

Multivariate Selection Response and Estimation of Fitness Surfaces

Bruce Walsh. July 2008. NSF Short Course on Statistical Genetics, Birmingham

This lecture serves as a brief introduction to two rather large topics: **multivariate selection response** (response when selection is acting on a vector of traits) and **fitness surface estimation** (what is the nature of selection acting on a vector of traits). These lecture notes are much more detailed than what I will actually cover in my 90 minutes, but are designed to be self-contained for future reference.

Review: Breeding Values and the Breeders' Equation

As way of background, we remind the reader of a couple of key concepts from single-trait quantitative genetics. The first is the idea of a **breeding value** for an individual (this is also called its **additive-genetic value**). An individual has a specific phenotypic value, say 200, but (by itself) this tells us nothing about what their offspring will look like. Given a specific population, each individual also has a breeding value, which tells us what fraction of their phenotype is transmitted to their offspring. In particular, the predicted mean value of offspring is simply the overall population mean plus the average breeding value of its two parents. Thus, breeding value is something that, in theory, can be directly measured. Indeed, this is extensively done in animal breeding.

The additive genetic variance σ_A^2 is the variance in breeding values in a population, and the **heritability** h^2 is defined as the fraction of total (phenotypic) variation that is due to variation in breeding values,

$$h^2 = \frac{\sigma_A^2}{\sigma_z^2}$$

The importance of h^2 for our purposes is the **Breeders' equation**,

$$R = h^2 S \tag{1}$$

which relates the between-generation change in mean (the **response** $R = \mu(t+1) - \mu(t)$) with the within-generation change in the mean from selection (the **selection differential** $S = \mu^*(t) - \mu(t)$, where μ^* is the trait mean among selected individuals). Thus, we can have extensive **selection** (non-zero S) and yet no **evolution** (a zero response) when there is very little variation in breeding values.

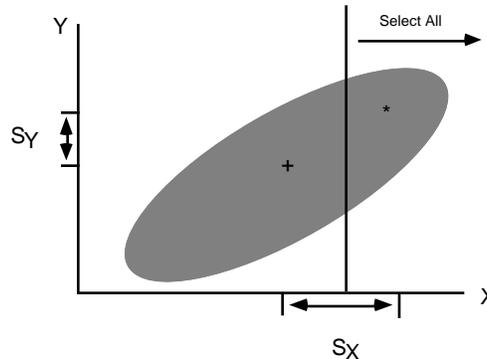
Multitrait Selection: Background

Now suppose we are interested in the joint response of (say) two traits: height and weight. Suppose h^2 for height is 0.8 and 0.4 for weight, while S is 5 for height and 10 for weight. At first glance, one might simply think that the joint response is $0.8 \cdot 5 = 4$ for height and $0.4 \cdot 10 = 4$ for weight. This is only correct if height and weight are *uncorrelated* for both phenotypic and breeding values.

For example, an unselected trait can have a non-zero S (within-generation change in mean) due to selection on a phenotypically-correlated trait. Likewise, if breeding values are correlated, we can see a **correlated selection response** where an unselected trait changes its mean over the next generation (nonzero R) due to selection on another, genetically correlated, trait. Much of this lecture deal with how to decouple these effects. Finally, a nonzero S for a particular trait *does not* always imply a nonzero R , *even* if that trait has a positive heritability – the univariate breeders equation (1) breaks down when selection occurs on multiple traits.

Key Concept: Phenotypic correlations influence within-generations changes, genetic correlations influence between-generation changes.

When selection is applied to character X , character Y will show a correlated selection differential when Y is phenotypically correlated with X . The figure below shows the change in the bivariate mean before (+) and after (*) selection. In this case, there is truncation selection only on trait X , but there is a correlated within-generation change in Y , with the mean of Y in the selected parents differing from the mean before selection.



Note that the correlated selection differential simply measures the *within-generation* change. Whether this translates into a *between-generation* change (i.e., a response to selection) depends on whether there is any *genetic* correlation between the characters. The reason for this can be illustrated by a hypothetical example, with truncation selection only on X . Suppose that $r_A = 0$ and $r_E > 0$ (i.e., no correlation in breeding values, but environmental values within an individual are correlated); therefore r_P and S_Y will be > 0 . The selected individuals, however, will not have higher than average breeding values for Y , in spite of their higher than average phenotypic values. Therefore the correlated response will be zero, regardless of the magnitude of h_Y^2 and S_Y .

Basic Matrix Algebra

The response to selection of several traits is best handled using matrix notation, which we briefly introduce here. A **matrix** is an array of elements, e.g.

$$\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} e & f \\ g & h \end{pmatrix}, \quad \mathbf{C} = \begin{pmatrix} i \\ j \end{pmatrix}$$

When the matrix consists of only a single row or a single column, it is called a **vector**, whereas **A** and **B** are **square matrices**. **C** above is a column vector (consisting of a single column of entries).

Matrix multiplication is defined in such a fashion that the order of multiplication of the matrices is critical. For the above matrices,

$$\mathbf{AB} = \begin{pmatrix} ae + bg & af + bh \\ ce + dg & cf + dh \end{pmatrix}, \quad \mathbf{BA} = \begin{pmatrix} ae + cf & eb + df \\ ga + ch & gd + dh \end{pmatrix}$$

and

$$\mathbf{AC} = \begin{pmatrix} ai + bj \\ ci + dj \end{pmatrix}, \quad \mathbf{BC} = \begin{pmatrix} ei + fj \\ gi + hj \end{pmatrix}$$

while the matrix products **CA** and **CB** are not defined. The **identity matrix**, **I**, which serves a role similar to 1 in scalar multiplication/division, is given by (for the 2×2 cases),

$$\mathbf{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Note that $\mathbf{AI} = \mathbf{IA} = \mathbf{A}$. Finally, the **Inverse** \mathbf{A}^{-1} of a square matrix \mathbf{A} is defined as satisfying

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{AA}^{-1} = \mathbf{I} \quad (2a)$$

For a 2×2 matrix,

$$\mathbf{A}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \quad (2b)$$

The role of the inverse and identity matrix is in solving systems of equations. Suppose we are trying to solve for \mathbf{x} in $\mathbf{AX} = \mathbf{C}$. Premultiplying both sides by the inverse of \mathbf{A} gives

$$\mathbf{A}^{-1}\mathbf{Ax} = \mathbf{A}^{-1}\mathbf{C}$$

noting that $\mathbf{A}^{-1}\mathbf{Ax} = \mathbf{Ix} = \mathbf{x}$ gives $\mathbf{x} = \mathbf{A}^{-1}\mathbf{C}$

The Multivariate Breeders' Equation

We now have all the pieces in place to express the response in matrix form. Suppose there are n traits under selection and we place the n selection differentials into a column vector \mathbf{S} ,

$$\mathbf{S} = \begin{pmatrix} S_1 \\ S_2 \\ \vdots \\ S_n \end{pmatrix}$$

Likewise, define the **phenotypic** and **additive genetic covariance matrices**, \mathbf{P} and \mathbf{G} , respectively, as matrices whose element in the i th row and j column is the covariance (phenotype or additive genetic) between traits i and j . Note that the diagonal elements are the variances. For example, for two characters

$$\mathbf{P} = \begin{pmatrix} \sigma^2(P_1) & \sigma(P_1, P_2) \\ \sigma(P_1, P_2) & \sigma^2(P_2) \end{pmatrix}, \quad \text{and} \quad \mathbf{G} = \begin{pmatrix} \sigma^2(A_1) & \sigma(A_1, A_2) \\ \sigma(A_1, A_2) & \sigma^2(A_2) \end{pmatrix}$$

Let \mathbf{R} denote the column vector of selection responses, so that the i th element in the list is R_i , the change in the mean of character i following one generation of selection. The response to selection becomes

$$\mathbf{R} = \mathbf{GP}^{-1}\mathbf{S} \quad (3)$$

This is the **multivariate breeders' equation**. Recall that the response for a single character under selection is $R = h^2S = \sigma_A^2(\sigma_P^2)^{-1}S$. In the multivariate case, the genetic and phenotypic variance are replaced by variance-covariance matrices and we use matrix inversion and multiplication.

The Directional Selection Gradient

The multivariate breeder's equation can also be written as

$$\mathbf{R} = \mathbf{G}\boldsymbol{\beta} \quad (4a)$$

where

$$\boldsymbol{\beta} = \mathbf{P}^{-1}\mathbf{S} \quad (4b)$$

is called the **directional selection gradient**. The i th element of $\boldsymbol{\beta}$, β_i , measures the amount of direct selection on trait X_i (i.e., the effects of correlated selection differentials are removed). From the rules of matrix multiplication, the response in trait j can be written as

$$R_j = \sigma^2(A_j)\beta_j + \sum_{i \neq j} \sigma(A_j, A_i)\beta_i \quad (5a)$$

where the first term is the change due to direct selection on trait j and the sum is the indirect contribution to the response due to the correlated response of selection on other traits. Likewise note that $\mathbf{P}\boldsymbol{\beta} = \mathbf{S}$, so that the selection differential on trait j can be written as

$$S_j = \sigma^2(P_j) \beta_j + \sum_{i \neq j} \sigma(P_j, P_i) \beta_i \quad (5b)$$

where the first term represents the contribution from direct selection on trait j and the sum term the contribution to the within-generation change due to direct selection on phenotypically correlated traits.

Realized Selection Gradients

The normal use of the multivariate breeder's equation is to predict response $\mathbf{R} = \mathbf{G}\boldsymbol{\beta}$ given the vector of directional selection gradients $\boldsymbol{\beta}$ and the additive-genetic covariance matrix \mathbf{G} . However, we can also use this equation to estimate the long-term selection gradient on a trait given some observed amount of population differentiation, with $\mathbf{R} = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$, where $\boldsymbol{\mu}_i$ is the mean of population i . Pre-multiplying both sides of the breeder's equation by \mathbf{G}^{-1} recovers

$$\boldsymbol{\beta} = \mathbf{G}^{-1}\mathbf{R} \quad (6)$$

Equation 6 requires the strong assumption of constancy of \mathbf{G} over long periods of time.

Example 1. Suppose you are looking at nose length (z_1) and head size (z_2) between a population of mice on the mainland and a divergent population on a nearby island. On the mainland, $\mu_1 = 20$ and $\mu_2 = 30$, while on the island $\mu_1 = 10$ and $\mu_2 = 35$, so that

$$\mathbf{R} = \begin{pmatrix} 20 - 10 \\ 30 - 35 \end{pmatrix} = \begin{pmatrix} 10 \\ -5 \end{pmatrix}$$

Suppose the genetic covariance matrix for these two traits has been stable over time and is

$$\mathbf{G} = \begin{pmatrix} 20 & -10 \\ -10 & 40 \end{pmatrix}$$

The realized selection gradient to obtain this response is estimated as

$$\boldsymbol{\beta} = \begin{pmatrix} 20 & -10 \\ -10 & 40 \end{pmatrix}^{-1} \begin{pmatrix} 10 \\ -5 \end{pmatrix} = \begin{pmatrix} 0.5 \\ 0 \end{pmatrix}$$

Thus, even though character 2 decreased on the island, there was no direct selection on this trait, rather this change is entirely due to a correlated response to direct selection to increase trait 1.

