Uyenonyama and Feldman 1980, 1981; Damuth and Heisler 1988).

Working with the mustard plant *Cakile edentual* (Brassicaceae) with the endearing name the Great Lake Sea Rocket, Donohue (2004) did indeed observe that plant density influenced the relative strengths of individual vs. group selection. For plant height and stem mass, she observed antagonistic individual vs. group selection at varying densities, with individual selection favoring smaller plants with more stem mass, while group selection favored larger plants with less stem mass. At all densities, individual selection remained significantly stronger than group selection. Surprisingly, group selection was strongest at the intermediate, as opposed to the high, density. Weinig et al. (2007) also found a density-dependent effect in *Arabidopsis thaliana*, where the density must be above some threshold before group selection is important.

Selection Differentials Can Be Misleading in Levels of Selection

One might be tempted to use selection differentials to assess levels of selection, an approach suggested by Price (1972) and Wade (1985) who showed how to decompose the differential into individual and group components. They did so by recalling the Robertson-Price identity (Equation 10.7), which states that the within-generation change in a trait can be written as a covariance, $S = \sigma(w, z)$. This also holds more generally for levels of selection. Let w_{ij} and z_{ij} denote the fitnesses and phenotypes of the jth individual in group i. Using the definition of the covariance, Price showed that the total selection differential can be decomposed as

$$\sigma(z, w) = E_i\big[\sigma(w_{ij}, z_{ij})\big] + \sigma(\overline{z}_i, w_{i\cdot}) \tag{29.31}$$

The first term is the covariance between individual phenotype and fitness within a group, and we take its average over all groups to give the within-group covariance between individual value and fitness. The second term is the covariance between the mean trait value $\overline{z}_i$ and the mean fitness $w_{i\cdot}$ of the group. Equation 29.31 decomposes the selection differential into the differential from individual effects within groups plus the differential for group effects.

Motivated by Equation 29.31, Price suggested that group selection implies a non-zero covariance between the mean trait value and mean fitness of groups (i.e., $\sigma(\overline{z}_i, w_{i\cdot}) \neq 0$). Under this definition, group selection cannot occur when there are no between-group differences in mean fitness. While this sounds reasonable, as we will see shortly, there are indeed situations where each group has the same mean fitness, but there is still selection on the group means. Hence, group selection can occur even when the covariance between mean trait value and mean fitness is zero. Likewise, even when there is selection only on individual phenotype, if group means differ, then (as we will show shortly) one can easily get a non-zero value for $\sigma(\overline{z}_i, w_{i\cdot})$.

Hence, this **covariance definition** (a non-zero group selection differential) for group selection is very misleading, as one can have a zero group selection differential when group selection *is* occurring, and a non-zero group selection differential when it is not. The root cause of this is a familiar problem: selection on phenotypically-correlated traits can modify the selection differential of a focal trait, in the extreme showing an indirect response (non-zero S) when there is no direct selection on the trait. This also applies when considering

selection differentials on group means. Indeed, by analogy to using a multiple regression to control for correlated traits, this was the motivation of Heisler and Damuth (1987, also Goodnight et al. 1992) for using contextual analysis. The partial regression coefficients in a contextual analysis control for any indirect effects of other correlated traits included *within the analysis*.

Example 29.10:   Recall the previous example, where Aspi et al. (2003) applied contextual analysis to height and stem number in *Silene tatarica*. They reported the within- and between-patch selection differentials as

Trait	Within-patch S	Between-patch S
Height	0.485***	0.025 ns
No. Stems	0.855***	-0.006 ns

The use of selection differentials correctly determines positive directional selection on individual height and stem number. However, they also suggest no between-patch effects (no selection on patch means), when in fact contextual analysis shows these are highly significant for both traits.

One consequence of indirect effects generating non-zero covariances (and hence non-zero selection differentials) is that an analysis restricted to one level of selection can be very misleading. For example, suppose an investigator assumes that group selection might be important, and only includes group means in their analysis. For a single trait, they might consider the slope of a simple (i.e., univariate) regression of mean fitness on group trait mean,

$$w_{i.} = 1 + \beta \, \overline{z}_i + e_i \tag{29.32a}$$

Recalling (LW Chapter 3) that the slope of a univariate regression is the covariance divided by variance of the predictor variable, the slope becomes

$$\beta = \frac{\sigma(\overline{z}_i, w_{i.})}{\sigma^2(\overline{z}_i)} \tag{29.32b}$$

Hence, a non-zero covariance implies a non-zero slope, and using this simple univariate regression, an investigator would conclude that group selection is occurring on this trait. In reality, as we see below, strict individual selection can generate a group-level covariance, and hence (in this case) a spurious assignment of a group-level effect to selection. By using a multiple repression including both individual and group mean values, a contextual analysis entirely corrects for this (provided selection is limited to only these two levels).

Hard, Soft, and Group Selection: A Contextual Analysis Viewpoint

Evolutionary biology has had an unfortunately history of, at times, being constrained by thinking about complex evolutionary processes in terms simple catch phrases. We have seen examples of this with kin versus group selection, where Chapter 18 shows that these are special cases of the more general problem of selection with interacting individuals. Likewise, in this (and the previous) chapter, we have seen that the widely-used terms of stabilizing and distributive selection are special cases of the more general situation of a quadratic term in the individual fitness surface. It should therefore not be surprising that there is some confusion with the exact meaning of group selection in the more general context of multilevel selection

(reviewed by Okasha 2004). In the Price covariance framework, group selection is indicated by a non-zero covariance between mean fitness and mean trait value of a group, so that $\sigma(\overline{z}_i, w_{i\cdot}) \neq 0$. This line of thought also implies that for group selection to exist, there must be variance across groups in mean fitness ($\sigma(W_{i\cdot}) \neq 0$). However (following Goodnight et al. 1992), we will now show that these covariance conditions for defining group selection are misleading. One can have group selection in the absence of variance in mean fitness across groups and strict individual selection can still generate a covariance between group mean fitness and group mean trait value. Just as the more general framework of section in interacting populations removes much of the historical confusion between group vs. kin selection, multilevel selection removes much of the confusion over whether or not group selection is occurring. Recall that under multilevel selection, we simply have that the fitness of a focal individual is also influenced by other individuals, and hence selection has both individual and group components. The later are very clearly defined by non-zero regression coefficients on group traits in a contextual analysis. As succinctly stated by Goodnight et al (1992), "*With contextual analysis, whether selection is acting at a particular level becomes a statistical rather than a philosophical question.*"

Wallace's (1968, 1976) distinction between hard and soft selection serves to highlight the differences between the covariance vs. multiple regression definitions of group selection (Goodnight et al. 1992). **Hard selection** occurs when the fitness of a focal individual is solely determined by its phenotype and is unaffected by any group members. This results in the mean fitness of groups differing, with some groups leaving more offspring than others. Conversely, **soft selection** occurs when each group contributes equally to the next generation, so that group mean fitness is a constant and there is no variance in mean fitness between groups. The law of constant yield is an example of soft selection, with fitness (number of seeds) being largely independent of patch density. Figure 29.14 shows the different consequences of hard, soft, and (strict) group selection on the within- and between-group regressions of phenotype on fitness.
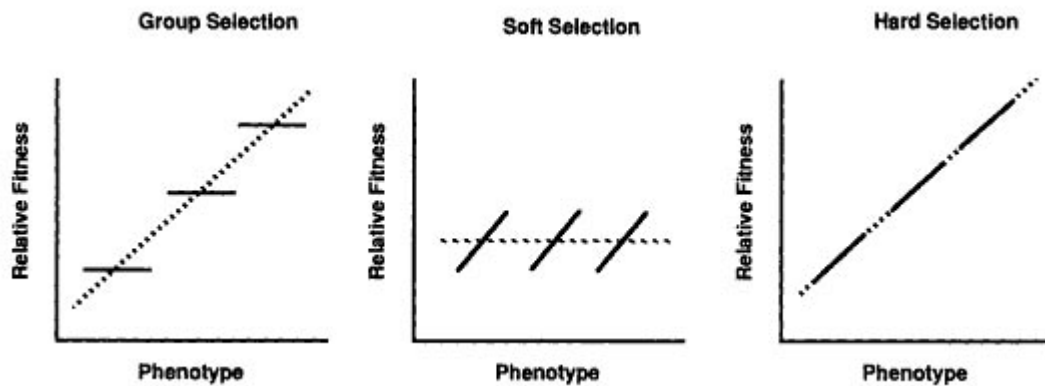