Evolutionary Constraints Imposed by Genetic Correlations

One immediate consequence of the multivariate Breeders' Equation (5a) is that a character under selection need not necessarily change in the direction most favored by natural selection if the correlated response on other characters under selection is stronger. For example, fitness may maximally increase if μ_2 decreases, so that $\beta_2 < 0$. However, if the sum of correlated responses is positive, then μ_2 may increase. Thus, a character may be dragged off its optimal value by correlated responses

on other traits. However, once these responses (driven by strong directional selection) reach their equilibrium, then $\beta_i \simeq 0$, at which point the response from β_2 dominates.

In general, $\mathbf{R} \neq \beta$, the direction of change that results in the largest increase in mean population fitness. Thus, the effect of the additive-genetic covariance matrix \mathbf{G} is to constrain the selection response from its optimal value. The mean vector changes in the direction most favored by selection if and only if

$$\mathbf{G}\beta = \lambda\beta \quad (7)$$

which only occurs when β is an eigenvector of \mathbf{G} (eigenvalues/eigenvectors are reviewed below). Note that even if \mathbf{G} is a diagonal matrix (there is no correlation between the additive genetic values of the characters under selection) Equation 7 is usually not satisfied. In fact, only when we can write $\mathbf{G} = \sigma_A^2 \mathbf{I}$ is Equation 7 satisfied for arbitrary β . Thus, only when both (i) all characters have the same additive genetic variance and (ii) there no additive genetic covariance between characters is the response to selection always in the direction (β) most favored by natural selection. Differences in the amounts of additive genetic variances between characters, and/or non-zero additive-genetic covariances, both impose constraints on character evolution.

Example 2. Consider the following phenotypic and genetic covariance matrices and vector of selection differentials,

$$\mathbf{S} = \begin{pmatrix} 10 \\ -10 \end{pmatrix}, \quad \mathbf{P} = \begin{pmatrix} 20 & -10 \\ -10 & 40 \end{pmatrix}, \quad \mathbf{G} = \begin{pmatrix} 20 & 5 \\ 5 & 10 \end{pmatrix}$$

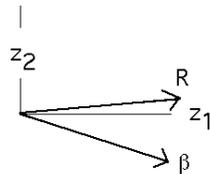
First, assuming no selection on other traits, what is the true nature of selection on these two traits?

$$\beta = \mathbf{P}^{-1}\mathbf{S} = \mathbf{P}^{-1} \begin{pmatrix} 10 \\ -10 \end{pmatrix} = \begin{pmatrix} 0.43 \\ -0.14 \end{pmatrix}$$

Thus, mean population fitness is maximized by increasing trait one and decreasing trait two. What does the actual response look like?

$$\mathbf{R} = \mathbf{G}\beta = \begin{pmatrix} 20 & 5 \\ 5 & 10 \end{pmatrix} \begin{pmatrix} 0.43 \\ -0.14 \end{pmatrix} = \begin{pmatrix} 7.86 \\ 0.71 \end{pmatrix}$$

Thus, the actual response to selection is to increase both traits. The figure below shows the directions of the vectors of optimal response β and the actual response \mathbf{R} .



Is There Genetic Variation in the Multivariate Direction of Selection?

One subtle, but extremely important, feature of multivariate response is that there can be genetic variation (i.e., non-zero heritabilities) in all traits under selection, but little or no genetics variation along *direction* that selection is trying to move the population. To see this point, consider the following example.

Example 3. Suppose the \mathbf{G} matrix is:

$$\mathbf{G} = \begin{pmatrix} 10 & 20 \\ 20 & 40 \end{pmatrix}$$

Suppose that selection is optimized by increasing trait one by two units for every unit trait two is decreased, so that

$$\boldsymbol{\beta} = \begin{pmatrix} 2 \\ -1 \end{pmatrix}$$

The resulting response is

$$\mathbf{R} = \mathbf{G}\boldsymbol{\beta} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Thus, even though there is considerable additive genetic variation in both traits 1 and 2, there is no response. This occurs because \mathbf{G} has a zero eigenvalue, with corresponding eigenvector that exactly corresponds to our $\boldsymbol{\beta}$. Hence, there is NO additive genetic variance along this particular direction, and hence no response. Likewise, if our $\boldsymbol{\beta}$ was only a few degrees away from the eigenvector, the resulting response would be very small.

Example 4. Blows et al. (2004) examined 8 cuticular hydrocarbons (CHCs) in the fly *Drosophila serata*, which are important cues in mate choice. The first two eigenvalues account for 78% of the original additive genetic variation. Blows also estimated the selection gradient on these traits, which play a role in sexual selection. The resulting two leading eigenvectors and $\boldsymbol{\beta}$ were as follows:

$$\mathbf{e}_1 = \begin{pmatrix} 0.232 \\ 0.132 \\ 0.255 \\ 0.536 \\ 0.449 \\ 0.363 \\ 0.430 \\ 0.239 \end{pmatrix}, \quad \mathbf{e}_2 = \begin{pmatrix} 0.319 \\ 0.182 \\ 0.213 \\ -0.436 \\ 0.642 \\ -0.362 \\ -0.014 \\ -0.293 \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} -0.099 \\ -0.055 \\ 0.133 \\ -0.186 \\ -0.133 \\ 0.779 \\ 0.306 \\ -0.465 \end{pmatrix}$$

Let's look at the angle θ between the direction of maximal genetic variation (\mathbf{e}_1) and the optimal direction favored by selection ($\boldsymbol{\beta}$). Recalling that the angle between two vectors is given by (see Equation 11 below),

$$\cos(\theta) = \frac{\mathbf{e}_1^T \boldsymbol{\beta}}{\|\mathbf{e}_1\| \|\boldsymbol{\beta}\|} = \frac{\mathbf{e}_1^T \boldsymbol{\beta}}{\sqrt{\mathbf{e}_1^T \mathbf{e}_1} \sqrt{\boldsymbol{\beta}^T \boldsymbol{\beta}}} = \frac{0.147496}{\sqrt{0.99896 \cdot 0.999502}} = 0.1476$$

Giving $\theta = \cos^{-1}(0.1476) = 81.5$ degrees. Thus, the vector of maximal genetic variation and the vector of optimal response are almost at right angles. Likewise, the angle between \mathbf{e}_2 and $\boldsymbol{\beta}$ is $\theta = 99.65$ degrees. Very little of the standing additive genetic variation is in the direction of the optimal selection response. While all of the CHC traits showed significant variation, and indeed responded to artificial selection, there is very little useable genetic variation in the direction that sexual selection is trying to move the population.

Schluter's Genetic Line of Least Resistance, \mathbf{g}_{max}

The notion of multivariate constraints to response, even when all of the component traits show additive variation, traces back to Dickerson (1955). It is somewhat surprising then that it took over forty years for the first attempts to describe the potential geometry of these constraints. Schluter (1996) defined the **genetic line of least resistance** \mathbf{g}_{max} as the eigenvector corresponding to the lead (largest) eigenvalue of \mathbf{G} (this is just the **first principal component** of \mathbf{G}). Schluter looked at morphological divergence data in a small set of vertebrates (stickleback fish, mice, and three species of birds). Let \mathbf{d} denote the vector of differences between the species means, which we can scale to unit length by

$$\mathbf{d}' = \frac{\mathbf{d}}{\sqrt{\mathbf{d}^T \mathbf{d}}} \quad (8)$$

Using this scaled divergence vector, we can look at the angle θ between \mathbf{d} and \mathbf{g}_{max} , namely the angle between the vector of divergence and the vector (or direction) of maximal additive genetic variation, where

$$\theta = \cos^{-1}(\mathbf{g}_{max}^T \mathbf{d}') \quad (9)$$

Schluter observed the following in his (small) data set

- (i). The smallest values of θ (departures from the direction of maximal genetic variation) occurred between the most recently diverged species.
- (ii). The greater the value of θ (angle between the maximal genetic variation and direction of actual divergence), the smaller the total amount of divergence.
- (iii). The effect \mathbf{g}_{max} on the absolute amount of divergence showed no tendency to weaken with time (out to at least 4 million years).

Thus there is strong empirical evidence that populations tend to evolve along lines of least genetic resistance (i.e., lines of maximal genetic variance). There are two ways to interpret this observation. The first is that such lines constraint selection. The second is that such lines are also the lines upon which maximal genetic drift occurs (the between-mean variance being proportion to the total amount of genetic variation).

McGuigan et al. (2005) offered a very interesting study that offered both some support and some counterexamples to Schluter's general findings. They looked at divergence in two species of Australian rainbow fish (genus *Melanotaenia*) that each have populations differentially adapted to lake vs. stream hydrodynamic environments. Divergence between species, as well as divergence within replicate hydrodynamic populations within each species, followed Schluter's results (small angular departures from \mathbf{d} and \mathbf{g}_{max}). However, hydrodynamic divergence between lake vs. stream populations in each species were along directions that were quite removed from \mathbf{g}_{max} (as well as the other eigenvectors of \mathbf{G} that described most of the genetic variation). Thus, the between- and within-species divergence within the same hydrodynamic environment are consistent with drift, while hydrodynamic divergence had to occur against a gradient of very little genetic variation. Of course, one cannot rule out that the adaptation to these environments resulted in a depletion of genetic variation along these directions.

*Quantifying Multivariate Constraints to Response

Due to time constraints, most of the material in this section will not be covered by my formal lecture. Here, we provide some of the background on the geometry of matrices and then use this to quantify constraints to selection response.

The Geometry of a Matrix

As many of the above examples highlight, we can think about the response to selection in terms of geometry. For example, constraints can be measured by the difference between the optimal vector and the actual response. Hence, we start with a slight digression on the geometry of matrices and vectors.