Figure 29.14:   The regression of fitness on trait value within each group (solid line) and between the group means (dotted line). Under strict group selection (**left**), phenotype within a group has no predictive value on fitness (within-group regression slope of zero), while the regression of fitness on group means is positive (non-zero regression of individual fitness on group mean). Under soft selection (**center**), fitness is entirely determined by the within-group deviation. The result is a significant within-group regression but no between-group regression. Finally, under hard selection (**right**) fitness is entirely a function of phenotype, so that both the within-group and between-group regressions have the same slope. After Goodnight et al. (1992).

Table 29.2.   Hard, soft, and strict group selection can all be expressed as special cases of a more general contextual analysis. Model one gives the regression coefficients on individual value z_{ij} and group mean $\overline{z}_i$ for the contextual regression given by Equation 29.33a, while Model two gives the regression coefficients when the contextual analysis is framed in terms of within-group deviations and group means (Equation 29.33b). To contrast these results with those from the covariance criteria for group selection, the table also shows whether the type of selection results in a non-zero within-group (S_I) or between-group (S_g) selection differential and whether the between-group variance in mean fitness $\sigma(W_{i\cdot})$ is non-zero.

Regression	Model 1		Model 2				
	z_i	$\overline{z}_i$	$(z_{ij} - \overline{z}_i)$	$\overline{z}_i$	$S_I \neq 0$	$S_g \neq 0$	$\sigma(W_{i\cdot}) \neq 0$
Hard	β	0	β	β	Yes	Yes	Yes
Soft	β	$-\beta$	β	0	Yes	No	No
Strict Group	0	β	0	β	Yes	Yes	Yes

As summarized in Table 29.2, hard, soft, and strict group selection can all be modeled as special cases of a contextual analysis regression. For simplicity, we focus on a single trait, with the resulting regression having the form

$$w_{ij} = 1 + \beta_1 z_{ij} + \beta_2 \overline{z}_i + e_{ij} \tag{29.33a}$$

We can also rewrite this regression in terms of selection on the within-group deviation plus selection on the group means. In terms of the partial regression coefficients in Equation 29.33a, this becomes

$$w_{ij} = 1 + \beta_1 (z_{ij} - \overline{z}_i) + \beta_2 \overline{z}_i + e_{ij} \tag{29.33b}$$

$$= 1 + \beta_1 z_{ij} + (\beta_2 - \beta_1) \overline{z}_i + e_{ij} \tag{29.33c}$$

Hard selection occurs when selection is entirely on individual value, so that $\beta_2 = 0$ and hence there is no group component to multilevel selection. However, since the mean of random groups likely differs, this generates variance in group mean fitness and also generates a non-zero covariance between group mean trait value and mean fitness (Goodnight et al 1992). Under these criteria, one would infer group selection, even though it is absent. A non-zero between-group selection differential arises as a correlated response to direct selection on individual value. In this simple case, the contextual analysis clearly shows no multilevel selection, but a simple univariate regression might suggest otherwise (e.g., Equation 29.32a).

The opposite conclusion arises with soft selection. Here selection is entirely based on within-group deviations (such as within-family selection, Chapter 17). Hence, $\beta_1 = -\beta_2$ as the regression of fitness becomes

$$w_{ij} = 1 + \beta (z_{ij} - z_{i\cdot}) + e_{ij}$$

$$= 1 + \beta z_{ij} - \beta z_{i\cdot} + e_{ij} \tag{29.34a}$$

Under soft selection, there is no variation in mean fitness across groups and $\sigma(\overline{z}_i, w_{i\cdot}) = 0$. Hence, under Price's covariance definition, one might infer no group selection. However, there is clearly multilevel selection occurring, as $\beta_2 \neq 0$.

Finally, consider strict group selection, when an individuals' fitness is entirely a function of its group mean,

$$w_{ij} = 1 + \beta z_{i\cdot} + e_{ij} \tag{29.34b}$$

while written here as a univariate regression, formally we have $\beta_1 = 0$ if the regression given by Equation 29.33a is fitted. However, if one simply fits an individual selection model

($w_{ij} = 1 + \beta\, z_{ij} + e_{ij}$), there is a non-zero regression slope if group effects are ignored, and hence $E_i\big[\sigma(w_{ij}, z_{ij})\big] \neq 0$ (Goodnight et al 1992). Again this occurs because individual value and group mean are correlated, and hence direct selection on one results in indirect selection (and hence a covariance) in the other. With contextual analysis, we reach the correct conclusion of selection only on group values.

Literature Cited

Arnold, S. J., M. E. Pfrender, and A. G. Jones. 2001. The adaptive landscape as a conceptual bridge between micro- and macroevolution. *Genetica* 112-113: 9–32. [29]

Aspi, J., A. Jäkäläniemi, J. Tuomi, and P. Siikamäki. 2003. Multilevel phenotypic selection on morphological characters in a metapopulation of *Silene tatarica. Evolution* 57: 509–517. [29]

Banschbach, V. S. and J. M. Herbers. 1996. Complex colony structure in social insects. II. Reproduction, queen-worker conflict, and levels of selection. *Evolution* 50: 298–307. [29]

Bisgaard, S., and B. Ankenman. 1996. Standard errors for the eigenvalues in second-order response surface models. *Technometrics* 38: 238–246. [29]

Blows. M. W. 2007a. A tale of two matrices: multivariate approaches in evolutionary biology. *J. Evol .Biol.* 20: 1–8. [29]

Blows. M. W. 2007b. Complexity for complexity's sake? *J. Evol .Biol.* 20: 39–44. [29]

Blows, M. W., and R. Brooks. 2003. Measuring nonlinear selection. *Amer. Natl.* 162: 815–820. [29]

Blows, M .W., R. Brooks, and P. G. Kraft. 2003. Exploring complex fitness surfaces: multiple ornamentation and polymorphism in male guppies. *Evolution* 57: 1622–1630. [29]

Blows, M. W., S. F. Chenoweth, and E. Hine. 2004. Orientation of the genetic variance-covariance matrix and fitness surface for multiple male sexually selected traits. *Amer. Natl.* 163: 329–340. [29]

Blows, M. W., and B. Walsh. 2008. Spherical cows grazing in Flatland: constraints to selection and adaptation. *In*, J. Van der Werk, H.-U. Graser, R. Frankham, and C. Gondro (Eds) *Adaption and fitness in animal populations: Evolutionary and breeding perspectives on genetic resource management*, pp 82 – 102. Springer. [29]

Box, G. E. P. and N. R. Draper. 1987. *Empirical model-building and response surfaces*. John Wiley and Sons, New York. [29]

Box, G. E. P., and J. S. Hunter. 1954. A confidence region for the solution of a set of simultaneous equations with an application to experimental designs. *Biometrika* 41: 190–199. [29]

Boyd, L. H., and G. R. Iversen. 1979. *Contextual analysis: concepts and statistical techniques.* Wadsworth, Belmont, Califorina. [29]

Breden, F. and M. J. Wade. 1989. Selection within and between kin groups of the imported willow leaf beetle. *Amer. Natl.* 134: 35–50. [29]

Brodie, E. D. 1992. Correlational selection for color pattern and antipredator behavior in the garter snake *Thamnophis ordinoides*. *Evolution* 46: 1284–1298. [29]

Brodie, E. D. III, A. J. Moore, and F.J. Janzen. 1995. Visualizing and quantifying natural selection. *Trends in Ecol. Evol.* 10: 313–318. [29]

Brooks, R., and J. A. Endler. 2001. Direct and indirect sexual selection and quantitative genetics of male traits in gupppies (*Poecilia reticulata. Evolution* 55: 1002 - 1015. Brooks and Endler (2001) [29]

Bumpus, H. C. 1899. The elimination of the unfit as illustrated by the introduced House Sparrow,*Passer domesticus. Biol. Lectures, Marine Biol. Lab., Woods Hole*: 209-226. [28]