Comparing Vectors: Lengths and Angles

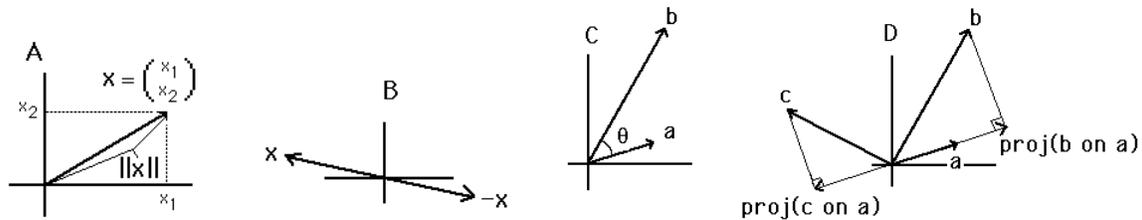


Figure 1. Some basic geometric concepts of vectors. While we use examples from two dimensions, these concepts easily extend to n dimensions. **A:** A vector \mathbf{x} can be thought of as an arrow from the origin to a point in space whose coordinates are given by the elements of \mathbf{x} . **B:**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

As Figure 1 shows, a vector \mathbf{x} can be treated as a geometric object, an arrow leading from the origin to the n dimensional point whose coordinates are given by the elements of \mathbf{x} . By changing coordinate systems, we change the resulting vector, potentially changing both its direction (**rotating** the vector) and length (**scaling** the vector). This geometric interpretation suggests several ways for comparing vectors, such as the *angle* between two vectors (Figure 1C) and the *projection* of one vector onto another (Figure 1D).

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

$$\|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2} = \sqrt{\mathbf{x}^T \mathbf{x}} \quad (10a)$$

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

$$\|\mathbf{x} - \mathbf{y}\|^2 = \sum_{i=1}^n (x_i - y_i)^2 = (\mathbf{x} - \mathbf{y})^T (\mathbf{x} - \mathbf{y}) = (\mathbf{y} - \mathbf{x})^T (\mathbf{y} - \mathbf{x}) \quad (10b)$$

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

$$\cos(\theta) = \frac{\mathbf{x}^T \mathbf{y}}{\|\mathbf{x}\| \|\mathbf{y}\|} = \frac{\mathbf{y}^T \mathbf{x}}{\|\mathbf{x}\| \|\mathbf{y}\|} \quad (11)$$

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

Another way to compare vectors, illustrated in Figure 1D, is to consider the **projection** of one vector on another. For any two n dimensional vectors, the projection of \mathbf{x} on \mathbf{y} generates a vector defined by

$$\text{Proj}(\mathbf{x} \text{ on } \mathbf{y}) = \frac{\mathbf{x}^T \mathbf{y}}{\mathbf{y}^T \mathbf{y}} \mathbf{y} = \frac{\mathbf{x}^T \mathbf{y}}{\|\mathbf{y}\|^2} \mathbf{y} = \left(\cos(\theta) \frac{\|\mathbf{x}\|}{\|\mathbf{y}\|} \right) \mathbf{y} \quad (12a)$$

If $\|\mathbf{y}\| = 1$, then

$$\text{Proj}(\mathbf{x} \text{ on } \mathbf{y}) = (\mathbf{x}^T \mathbf{y}) \mathbf{y} = (\cos(\theta) \|\mathbf{x}\|) \mathbf{y} \quad (12b)$$

Note that since the term involving cosines in Equations 12a/b is a scalar, the vector resulting from the projection of \mathbf{x} on \mathbf{y} is in the same direction as \mathbf{y} , unless $90^\circ < \theta < 270^\circ$ in which case $\cos(\theta) < 0$ and the projection vector is in exactly the opposite direction (the reflection of \mathbf{y} about the origin). The length of the projection vector is

$$\|\text{Proj}(\mathbf{x} \text{ on } \mathbf{y})\| = |\cos(\theta)| \|\mathbf{x}\| \quad (12c)$$

If two vectors lie in exactly the same direction, the projection of one on the other just recovers the original vector (as $\text{Proj}(\mathbf{x} \text{ on } \mathbf{y}) = \mathbf{x}$). Conversely, if two vectors are orthogonal, then the projection of one on the other yields a vector of length zero. An important use of projection vectors is that if $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n$ is any set of mutually orthogonal n dimensional vectors, then any n dimensional vector \mathbf{x} can be represented as the sum of projections of \mathbf{x} onto the members of this set,

$$\mathbf{x} = \sum_{i=1}^n \text{Proj}(\mathbf{x} \text{ on } \mathbf{y}_i) \quad (13)$$

Matrices Describe Vector Transformations

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

Example 5. Suppose

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

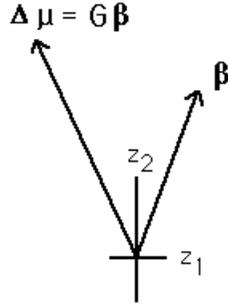
hence

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

The resulting change in character means are different from those most favored by natural selection. Selection favors an increase in z_1 , but when the genetic variance-covariance structure is taken into account, the resulting change in z_1 is negative. Taking the appropriate inner products, we find $\|\boldsymbol{\beta}\| = \sqrt{10}$, $\|\mathbf{R}\| = \sqrt{20}$, and $\boldsymbol{\beta}^T \mathbf{R} = 10$. Applying Equation 11,

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

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



Eigenvalues and Eigenvectors

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

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

$$\mathbf{A}\mathbf{y} = \lambda\mathbf{y} \quad (14)$$

for some scalar value λ . Geometrically, this means that the new vector resulting from transformation of \mathbf{y} by \mathbf{A} points in the same direction (or is exactly reflected about the origin if $\lambda < 0$) as \mathbf{y} . For such vectors, the only action of the matrix transformation is to scale them by some amount λ . Hence, it is natural that the system of such vectors along with their corresponding scalar multipliers completely describes the geometry of the transformation associated with \mathbf{A} . Vectors satisfying Equation 14 are referred to as *eigenvectors* and their associated scaling factors are *eigenvalues*. If \mathbf{y} is an eigenvector, then $a\mathbf{y}$ is also an eigenvector as $\mathbf{A}(a\mathbf{y}) = a(\mathbf{A}\mathbf{y}) = \lambda(a\mathbf{y})$. Note, however, that the associated eigenvalue remains unchanged. Hence, we typically scale eigenvectors to be of unit length to give **unit** or **normalized** eigenvectors. In particular, if \mathbf{u}_i is the eigenvector associated with the i th eigenvalue, then the associated normalized eigenvector is $\mathbf{e}_i = \mathbf{u}_i / \|\mathbf{u}_i\|$.

The eigenvalues of square matrix \mathbf{A} of dimension n are solutions of Equation 14, which is usually expressed as the **characteristic equation** $|\mathbf{A} - \lambda\mathbf{I}| = 0$. This can also be expressed using the **Laplace expansion** as

$$|\mathbf{A} - \lambda\mathbf{I}| = (-\lambda)^n + S_1(-\lambda)^{n-1} + \dots + S_{n-1}(-\lambda)^1 + S_n = 0 \quad (15)$$

where S_i is the sum of all **principal minors** (minors including diagonal elements of the original matrix) of order i (minors are defined in Chapter 8 of Lynch and Walsh). Finding the eigenvalues thus requires solving a polynomial equation of order n . In practice, for $n > 2$ this is usually done numerically, and most statistical and numerical analysis packages offer routines to accomplish this task.

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

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

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

functions of the eigenvalues,

$$\text{tr}(\mathbf{A}) = \sum_{i=1}^n \lambda_i \quad \text{and} \quad |\mathbf{A}| = \prod_{i=1}^n \lambda_i \quad (16)$$

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

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

$$\mathbf{A} = \lambda_1 \mathbf{e}_1 \mathbf{e}_1^T + \lambda_2 \mathbf{e}_2 \mathbf{e}_2^T + \cdots + \lambda_n \mathbf{e}_n \mathbf{e}_n^T \quad (17a)$$

This decomposition is called the **spectral decomposition** of \mathbf{A} . Hence,

$$\begin{aligned} \mathbf{A}\mathbf{x} &= \lambda_1 \mathbf{e}_1 \mathbf{e}_1^T \mathbf{x} + \lambda_2 \mathbf{e}_2 \mathbf{e}_2^T \mathbf{x} + \cdots + \lambda_n \mathbf{e}_n \mathbf{e}_n^T \mathbf{x} \\ &= \lambda_1 \text{Proj}(\mathbf{x} \text{ on } \mathbf{e}_1) + \lambda_2 \text{Proj}(\mathbf{x} \text{ on } \mathbf{e}_2) + \cdots + \lambda_n \text{Proj}(\mathbf{x} \text{ on } \mathbf{e}_n) \end{aligned} \quad (17b)$$

The last equality follows since $\mathbf{e}_i(\mathbf{e}_i^T \mathbf{x}) = (\mathbf{e}_i^T \mathbf{x})\mathbf{e}_i$ as $\mathbf{e}_i^T \mathbf{x}$ is a scalar.

Example 6. Determine the **eigenstructure** (the set of eigenvalues and their associated unit eigenvectors) of the genetic variance-covariance matrix \mathbf{G} given in Example 5. Writing the characteristic equation,

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

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

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

The associated unit eigenvectors are

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

These are orthogonal as $\mathbf{e}_1^T \mathbf{e}_2 = 0$. From Example 5, $\|\boldsymbol{\beta}\| = \sqrt{10}$, while $\mathbf{e}_1^T \boldsymbol{\beta} \simeq 0.727$ and $\mathbf{e}_2^T \boldsymbol{\beta} \simeq 3.079$. Since $\|\mathbf{e}_1\| = \|\mathbf{e}_2\| = 1$,

$$\cos(\theta|\mathbf{e}_1, \boldsymbol{\beta}) \simeq \frac{0.727}{\sqrt{10}} \simeq 0.201 \quad \text{and} \quad \cos(\theta|\mathbf{e}_2, \boldsymbol{\beta}) \simeq \frac{3.079}{\sqrt{10}} \simeq 0.974$$

giving the angle between \mathbf{e}_1 and $\boldsymbol{\beta}$ as $\theta(\mathbf{e}_1, \boldsymbol{\beta}) \simeq 78.4^\circ$, while $\theta(\mathbf{e}_2, \boldsymbol{\beta}) \simeq 13.2^\circ$. The corresponding scaled projections of $\boldsymbol{\beta}$ on these eigenvectors are

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

From Equation 17b, $\boldsymbol{\beta}$ is the sum of these two projections. As Figure 2 shows, the eigenstructure of \mathbf{G} explains the unusual behavior of response seen in Example 5. The eigenvector associated with the **leading eigenvalue** λ_1 accounts for most of the variation inherent in \mathbf{G} , and this eigenvector corresponds to a strong negative correlation between the additive genetic values of z_1 and z_2 . Hence,

even though β points in very much the same direction as e_2 , because $\lambda_1 \gg \lambda_2$ the projection of β on e_1 gives a vector of greater length than the projection on e_2 , and it is this projection vector that results in the decrease in μ_{z_1} .