Cheverud, J. M. 1984. Quantitative genetics and developmental constraints on evolution by selection. *J. Theor. Biol.* 110: 155–171. [29]

Conner, J. K. 2001. How strong is natural selection? *Trends in Ecology and Evolution* 16: 215–217. [29]

Conner, J. K. 2007. A tale of two methods: putting biology before statistics in the study of phenotypic evolution. *J. Evol .Biol.* 20: 17–19. [29]

Crespi, B. J. 1990. Measuring the effect of natural selection on phenotypic interaction systems. *Amer. Natl.* 135: 32–47. [29]

Crespi, B. J. and F. L. Bookstein. 1988. A path-analytic model for the measurement of selection on morphology. *Evolution* 43: 18-28. [29]

Del Castillo, E., and S. Cahya. 2001. A tool for computing confidence regions on the stationary point of a response surface. *Amer. Stat.* 55: 358–365. [29]

Donohue, K. 2003. The infleunce of neighbor relatedness on multilevel selection in the Great Lakes Sea Rocket. *Amer. Natl.* 162: 77–92. [29]

Donohue, K. 2004. Density-dependent multilevel selection in the Great Lakes Sea Rocket. *Ecology* 85: 180–191. [29]

Endler, J. A., 1986. *Natural selection in the wild*. Princeton University Press, Princeton, New Jersey. [29]

Friedman, J. A, and W. Stuetzle. 1981. Projection pursuit regression. *J. Amer. Stat Assoc.* 76: 817–823. [29]

Garant, D., L. E. B. Kruuk, T. A. Wilkin, R. H. McCleery, and B. C. Sheldon. 2005. Evolution driven by differential dispersal within a wild bird population. *Nature* 433: 60–64. [29]

Garant, D., L. E. B. Kruuk, R. H. McCleery, and B. C. Sheldon. 2007. The effects of environmental heterogeneity on multivariate selection on reproductive traits in female great tits. *Evolution* 61: 1546–1559. [29]

Goodnight, C. J., J. M. Schwartz, and L. Stevens. 1992. Contextual analysis of models of group selection, soft selection, hard selection, and the evolution of altruism. *Amer. Natl.* 140: 743–761. [29]

Jolicoeur, P. 1963. The multivariate generalization of the allometry equation. *Biometrics* 19: 497–499. [29]

Jordan, N. 1991. Multivariate analysis of selection in experimental populations derived from hybridization of two ecotypes of the annual plant *Diodia teres*. *Evolution* 45: 1760–1772. [29]

Hadfield, J. D. 2008. Estimating evolutionary parameters when viability selection is operating. *Proc. Roy. Soc. B* 275: 723–734. [29]

Harper, J. L. 1977. *Population biology in plants.* Academic Press, New York. [29]

Hedges, L. V., and I. Olkin. 1985. *Statistical Methods for Meta-Analysis*. Academic Press. [29]

Heisler, I. L., and J. Damuth. 1987. A method of analyzing selection in hierachically structured populations. *Amer. Natl.* 130: 582–602. [29]

Hendry, A. P., and M. T. Kinnison. 1999. The pace of modern life: measuring rates of contemporary microevolution. *Evolution* 53: 1637–1653. [29]

Hereford, J., T. F. Hansen, and D. Houle. 2004. Comparing strenghts of directional selection; how strong is strong? *Evolution* 58: 2133–2143. [29]

Hersch, E. I., and P. C. Phillips. 2004. Power and potential bias in field studies of natural selection. *Eovlution* 58: 479–485. [29]

Hoekstra, H. E., J. M. Hoekstra, D. Berrigan, S. N. Vignieri, A. Hoang, C. E. Hill, P. Beerli, and J. G. Kingsolver. 2001. Strength and tempo of directional selection in the wild. *PNAS* 98: 9157–9160 [29]

Hunt, J., J. B. Wolf, and A. J. Moore. 2007. The biology of multivariate evolution. *J. Evol .Biol.* 20: 24–27. [29]

Hunter, J. E., and F. L. Schmidt. 2005. *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*. Sage . [28]