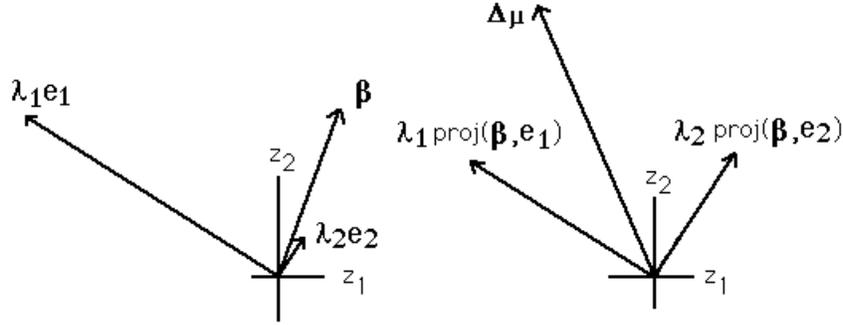


Figure 2. Left: The scaled eigenvectors associated with the variance-covariance matrix \mathbf{G} from Example 5. Note that e_1 and e_2 are orthogonal and hence can be thought of as describing a new coordinate system. **Right:** Since $\lambda_1 \gg \lambda_2$, the leading eigenvector largely dominates the transformation. This is shown by taking the projections of β on each of these eigenvectors. Even though β is nearly orthogonal to e_2 , the projection of β on e_1 yields a vector of greater length than the projection of β on e_2 . From Equation 17b, the response to selection \mathbf{R} is the sum of these two projections.

Blow's Matrix Subspace Projection

Schluter's approach only deals with the leading eigenvector of \mathbf{G} . If the leading eigenvector dominates all of the others (and hence accounts for most of the variance), then this makes sense. (The fraction of variation accounted for along the direction of the k -th eigenvalue is just $\lambda_k / \sum \lambda_i = \lambda_k / \text{trace}(\mathbf{G})$). However, in many cases the first few eigenvalues together may account for most of the variation, so that focusing only on the largest may miss a significant fraction of the variation. Blows et al. (2004) suggested a matrix subspace projection approach to consider more of the eigenvectors of \mathbf{G} .

A common problem is that the \mathbf{G} matrix is **ill-conditioned**, in that $\lambda_{max}/\lambda_{min}$ is large. In such cases (as well as others!) estimation of the \mathbf{G} matrix may result in estimates of eigenvalues that are very close to zero or even negative. Negative estimates arise due to sampling, but values near zero may reflect the true biology in that although n traits may be measured, there is very little variation in certain dimensions. In such cases, one might extract a subset of \mathbf{G} , for example by taking the leading k eigenvectors. This set forms a **subspace** of the full genetic variance described by \mathbf{G} . It is usually the case the \mathbf{G} contains several (perhaps most!) eigenvalues that account for almost no variation (i.e., $\lambda_i / \text{tr}(\mathbf{G}) \simeq 0$). In such cases, most of the genetic variation resides on a lower-dimensional subspace.

We can examine the genetic constraints on this subspace by looking at the **projection** of the full space into this subspace (this is just the matrix extension to the projection of one vector onto another that was discussed earlier). Suppose we have included the first k eigenvectors in our analysis. We can use these to form a **projection matrix** by first defining the matrix \mathbf{A} , where

$$\mathbf{A} = (\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_k) \quad (18)$$

so that the \mathbf{A} matrix consists of the first k eigenvectors of \mathbf{G} . The resulting projection matrix becomes

$$\mathbf{P}_{roj} = \mathbf{A} (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \quad (19a)$$

and in particular, the projection of the optimal vector of selection response β onto this subspace of \mathbf{G} (the subspace that essentially contains all of the useable additive variation) is given by

$$\mathbf{p} = \mathbf{P}_{roj}\beta = \mathbf{A} (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \beta \quad (19b)$$

Example 7. Let's reconsider Blow's CHC data. The first two eigenvalues account for roughly 80% of the total variation in \mathbf{G} , i.e., $(\lambda_1 + \lambda_2) / \sum \lambda_i = 0.78$. The resulting \mathbf{A} matrix becomes

$$\mathbf{A} = (\mathbf{e}_1, \mathbf{e}_2) = \begin{pmatrix} 0.232 & 0.319 \\ 0.132 & 0.182 \\ 0.255 & 0.213 \\ 0.536 & -0.436 \\ 0.449 & 0.642 \\ 0.363 & -0.362 \\ 0.430 & -0.014 \\ 0.239 & -0.293 \end{pmatrix}$$

Applying Equation 19a gives an 8×8 projection matrix (not show here), and Equation 19b gives the projection vector \mathbf{b} of β onto the subspace given by \mathbf{A} as

$$\mathbf{p} = \mathbf{P}_{roj}\beta = \begin{pmatrix} -0.0192 \\ -0.0110 \\ 0.0019 \\ 0.1522 \\ -0.0413 \\ 0.1142 \\ 0.0658 \\ 0.0844 \end{pmatrix}$$

The angle θ between β and the projection of β into the subspace of the genetic variance is given by

$$\theta = \cos^{-1} \left(\frac{\mathbf{p}^T \beta}{\sqrt{\mathbf{p}^T \mathbf{p}} \sqrt{\beta^T \beta}} \right) = \cos^{-1} (0.223) = 77.1 \text{ degrees}$$

Thus the direction of optimal response is 77 degrees away from the total genetic variation (78%) described by this subspace.

Describing Phenotypic Selection: Introductory Remarks

How do particular character values influence the fitness of an individual? Our interest in a particular character may be in predicting how selection changes it over time, which requires knowledge of the genetics of that character. Alternately, we may simply wish to explore the ecological implications of a character by examining how expected fitness changes with character value.

One general way of detecting selection on a character is to compare the (fitness-weighted) phenotypic distribution before and after some episode of selection. One important caveat is that growth or other ontogenetic changes, immigration, and environmental changes can also change the

phenotypic distribution. We must take great care to account for these factors. Typically, selection on a character is measured by considering changes in the mean and variance, rather than changes in the entire distribution. In many cases the entire selection response following a generation of selection can be reasonably predicted from the within-generation change in the mean and variance. As we will see below, there are a number of subtle issues in assigning fitnesses to phenotypes. Again, we remind the reader of the most important caveat, that of dealing with the effects of selection on correlated characters, which will be considered in shortly.

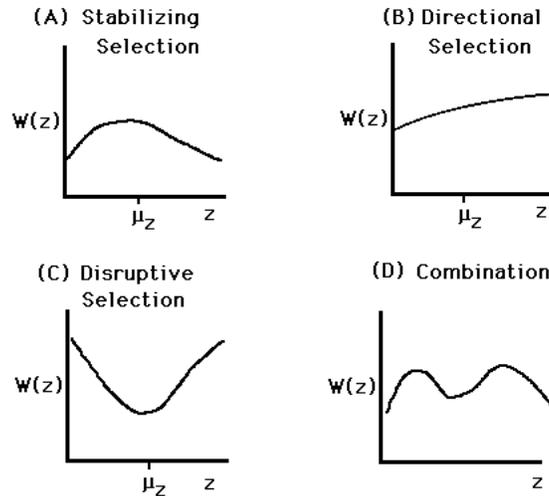


Figure 3. Selection is usually classified into three basic forms depending on the local geometry of $W(z)$: stabilizing (A), directional (B), and disruptive (C). As D illustrates, populations can simultaneously experience multiple forms of selection.

Fitness Surfaces

$W(z)$, the expected fitness of an individual with phenotype z , describes a **fitness surface** (or equivalently a **fitness function**), relating fitness and character value. The **relative fitness surface** $w(z) = W(z)/\bar{W}$ is often more convenient than $W(z)$, and we use these two somewhat interchangeably. The nature of selection on a character in a particular population is determined by the local geometry of the individual fitness surface over that part of the surface spanned by the population (Figure 3). If fitness is increasing (decreasing) over some range of phenotypes, a population having its mean value in this interval experiences **directional selection**. If $W(z)$ contains a local maximum, a population with members within that interval experiences **stabilizing selection**. If the population is distributed around a local minimum, **disruptive selection** occurs. As is illustrated in Figure 4D, when the local geometry of the fitness surface is complicated (e.g., multimodal) the simplicity of description offered by these three types of selection breaks down, as the population can experience all three simultaneously.

Mean population fitness \bar{W} is also a fitness surface, describing the expected fitness of the population as a function of the distribution of phenotypes in that population,

$$\bar{W}(\theta) = \int W(z) p(z, \theta) dz$$

Hence, mean fitness can be thought of as a function of the parameters θ of the phenotypic distribution. For example, if z is normally distributed, mean fitness is a function of the mean μ_z and variance σ_z^2 for that population.

To stress the distinction between the $W(z)$ and \bar{W} fitness surfaces, we call the former the **individual fitness surface**, latter as the **mean fitness surface**. Knowing the individual fitness surface allows us to compute the mean fitness surface for any specified distribution of phenotypes $p(z)$, but the converse is not true. The importance of the mean fitness surface is that it provides one way of describing how the population changes under selection. When the breeders' equation holds, the first two partial derivatives of \bar{W} with respect to μ_z describe the change in mean and variance. Mean fitness surfaces are considerably smoother than the individual fitness surfaces that generate them (Figure 4). The individual fitness surface may have discontinuities and rough spots — regions where very small changes in phenotypic value result in large changes in individual fitness. Mean population fitness averages over $W(z)$, smoothing out these rough spots.

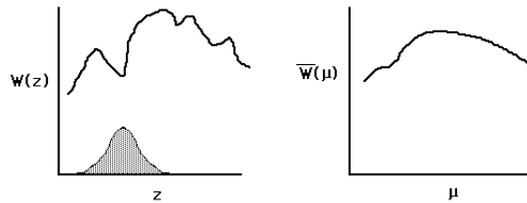


Figure 4. In this example, a small change in z can result in a large change in the individual fitness surface $W(z)$. However, since the mean population fitness $\bar{W}(\mu_z)$ averages individual fitnesses over the phenotypic distribution, shown as the stippled curve, small changes in μ_z result in only small changes in $\bar{W}(\mu_z)$.

Describing Phenotypic Selection: Changes In Phenotypic Moments

Selection for particular phenotypes changes in the phenotypic distribution (although it need not change all moments, for example, the mean may be unchanged). Thus, selection is detected by testing for differences between the distribution of phenotypes before and after some episode of selection. Nonparametric tests such as the Kolmogorov-Smirnov test have the advantage of making no assumptions about the form of the distribution, but suffer from low power. While complete distributions can be compared, the most common procedure for detecting selection is to test for changes in phenotypic moments. Standard statistical tests for differences in means (t -tests) and variances (F -tests) can be used, but these tests rely on normality assumptions that are often violated, and nonparametric tests are often more appropriate. Differences in means can be tested using the Wilcoxon-Mann-Whitney test, while Conover's squared rank test (Conover 1980) can be used to test for changes in variances. Other nonparametric tests for changes in variance exist, but care must be exercised, as some (e.g., the Siegel-Tukey test) are quite sensitive to differences in means. While these issues are important, the main problem in detecting selection on a character is that changes in the moments may be due entirely to selection on phenotypically correlated characters (to be discussed shortly). Keeping this important caveat in mind, we now examine measures of selection for single characters.

The Robertson-Price Identity

If $p(z)$ is the density of phenotype z before selection, then the density after selection is

$$p_s(z) = \frac{W(z)p(z)}{\int W(z)p(z) dz} = \frac{W(z)p(z)}{\bar{W}} = w(z)p(z) \quad (20a)$$

The mean trait value following selection becomes

$$\mu_s = \int z p_s(z) dz = \int z w(z) p(z) dz = E[z w(z)] \quad (20b)$$

Note also that

$$\bar{w} = \int w(z)p(z) dz = \frac{1}{\bar{W}} \int W(z)p(z) dz = \bar{W}/\bar{W} = 1$$

i.e., the mean relative fitness in a population is always equal to one, and that since $\mu = E(z) \cdot E(w) = E(z) \cdot 1$, the directional selection differential may be rewritten as

$$S = \mu_s - \mu = E[zw(z)] - E(z)E(w) = \sigma[z, w(z)] \quad (21)$$

Thus, the *directional selection differential is equivalent to the covariance of phenotype and relative fitness*. This relationship, first noted by Robertson (1966), was greatly elaborated on by Price (1970, 1972). We refer to this very useful result as the **Robertson-Price identity**.

Directional Selection: Differentials (S) and Gradients (β)

We have extensively discussed the standard measure of selection on the mean, the **directional selection differential**, $S = \mu_s - \mu$, the within-generation change in the mean. An alternative measure for the change in the mean follows from the Robertson-Price identity. Define the **directional selection differential** β by

$$\beta = \frac{S}{\sigma_z^2} = \frac{\sigma(z, w)}{\sigma_z^2} \quad (22)$$

Recall that the slope of the regression of y given x is $\sigma(x, y)/\sigma_x^2$. Thus, β is the slope of the least-squares linear regression of relative fitness (w) on phenotypic value (z), $w(z) = 1 + \beta(z - \mu_z) + e$. Substituting $S = \sigma_z^2 \beta$ into $R = h^2 S$ gives

$$R = \sigma_A^2 \beta \quad (23a)$$

Noting that $\sigma_A^2 = h^2 \sigma_z^2$, we can rearrange 23a to give

$$\frac{R}{\sigma_z^2} = h^2 \beta \quad (23b)$$

so that $h^2 \beta$ is the expected response in units of phenotypic variance. The importance of the selection gradient is that under appropriate conditions it relates how a change in the trait mean maps into a change in the mean fitness of a population. In particular, if $w(z)$ denotes the expected fitness of an individual with character value z , then when phenotypes are normally distributed, and fitness is frequency-independent (individual fitnesses are not a function of the mean of the character), Lande (1976) showed that the directional selection gradient satisfies

$$\beta = \frac{\partial \ln \bar{W}}{\partial \mu} \quad (24a)$$

Hence we can express the breeders equation as

$$R = \sigma_A^2 \left(\frac{\partial \ln \bar{W}}{\partial \mu} \right) \quad (24b)$$

Quadratic Selection: Differentials (C) and Gradients (γ)

Analogous measures to S and β can be defined to quantify the change in variance. At first glance this seems best described by $\sigma_{z^*}^2 - \sigma_z^2$, where $\sigma_{z^*}^2$ is the phenotypic variance following selection.

The problem with this measure is that directional selection reduces the variance. Lande and Arnold (1983) showed (see Example 10) that

$$\sigma_{z^*}^2 - \sigma_z^2 = \sigma [w, (z - \mu_z)^2] - S^2 \quad (25a)$$

Hence, directional selection decreases the phenotypic variance by S^2 . With this in mind, Lande and Arnold suggest a corrected measure, the **stabilizing selection differential**

$$C = \sigma_{z^*}^2 - \sigma_z^2 + S^2 \quad (25b)$$

that describes selection acting directly on the variance. As we will see below, the term stabilizing selection differential may be slightly misleading, so following Phillips and Arnold (1989) we refer to C as the **quadratic selection differential**. Correction for the effects of directional selection is important, as claims of stabilizing selection based on a reduction in variance following selection can be due entirely to reduction in variance caused by directional selection. Similarly, disruptive selection can be masked by directional selection (e.g., Example 8).

Example 8. Boag and Grant (1981) observed intense natural selection in *Geospiza fortis* (Darwin's medium ground finch) during a severe drought on Daphne Major Island in the Galápagos. The estimated mean weight and variance of 642 adults before the drought were respectively, 15.79 and 2.37, while the estimated mean and variance of 85 surviving adults after the drought was 16.85 and 2.43. Thus $\hat{S} = 16.85 - 15.79 = 1.06$ (with a standard error of 0.180), implying that the directional selection differential on body size is significantly positive. There appears to be very little selection on the variance when the uncorrected change in variance $\text{Var}(z^*) - \text{Var}(z) = 2.43 - 2.37 = 0.06$ is used. However,

$$\hat{C} = 0.06 + 1.06^2 = 1.14$$

consistent with disruptive selection in addition to directional selection. The standard error for \hat{C} is 0.549. Thus \hat{C} is 2.08 standard errors above zero, suggesting that it is also significant.

Provided that selection does not act on characters phenotypically correlated with the one under study, C provides information on the nature of selection on the variance. Positive C indicates selection to increase the variance (as would occur with disruptive selection), while negative C indicates selection to reduce the variance (as would occur with stabilizing selection). As we discuss shortly, $C < 0$ ($C > 0$) is *consistent* with stabilizing (disruptive) selection, but not *sufficient*. A further complication in interpreting C is that if the phenotypic distribution is skewed, selection on the variance changes the mean. This causes a non-zero value of S that in turn inflates C (Figure 5).

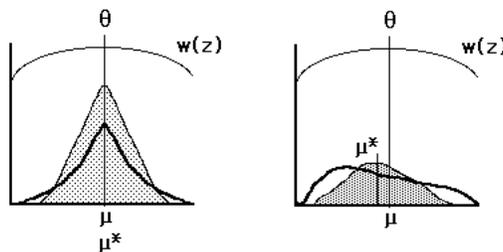


Figure 5. Even when a population is under strict stabilizing selection, the mean can change if the phenotypic distribution is skewed. A standard fitness function for stabilizing selection is $W(z) = 1 - b(\theta - z)^2$. O'Donald (1968) found that, even if the population mean is at the optimum value ($\mu_z = \theta$), S is nonnegative if the skew is nonzero ($\mu_{3,z} \neq 0$) as $S = -(b\mu_{3,z})/(1 - b\sigma_z^2)$. **Left:** If phenotypes are distributed symmetrically about the mean ($\mu_{3,z} = 0$), the distribution after selection (stippled) has the same mean when $\mu_z = \theta$. **Right:** If, however, the distribution is skewed, more of the distribution lies on one side of the mean than the other. Since the distribution is unbalanced, the longer tail experiences more selection than the shorter tail, changing the mean.

Analogous to S equaling the covariance between z and relative fitness, Equation 25 implies C is the covariance between relative fitness and the squared deviation of a character from its mean

$$C = \sigma [w, (z - \mu)^2] \quad (26)$$

The quadratic analogue of β , the **quadratic (stabilizing) selection gradient** γ , was suggested by Lande and Arnold (1983),

$$\gamma = \frac{\sigma [w, (z - \mu)^2]}{\sigma_z^4} = \frac{C}{\sigma_z^4} \quad (27)$$

β and γ Measure the Geometry of the Fitness Surface

A conceptual advantage of β and γ is that they describe the average local geometry of the fitness surface when phenotypes are normally distributed. When z is normal and individual fitness are not frequency-dependent, Equation 24a implies that β can be expressed in terms of the geometry of the *mean* fitness surface,

$$\beta = \frac{\partial \ln \bar{W}}{\partial \mu_z} = \frac{1}{\bar{W}} \frac{\partial \bar{W}}{\partial \mu_z}$$

the slope of the \bar{W} surface with respect to population mean. β can also be expressed as a function of the *individual* fitness surface. Lande and Arnold (1983) showed, provided z is normally distributed, that

$$\beta = \int \frac{\partial w(z)}{\partial z} p(z) dz \quad (28a)$$

implying that β is the average slope of the individual fitness surface, the average being taken over the population being studied. Likewise, if z is normally distributed,

$$\gamma = \int \frac{\partial^2 w(z)}{\partial z^2} p(z) dz \quad (28b)$$

which is the average curvature of the individual fitness surface (Lande and Arnold 1983). Thus, β and γ provide a measure of the geometry of the individual fitness surface averaged over the population being considered.

β and γ Describe the Selection Dynamics

A final advantage of β and γ is that they appear as the only measures of phenotypic selection in equations describing selection response. We have already seen (Equation 23a) that under the constraints of the breeders' equation, $\Delta\mu = \sigma_A^2\beta$, which is independent of any other measure of the phenotype (note that σ_z^2 is missing). Similarly, for predicting changes in variance under the infinitesimal model, the expected change in variance from a single generation of selection is

$$\Delta\sigma_A^2 = \frac{\sigma_A^4}{2} (\gamma - \beta^2) \quad (29)$$

While the distinction between differentials and gradients seems almost trivial in the univariate case (only a scale difference), *the multivariate versions are considerably different*. As we will see in later lectures, gradients have the extremely important feature of removing the effects of phenotypic correlations.

Describing Phenotypic Selection: Individual Fitness Surfaces

We can decompose the fitness W of an individual with character value z into the sum of its expected fitness $W(z)$ plus a residual deviation e ,

$$W = W(z) + e$$

The residual variance for a given z , $\sigma_e^2(z)$, measures the variance in fitness among individuals with phenotypic value z . Estimation of the individual fitness surface is thus a generalized regression problem, the goal being to choose a candidate function for $W(z)$ that minimizes the average residual variance $E_z[\sigma_e^2(z)]$. Since the total variance in fitness σ_W^2 equals the sum of the within-group (phenotype) and between-group variance in fitness,

$$\frac{\sigma_W^2 - E_z[\sigma_e^2(z)]}{\sigma_W^2}$$

is the fraction of individual fitness variation accounted for by a particular estimate of $W(z)$, and this provides a measure for comparing different estimates. In the limiting case where fitness is independent of z (and any characters phenotypically correlated with z), $W(z) = \bar{W}$, so that the between-phenotypic variance is zero while $\sigma_e^2(z) = \sigma_W^2$.

There are at least two sources of error contributing to e . First, there can be errors in measuring the actual fitness of an individual (these are almost always ignored). Second, the *actual* fitness of a particular individual can deviate considerably from the *expected value* for its phenotype due to chance effects and selection on other characters besides those being considered. Generally, these residual deviations are heteroscedastic. To see how this arises naturally, suppose fitness is measured by survival to a particular age. While $W(z) = p_z$ is the probability of survival for an individual with character value z , the fitness for a particular individual is either 0 (does not survive) or 1 (survives). Thus the residual has only two possible values, $e = 1 - p_z$ with probability p_z and $e = -p_z$ with probability $1 - p_z$, giving $\sigma_e^2(z) = p_z(1 - p_z)$. Unless p_z is constant over z , the residuals are heteroscedastic. Note in this case that even after removing the effects attributable to differences in phenotypes, there still is substantial variance in individual fitness.