Kalisz, S. 1986. Variable selection on the timing of germination in *Collinsia verna (Scrophulariaceae)*. *Evolution* 40: 479–491. [29]

Kendall, M., and A. Stuart. 1983. *The advanced theory of statistics. Volume 3: Design and analysis, and time-series*. 4th Ed. Griffin, London. [29]

Kingsolver, J. G., H. E. Hoekstra, J. M. Hoekstra, D. Berrigan, S. N. Vignieri, C. E. Hill, A. Hoang, P. Gilbeert, and P. Beerli. 2001. The strenght of phenotypic selection in natural populations. *Amer. Natl.* 157: 245–261. [29]

Kingsolver, J. G., and D. W. Pfenning. 2004. Individual-level selection as a cause of Cope's rule of phyletic size increase. *Evolution* 58: 1608–1612. [29]

Kingsolver, J. G., and D. W. Pfenning. 2007. Patterns and power of phenotypic selection in nature. *BioScience* 57: 561–572. [29]

Kingsolver, J. G., and D. W. Schemske. 1991. Path analysis of selection. *Trends Ecol. Evol.* 6: 276–280. [29]

Kinnison, M. T., and A. P. Hendry. 2001. The pace of modern life II: from rates of contemporary microevolution to pattern and process. *Genetica* 112-113: 145–164. [29]

Knapczyk, F. N., and J. K. Conner. 2007. Estimates of the average strenght of natural selection are not inflated by samplng error or publication bias. *Amer. Natl.* 170: 501–508. [29]

Kruuk. L. E. B., and D. Garant. 2007. A wake-up call for studies of natural selection? *J. Evol. Biol.* 20: 30–33. [29]

Krzanowski, W. J. 1979. Between-group comparisons of principal components. *J. Amer. Stat. Assoc* 74: 703–707. [29]

Lande, R. 1976. Natural selection and random genetic drift in phenotypic evolution. *Evolution* 30: 314–334. [29]

Lande, R. 1979. Quantitative genetic analysis of multivariate evolution, applied to brain:body size allometry. *Evolution* 33: 402–416. [29]

Lande, R., and S. J. Arnold. 1983. The measurement of selection on correlated characters. *Evolution* 37: 1210–1226. [29]

Maddox, G. D. and J. Antonovics. 1983. Experimental ecological genetics in *Plantago*: a structural equation approach to fitness components in *P. aristata* and *P. patagonica*. *Ecology* 64: 1092–1099. [29]

Mercer, A. M. and P. R. Mercer. 2000. Cauchy's interlace theorem and lower bounds for the spectral radius. *Inter. J. Mat. Math Sci.* 23: 563–566. [29]

Mitchell-Olds, T. 1987. Analysis of local variation in plant size. *Ecology* 68: 82–87. [29]

Mitchell-Olds, T. 1989. FreeStat Users' Manual. Technical Bulletin No. 101, University of Montana Herbarium. Missoula, Mont. [29]

Mitchell-Olds, T., and J. Bergelson. 1990. Statistical genetics of *Impatiens capensis.* II. Natural selection. *Genetics* 124: 417–421. [29]

Mitchell-Olds, T. and R. G. Shaw. 1987. Regression analysis of natural selection: statistical inference and biological interpretation. *Evolution* 41: 1149–1161. [29]

Okasha, S. 2004. Mutilevel selection and the partitioning of covariance: A comparison of three approaches. *Evolution* 58: 486–494. [29]

Peterson, J. J., S. Cahya, and E. del Castillo. 2002. A general approach to confidence regions for optimal factor levels of response surfaces. *Biometrics* 58: 422–431. [29]

Phillips, P. C. and S. J. Arnold. 1989. Visualizing multivariate selection. *Evolution* 43: 1209–1222. [29]

Postma, E. 2006. Implications of the differences between true and predicted breeding values for the study of natural selection and micro-evolution. *J. Evol. Biol.* 19: 309–320. [29]

Rausher, M. D. 1992. The measurement of selection on quantitative traits: biases due to environmental covariances between traits and fitness. *Evolution* 46: 616–626. [29]

Rpsenberg, M. S. 2005. The file-drawer problem revisted: a general weighted method for calculating fail-safe numbers in meta-analysis. *Evolutio* 59: 464–468. [28]

Rosenthal, R. 1979. The "file drawer problem" and tolerance for null results. *Physchol. Bull.* 86: 638–641. [28]

Scheiner, S. M. 1989. Variable selection along a successional gradient. *Evolution* 43: 548–562. [29]

Scheiner, S. M., K. Donohue, L. A. Dorn, S. J. Mazer, and L. M. Wolfe. 2002. Reducing environmental bias when measuring natural selection. *Evolution* 56: 2156–2167. [29]

Schluter, D. 1988. Estimating the form of natural selection on a quantitiative triat. *Evolution* 42: 849–861. [29]

Schluter, D., and D. Nychka. 1994. Exploring fitness surfaces. *Amer. Natl.* 143: 597–616. [29]

Simms, E. L. 1990. Examining selection on the multivariate phenotype: plant resistance to heribvores. *Evolution* 44: 1177–1188. [29]

Simms, E. L., and M. D. Rausher. 1993. Patterns of selection on phytophage resistance in *Ipomoea purpurea. Evolution* 47: 970–976 [29]

Somers, K. M. 1986. Multivariate allometry and removal of size with principal components analysis. *Sys. Zoo.* 35: 359–368. [29]

Stevens, L., C. J. Goodnight, and S. Kalisz. 1995. Multilevel selection in natural populations of *Imptatiens capensis. Amer. Natl.* 145: 513–526. [29]

Stewart, S. C. and D. J. Schoen. 1987. Patterns of phenotypic variability and fecundity selection in a natural population of *Impatiens pallida*. *Evolutioon* 41: 1290–1301. [29]

Stinchcombe, J. R., A. F. Agrawl, P. A. Hohenlohe, S. J. Arnold and M. W. Blows. 2008. Estimating non-linear selection graidents using quadratic regression coefficients: double or nothing? *Evolution* 62: 2435–2440. [29]

Stinchcombe, J. R., M. T. Rutter, D. S. Burdick, P. Tiffin, M. D. Ruahser, and R. Mauricio. 2002. Testing for environmentally induced bias in phenotypic estimates of natural selection: theory and practice. *Amer. Natl.* 160: 511–523. [29]

Uyenoyama, M. K., and M. W. Feldman. 1980. Theories of kin and group selection: a populaion genetics perspective. *Theor. Pop. Biol.* 17: 380–414. [29]

Uyenoyama, M. K., and M. W. Feldman. 1981. On relatedness and adaptive topography in kin selection. *Theor. Pop. Biol.* 19: 87–123. [29]

van Tienderen, P. H. 2000. Elasticities and the link between demographic and evolutionary dynamics. *Ecology* 81: 666–679. [29]

Wade, M. J. 1985. Soft selection, hard selection, kin selection, and group selection. *Amer. Natl.* 125: 61–73. [29]

Wallace, B. 1968. Polymorphism, population size, and genetic load. *in* R. C. Lewontin (ed), }si Popual-tion biology and evolution,, pp. 87–108. Syracuse University Press, Syracuse, N. Y. [29]

Wallace, B. 1975. Hard and soft selection revisited. *Evolution* 29: 465–473. [29]

Walsh, B. 2007. Escape from flatland. *J. Evol. Biol.* 20: 36–38. [29]

Weis, A. E., and A. Kapelinski. 1994. Variable selection on *Eurosta*'s gall size. II. A path analysis of the ecological factors behind selection. *Evolution* 48: 734–746. [29]

Weiner, J., and R. P. Freckleton. 2010. Constant final yield. *Ann. Rev. Eco. Evol. Sys.* xx: xxx-xxx. [29]

Weldon, H. F. R. 1901. A first study of natural selection in *clausilia laminata* (Montagu). *Biometrika* 1: 109–124. [29]

Willis, J. H. 1996. Measures of phenotypic selection are biased by partial inbreeding. *Evolution* 50: 1501–1511 [29]

Wright, S. 1935. Evolution in populations in approximate equilibrium. *J. Geneti.* 30: 257–266. [28] [29]