Inferences about the individual fitness surface are limited by the range of phenotypes in the population. Unless this range is very large, only a small region of the fitness surface can be estimated with any precision. Estimates of the fitness surface at the tails of the current phenotypic distribution are extremely imprecise, yet potentially very informative, suggesting what selection pressures populations at the margin of the observed range of phenotypes may be under. A further complication is that the fitness surface changes as the environment changes so that year to year estimates can differ and cannot be lumped together to increase sample size.

Linear and Quadratic Approximations of $W(z)$

The individual fitness surface $W(z)$ can be very complex and a wide variety of functions may be chosen to approximate it. The simplest and most straightforward approach is to use a low-order polynomial (typically linear or quadratic).

Consider first the simple linear regression of *relative* fitness w as a function of phenotypic value z . Since the directional selection gradient $\beta = S/\sigma_z^2 = \sigma(w, z)/\sigma_z^2$, it follows from regression theory that β is the slope of the least-squares linear regression of relative fitness on z ,

$$w = a + \beta z + e \tag{30a}$$

Hence the best linear predictor of relative fitness is $w(z) = a + \beta z$. Since the regression passes through the expected values of w and z , this can be written as

$$w = 1 + \beta(z - \mu_z) + e \quad (30b)$$

giving $w(z) = 1 + \beta(z - \mu_z)$. Assuming the fitness function is well described by a linear regression, β is the expected change in relative fitness given a unit change in z . From standard regression theory, the fraction of variance in individual fitness accounted for by this regression is

$$r_{z,w}^2 = \frac{\text{Cov}^2(z, w)}{\text{Var}(z) \cdot \text{Var}(w)} = \hat{\beta}^2 \frac{\text{Var}(z)}{\hat{I}} \quad (31)$$

If the fitness surface shows curvature, as might be expected if there is stabilizing selection and/or disruptive selection, a **quadratic regression** is more appropriate,

$$w = a + b_1 z + b_2 z^2 + e \quad (32a)$$

We can also express this as

$$w = \alpha + b_1 z + b_2 (z - \mu)^2 + e \quad (32b)$$

The regression coefficients are the same in 32a and 32b, while the intercepts differ.

The regression coefficients b_1 and b_2 nicely summarize the local geometry of the fitness surface. Evaluating the derivative of Equation 32 at $z = \mu_z$ gives

$$\left. \frac{\partial w(z)}{\partial z} \right|_{z=\mu_z} = b_1 \quad \text{and} \quad \left. \frac{\partial^2 w(z)}{\partial z^2} \right|_{z=\mu_z} = 2b_2 \quad (33)$$

Hence b_1 is the slope and $2b_2$ the second derivative (curvature) of the best quadratic fitness surface around the population mean. $b_2 > 0$ indicates that the best-fitting quadratic of the individual fitness surface has an upward (positive) curvature (**concave selection**), while $b_2 < 0$ implies the curvature is downward (negative, for **convex selection**). Lande and Arnold (1983) suggest that $b_2 > 0$ indicates disruptive selection, while $b_2 < 0$ indicates stabilizing selection. Their reasoning follows from elementary geometry in that a *necessary* condition for a local minimum is that a function curves upward in some interval, while a necessary condition for a local maximum is that the function curves downward. Mitchell-Olds and Shaw (1987) and Schluter (1988) argue that this condition is not *sufficient*. Stabilizing selection is generally defined as the presence of a local maximum in $w(z)$ and disruptive selection by the presence of a local minimum, while b_2 indicates curvature, rather than the presence of local extrema. Hence, we use concave and convex selection to describe the sign of curvature (i.e., the sign of the second derivative of the individual fitness surface).

We solve for the regression coefficients b_1 and b_2 by transforming Equation 32 into a standard multiple regression problem by setting $x_1 = z$ and $x_2 = (z - \mu_z)^2$. To proceed, we need expressions for $\sigma(x_1, x_2)$, $\sigma(x_1, w)$, and $\sigma(x_2, w)$. From Lynch and Walsh's Equation A1.14, $\sigma(x_1, x_2) = \sigma(z, (z - \mu_z)^2) = \mu_{3,z}$, the skew of the phenotypic distribution before selection. Likewise, from Equations 21 and 25a, $\sigma(x_1, w) = \sigma(z, w) = S$ and $\sigma(x_2, w) = \sigma((z - \mu_z)^2, w) = C$. Substituting these into standard expressions for a bivariate regression (Lynch and Walsh Example 8.3), and noting that $\sigma^2((z - \mu_z)^2) = \mu_{4,z} - \sigma_z^4$, gives

$$b_1 = \frac{\sigma^2(x_2) \cdot \sigma(x_1, w) - \sigma(x_1, x_2) \cdot \sigma(x_2, w)}{\sigma^2(x_1) \cdot \sigma^2(x_2) - \sigma^2(x_1, x_2)} = \frac{(\mu_{4,z} - \sigma_z^4) \cdot S - \mu_{3,z} \cdot C}{\sigma_z^2 \cdot (\mu_{4,z} - \sigma_z^4) - \mu_{3,z}^2} \quad (34a)$$

$$b_2 = \frac{\sigma^2(x_1) \cdot \sigma(x_2, w) - \sigma(x_1, x_2) \cdot \sigma(x_2, w)}{\sigma^2(x_1) \cdot \sigma^2(x_2) - \sigma^2(x_1, x_2)} = \frac{\sigma_z^2 \cdot C - \mu_{3,z} \cdot S}{\sigma_z^2 \cdot (\mu_{4,z} - \sigma_z^4) - \mu_{3,z}^2} \quad (34b)$$

The estimators of b_1 and b_2 are obtained by replacing $\mu_{k,z}$ with their sample estimates and using \hat{C} and \hat{S} .

Provided z is normally distributed before selection, $\mu_{3,z} = 0$ and $\mu_{4,z} - \sigma_z^4 = 2\sigma_z^4$. In this case, Equations 22 and 27 imply, respectively, that $b_1 = \beta$ and $b_2 = \gamma/2$, giving the univariate version of the **Lande-Arnold regression**,

$$w = \alpha + \beta z + \frac{\gamma}{2} z^2 + e \quad (35)$$

developed by Lande and Arnold (1983), motivated by Pearson (1903). The Lande-Arnold regression thus provides a connection between selection differentials (directional and stabilizing) and quadratic approximations of the individual fitness surface.

An important point from Equation 34a is that if skew is present ($\mu_{3,z} \neq 0$), $b_1 \neq \beta$ and the slope term in the linear regression (the best *linear* fit) of $w(z)$ differs from the slope term in the quadratic regression (the best *quadratic* fit) of $w(z)$. This arises because the presence of skew generates a covariance between z and $(z - \mu_z)^2$. The biological significance of this can be seen by reconsidering Figure 5, wherein the presence of skew in the phenotypic distribution results in a change in the mean of a population under strict stabilizing selection (as measured by the population mean being at the optimum of the individual fitness surface). Skew generates a correlation between z and $(z - \mu_z)^2$ so that selection acting only $(z - \mu_z)^2$ generates a correlated change in z . From the Robertson-Price identity (Equation 21), the within-generation change in mean equals the covariance between phenotypic value and relative fitness. Since covariances measure the amount of *linear* association between variables, in describing the change in mean, the correct measure is the slope of the best *linear* fit of the individual fitness surface (Equation 30b). If skew is present, b_1 from the quadratic regression (Equation 32) is a biased estimator of β .

Complications from Unmeasured Correlated Variables

A major complication with the estimation of both linear and quadratic gradients are correlated traits (or other factors) not included in the analysis. For example, selection might be entirely based on height, while we measure body weight. The two traits are correlated, and hence we would observe a relationship between relative fitness and body weight. Shortly, we show that by including all traits under selection, unbiased estimates of β and γ are obtained. For our hypothetical example, using this approach we would obtain a non-zero estimate of the gradient associated with height, but not with weight.

A more interesting complication is when the environment influences both fitness and the trait of interest. For example, suppose that plants growing in rich soil leave more seed and also are able to generate more plant secondary compounds. If we were simply focused on the secondary compounds, we might find an association between them and fitness, entirely due to the shared environmental effect. As with phenotypically-correlated traits, if we knew these environmental factors, we could include them in the model, which would result in unbiased estimates of gradients. In reality, we are unlikely to know critical environmental factors.

Example 9. Kruuk et al. (2002) examined antler size in red deer on the Inner Hebrides Isle of Rum in Scotland. They found that males with larger antlers enjoy increased lifetime breeding success (antlers being involved in male-male competition), resulting in a $\beta = 0.44 \pm 0.18$. Further, antler size is also heritable, $h^2 = 0.22 \pm 0.12$, so that one would expect a significant response to natural selection, given both heritability in the trait and strong selective pressure for change. However, despite selection and heritability, no significant changes were observed response over 30 years of study. The authors suggest that antler size and male fighting ability is heavily dependent upon an individual's nutritional state, and hence selection is on this environmental component, rather than any genetic component. As a result, there is no response to selection.

Strenght of Selection in Natural Populations

Just how strong is selection in natural populations? Attempts at measuring selection on quantitative traits in nature trace back to Bumpus (1899) and Weldon (1901). Endler (1986) was the first to attempt to summarize the average strenght of selection, while more recently Kingsolver (Kingsolver et al. 2001, Hoekstra et al. 2001) summarized over 2,500 estimates of β and γ from natural populations (Figure 6).

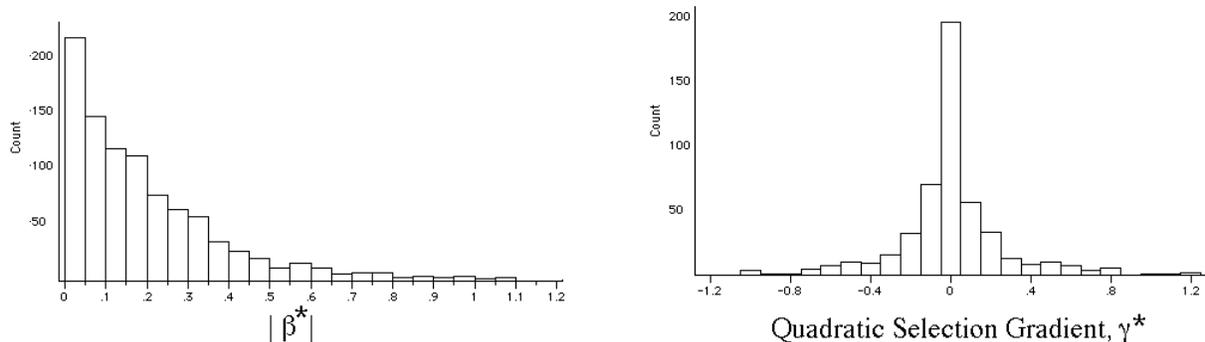


Figure 6. Summary of estimates of directional and quadratic selection in the wild. All estimates are scaled in terms of phenotypic standard deviations ($\beta^* = \beta/\sigma_z$, $\gamma^* = \gamma/\sigma_z$). Data from the summary by Kingsolver et al. 2001. **Left:** Plot of roughly 1,000 estimated 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.

A couple of surprising results emerge from this meta-analysis by Kingsolver. First, the distribution of the absolute value of (scaled) directional selection gradients essentially follows an exponential distribution with a medium (50% value) of 0.16. Note that a β^* of 0.16 implies at a one standard deviation change in the trait changes relative fitness by 16%. Thus, 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). 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^* value are simply a consequence of sampling error due to small sample size. Second, the wide-spread belief is that stabilizing selection is far more common than disruptive selection. However, Kingsolver observed an essentially symmetric distribution of γ^* values, mean positively- and negatively-curved fitness surfaces were equally common. Further, the average strenght of quadratic selection was weak, with $|\gamma^*|$. However, Blows and Brooks (2003) point out that the univariate estimators of γ likely significantly *underestimate* the strenght of quadratic selection when multiple traits are considered. We discuss this point shortly.

As pointed out by Connner (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 23b, the single-generation change in phenotypic standard deviations is $h^2 \cdot 0.16$. Hence, with a typically heritability of 0.4, only 16 generations of selection are required to shift the population mean by one standard deviation.

Selection On Multivariate Phenotypes: Differentials And Gradients

Above we described a variety of measures of univariate selection, with an emphasis on approximating the individual fitness function. In extending these methods to multiple characters, our main

concern is the effect of phenotypic correlations (we have previously examined the consequences of genetic correlations on the selection response). The phenotype of an individual is now a vector $\mathbf{z} = (z_1, z_2, \dots, z_n)^T$ 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).

Changes in the Mean Vector: The Directional Selection Differential, \mathbf{S}

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

$$\mathbf{S} = \boldsymbol{\mu}^* - \boldsymbol{\mu} \quad (36)$$

whose i th element is S_i , the differential for character z_i . As with the univariate case, the Robertson-Price identity (Equation 21) holds, so that the elements of \mathbf{S} represent the covariance between character value and relative fitness, $S_i = \sigma(z_i, w)$.

As is illustrated in Figure 7, \mathbf{S} confounds the direct effects of selection on a character 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 7 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 in its phenotypic distribution due to selection on a phenotypically correlated character (indirect 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 \mathbf{z} .

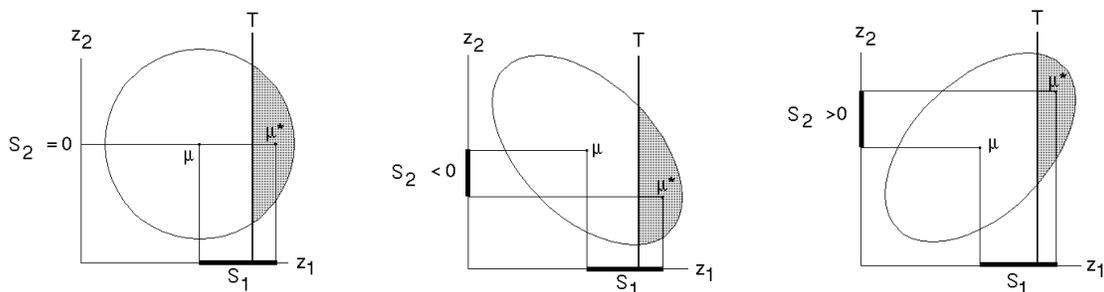


Figure 7. Selection on a character can result in a within-generation change in the mean of other phenotypically correlated characters not themselves under direct selection. Assume that character 1 is under simple truncation selection so that only individuals with $z_1 > T$ reproduce. **Left:** When z_1 and z_2 are uncorrelated, $S_2 = 0$. **Center:** When z_1 and z_2 are negatively correlated, $S_2 < 0$. **Right:** When z_1 and z_2 are positively correlated, $S_2 > 0$.

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

As we discuss shortly, the directional selection gradient $\boldsymbol{\beta}$ removes the effects of phenotypic correlations because it is a vector of partial regression coefficients. From multiple regression theory, 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

$$\mathbf{P}^{-1}\boldsymbol{\sigma}(\mathbf{z}, w) = \mathbf{P}^{-1}\mathbf{S} = \boldsymbol{\beta} \quad (37)$$

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

$$w(\mathbf{z}) = a + \sum_{j=1}^n \beta_j z_j = a + \boldsymbol{\beta}^T \mathbf{z} \quad (38a)$$

Our main interest in this equation is the vector $\boldsymbol{\beta}$ of partial regression coefficients – the slopes for the individual z_j . There are several equivalent ways writing this expression and these various forms interchangeably appear in the literature. First, noting that the regression passes through the mean of w and the mean of \mathbf{z} , namely $(1, \boldsymbol{\mu})$,

$$w(\mathbf{z}) = 1 + \sum_{j=1}^n \beta_j (z_j - \mu_j) = 1 + \boldsymbol{\beta}^T (\mathbf{z} - \boldsymbol{\mu}) \quad (38b)$$

We can also translate the \mathbf{z} values before the analysis to set $\boldsymbol{\mu} = 0$, in which case

$$w(\mathbf{z}) = 1 + \sum_{j=1}^n \beta_j z_j = 1 + \boldsymbol{\beta}^T \mathbf{z} \quad (38c)$$

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 . Provided we can exclude the possibility of unmeasured characters influencing fitness that are 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. Unmeasured traits under selection that are phenotypically correlated with those we measure and/or unmeasured environmental factors that influence both fitness and the values of our measured traits result in $\boldsymbol{\beta}$ being biased measure of the amount of directional selection acting on each measured character.

Since $\mathbf{S} = \mathbf{P}\boldsymbol{\beta}$, we have

$$S_i = \sum_{j=1}^n \beta_j P_{ij} = \beta_i P_{ii} + \sum_{j \neq i}^n \beta_j P_{ij} \quad (39)$$

illustrating that the directional selection differential confounds direct selection on that character with indirect contributions due to selection on phenotypically correlated characters. Equation 39 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. However, recall that if i and j are *genetically* correlated (non-zero additive genetic covariance), then selection on one trait results in a correlated response in the other, *even if there is no phenotypic within-generation change in the mean*.

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 and mean population fitness surfaces. Recall from vector calculus that the gradient vector $\nabla_{\mathbf{x}} f(\mathbf{x})$ is defined at

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

Further recall that the gradient vector of a function points to the direction of change in the variables that will give the greatest (local) increase in the function. Lande (1976, 1979) showed that

$$\beta = \nabla_{\boldsymbol{\mu}}[\ln \bar{W}(\boldsymbol{\mu})] = \bar{W}^{-1} \cdot \nabla_{\boldsymbol{\mu}}[\bar{W}(\boldsymbol{\mu})] \quad (40)$$

which holds provided fitnesses are frequency-independent (Lande 1976, 1979). In this case β is the gradient of **mean population** fitness with respect to the mean vector $\boldsymbol{\mu}$. Since β gives the direction of steepest increase in the mean population fitness surface, mean population fitness increases most rapidly when $\mathbf{R} = \beta$, i.e., when the between-generation change in means is in the same direction as the selection gradient.

Finally, while our focus has been on the role β plays in measuring phenotypic selection, it also plays an important role in the response to selection. If we can assume the breeders' equation holds, β is the only measure of phenotypic selection required to predict the response in mean as $\mathbf{R} = \mathbf{G}\beta$. Cheverud (1984) makes the important point that although it is often assumed a set of phenotypically correlated traits responds to selection in a coordinated fashion, this is not necessarily the case. Since β 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.

Changes in the Covariance Matrix: The Quadratic Selection Differential \mathbf{C}

Motivated by the univariate case wherein $C = \sigma[w, (z - \mu)(z - \mu)]$, 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 $(z_i - \mu_{z_i})(z_j - \mu_{z_j})$ and relative fitness w , viz.,

$$C_{ij} = \sigma[w, (z_i - \mu_{z_i})(z_j - \mu_{z_j})] \quad (41a)$$

As is derived below (Example 10), Lande and Arnold (1983) showed that

$$\mathbf{C} = \sigma[w, (\mathbf{z} - \boldsymbol{\mu})(\mathbf{z} - \boldsymbol{\mu})^T] = \mathbf{P}^* - \mathbf{P} + \mathbf{S}\mathbf{S}^T \quad (41b)$$

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 41b gives

$$P_{ij}^* - P_{ij} = -S_i S_j \quad (42)$$

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

Example 10. We wish to show $\mathbf{P}^* - \mathbf{P} = \sigma[w, (\mathbf{z} - \boldsymbol{\mu})(\mathbf{z} - \boldsymbol{\mu})^T] - \mathbf{S}\mathbf{S}^T$, which implies Equation 41b. From the definition of the variance-covariance matrix,

$$\begin{aligned} \mathbf{P} &= E[(\mathbf{z} - \boldsymbol{\mu})(\mathbf{z} - \boldsymbol{\mu})^T] = \int (\mathbf{z} - \boldsymbol{\mu})(\mathbf{z} - \boldsymbol{\mu})^T p(\mathbf{z}) d\mathbf{z} \\ \mathbf{P}^* &= E[(\mathbf{z}^* - \boldsymbol{\mu}^*)(\mathbf{z}^* - \boldsymbol{\mu}^*)^T] = \int (\mathbf{z} - \boldsymbol{\mu}^*)(\mathbf{z} - \boldsymbol{\mu}^*)^T p^*(\mathbf{z}) d\mathbf{z} \end{aligned}$$

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

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

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

Combining these results,

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

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

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

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

The Quadratic Selection Gradient, $\boldsymbol{\gamma}$

Like \mathbf{S} , \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,

$$w(\mathbf{z}) = a + \sum_{j=1}^n b_j z_j + \frac{1}{2} \sum_{j=1}^n \sum_{k=1}^n \gamma_{jk} (z_j - \mu_j)(z_k - \mu_k) \quad (43a)$$

$$= a + \mathbf{b}^T \mathbf{z} + \frac{1}{2} (\mathbf{z} - \boldsymbol{\mu})^T \boldsymbol{\gamma} (\mathbf{z} - \boldsymbol{\mu}) \quad (43b)$$

Where the ij -th element of the matrix $\boldsymbol{\gamma}$ is γ_{ij} . Again, we can expression this in a simpler form by translating \mathbf{z} so all traits have mean zero, in which case

$$w(\mathbf{z}) = 1 + \mathbf{b}^T \mathbf{z} + \frac{1}{2} \mathbf{z}^T \boldsymbol{\gamma} \mathbf{z} \quad (43c)$$

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

$$\boldsymbol{\gamma} = \mathbf{P}^{-1} \boldsymbol{\sigma} [w, (\mathbf{z} - \boldsymbol{\mu}) (\mathbf{z} - \boldsymbol{\mu})^T] \mathbf{P}^{-1} = \mathbf{P}^{-1} \mathbf{C} \mathbf{P}^{-1} \quad (44)$$

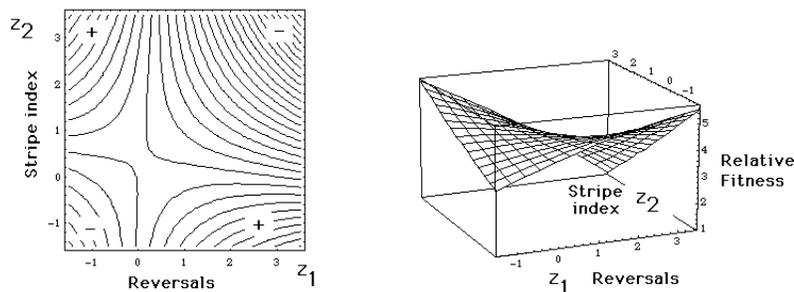
This is the **quadratic selection gradient** and (like $\boldsymbol{\beta}$) removes the effects of phenotypic correlations, providing a more accurate picture of how selection is operating on the multivariate phenotype.

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 34 shows (for the univariate case) that if the phenotypic distribution is skewed, b is a function of both S and C , while $\boldsymbol{\beta}$ 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,

$$w(\mathbf{z}) = 1 + \boldsymbol{\beta}^T \mathbf{z} + \frac{1}{2} \mathbf{z}^T \boldsymbol{\gamma} \mathbf{z} \quad (45)$$

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 pairwise combinations of characters constant, relative fitness is expected to increase 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 γ 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 previously, 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 **correlational selection** — $\gamma_{jk} > 0$ suggests selection for a positive correlation between characters j and k , while $\gamma_{jk} < 0$ suggests selection for a negative correlation.

Example 11. 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 $\gamma_{ij} = -0.268 \pm 0.097$. As is shown below, 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.



The fitness surface. 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 a + and valleys by a - 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 convex selection along one canonical axis (given by the index $0.77 \cdot z_1 - 0.64 \cdot z_2$) and concave selection along the other ($0.64 \cdot z_1 + 0.77 \cdot z_2$).

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. 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 $\boldsymbol{\gamma}$ by \mathbf{P} gives $\mathbf{C} = \mathbf{P} \boldsymbol{\gamma} \mathbf{P}$, or

$$C_{ij} = \sum_{k=1}^n \sum_{\ell=1}^n \gamma_{k\ell} P_{ik} P_{\ell j} \quad (46)$$

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, $\boldsymbol{\gamma}$ provides a measure of the average curvature of the **individual** fitness surface, as

$$\boldsymbol{\gamma} = \int \mathbf{H}_{\mathbf{z}}[W(\mathbf{z})] \phi(\mathbf{z}) d\mathbf{z} \quad (47a)$$

where $\mathbf{H}_{\mathbf{z}}[f]$ denotes the Hessian matrix with respect to \mathbf{z} and is a multivariate measure of the quadratic curvature of a function (the ij -th element of $\mathbf{H}_{\mathbf{z}}[f]$ is $\partial^2 f / \partial z_i \partial z_j$). This result, due to Lande and Arnold (1983), can be obtained by an integration by parts. When fitnesses are frequency-independent (again provided $\mathbf{z} \sim \text{MVN}$), $\boldsymbol{\gamma}$ provides a description of the curvature of the **log mean population** fitness surface, with

$$\mathbf{H}_{\boldsymbol{\mu}}[\ln \bar{W}(\boldsymbol{\mu})] = \boldsymbol{\gamma} - \boldsymbol{\beta} \boldsymbol{\beta}^T \quad (47b)$$

In particular,

$$\frac{\partial \ln \bar{W}(\boldsymbol{\mu})}{\partial \mu_i \partial \mu_j} = \gamma_{ij} - \beta_i \beta_j \quad (47c)$$

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 phenotypic selection on the additive-genetic covariance matrix. The difference between the additive genetic covariance \mathbf{G}^* after selection (but before reproduction) and the covariance matrix \mathbf{G} before selection is

$$\mathbf{G}^* - \mathbf{G} = \mathbf{G} \mathbf{P}^{-1} (\mathbf{P}^* - \mathbf{P}) \mathbf{P}^{-1} \mathbf{G} \quad (48a)$$

We can express this in terms of gradients as follows:

$$\begin{aligned} \mathbf{G}^* - \mathbf{G} &= \mathbf{G} \mathbf{P}^{-1} (\mathbf{P}^* - \mathbf{P}) \mathbf{P}^{-1} \mathbf{G} \\ &= \mathbf{G} \mathbf{P}^{-1} (\mathbf{C} - \mathbf{S} \mathbf{S}^T) \mathbf{P}^{-1} \mathbf{G} \\ &= \mathbf{G} (\mathbf{P}^{-1} \mathbf{C} \mathbf{P}^{-1} - (\mathbf{P}^{-1} \mathbf{S}) (\mathbf{P}^{-1} \mathbf{S})^T) \mathbf{G} \\ &= \mathbf{G} (\boldsymbol{\gamma} - \boldsymbol{\beta} \boldsymbol{\beta}^T) \mathbf{G} \end{aligned} \quad (48b)$$

Hence, the within-generation change in \mathbf{G} has a component from directional selection ($\boldsymbol{\beta}$) and a second due from quadratic selection ($\boldsymbol{\gamma}$),

$$\begin{aligned} \mathbf{G}^* - \mathbf{G} &= -\mathbf{G} \boldsymbol{\beta} \boldsymbol{\beta}^T \mathbf{G} + \mathbf{G} \boldsymbol{\gamma} \mathbf{G} \\ &= -\mathbf{R} \mathbf{R}^T + \mathbf{G} \boldsymbol{\gamma} \mathbf{G} \end{aligned} \quad (48c)$$

In terms of the change in covariance for two particular characters, this can be factored as

$$\begin{aligned} G_{ij}^* - G_{ij} &= - \left(\sum_{k=1}^n \beta_k G_{ik} \right) \left(\sum_{k=1}^n \beta_k G_{jk} \right) + \sum_{k=1}^n \sum_{\ell=1}^n \gamma_{k\ell} G_{ik} G_{\ell j} \\ &= -R_i \cdot R_j + \sum_{k=1}^n \sum_{\ell=1}^n \gamma_{k\ell} G_{ik} G_{\ell j} \end{aligned} \quad (48d)$$

Thus the within-generation change in the additive genetic variance of character i is given by

$$G_{ii}^* - G_{ii} = -(R_i)^2 + \sum_{k=1}^n \sum_{\ell=1}^n \gamma_{k\ell} G_{ik} G_{i\ell} \quad (48e)$$

Summary

Table 1 (next page) summarizes the main features of differentials and gradients.

Multidimensional Quadratic Fitness Regressions

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 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 problems. Unless we test for, and confirm, multivariate normality, β and γ must be estimated from separate regressions — β from the best linear regression, γ from the best quadratic regression. In either case, there are a large number of parameters to estimate — γ has $n(n+1)/2$ terms and β has n terms, for a total $n(n+3)/2$. With 5, 10, and 25 characters, this corresponds to 20, 65 and 350 parameters. The number of observations should greatly exceed $n(n+3)/2$ in order estimate these parameters with any precision.

A second problem is **multicollinearity** — if many of the characters being measured are highly correlated with each other, 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} , which in turn gives a very large sampling variance for the estimate of β and γ . One possibility is to use principal components 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 the perform fitness regressions using the first few PCs as the characters (Lande and Arnold 1983). This approach also reduces the problem of the number of parameters to estimate, but some have expressed concern that it risks the real possibility of removing the most important characters and PCs are often difficult to interpret biologically. While the first PC for morphological characters generally corresponds to a general measure of size, the others are often much more problematic. Finally, using PCs can spread the effects of selection on one character over several PCs, further complicating interpretation.

An alternative opinion is that most traits we measure are themselves artificial, and we are naive to assume that they neatly correspond to the actual characters under selection. In such cases, a lower dimensional subspace of \mathbf{P} may contain most of the variation. As was the case for \mathbf{G} , by considering such a subspace we are likely to gain a much better (and less biased) perspective on the nature of selection. We return to this point shortly.

Table 1. Analogous features of directional and quadratic differentials and gradients.

Changes in Means (Directional Selection)	Changes in Covariances (Quadratic Selection)
Differentials measure the covariance between relative fitness and phenotype	
$S_i = \sigma [w, z_i]$	$C_{ij} = \sigma [w, (z_i - \mu_i)(z_j - \mu_j)]$
Differentials confound direct and indirect selection	
$\mathbf{S} = \boldsymbol{\mu}^* - \boldsymbol{\mu} = \mathbf{P}\boldsymbol{\beta}$	$\mathbf{C} = \mathbf{P}^* - \mathbf{P} + \mathbf{S}\mathbf{S}^T = \mathbf{P}\boldsymbol{\gamma}\mathbf{P}$
Gradients measure the amount of direct selection	
$\boldsymbol{\beta} = \mathbf{P}^{-1}\mathbf{S}$	$\boldsymbol{\gamma} = \mathbf{P}^{-1}\mathbf{C}\mathbf{P}^{-1}$
Gradients describe the slope and curvature of the mean population fitness surface, provided $\mathbf{z} \sim \text{MVN}$ and fitnesses are frequency-independent	
$\beta_i = \frac{\partial \ln \bar{W}(\boldsymbol{\mu})}{\partial \mu_i}$	$\gamma_{ij} = \frac{\partial^2 \ln \bar{W}(\boldsymbol{\mu})}{\partial \mu_i \partial \mu_j} + \beta_i \beta_j$
Gradients describe the average slope and average curvature of the individual fitness surface, provided $\mathbf{z} \sim \text{MVN}$	
$\beta_i = \int \frac{\partial w(\mathbf{z})}{\partial z_i} \phi(\mathbf{z}) d\mathbf{z}$	$\gamma_{ij} = \int \frac{\partial^2 w(\mathbf{z})}{\partial z_i \partial z_j} \phi(\mathbf{z}) d\mathbf{z}$
Gradients appear as coefficients in fitness regressions	
$w(\mathbf{z}) = a + \boldsymbol{\beta}^T(\mathbf{z} - \boldsymbol{\mu})$ $\boldsymbol{\beta}$ = slope of best linear fit	$w(\mathbf{z}) = a + \mathbf{b}^T(\mathbf{z} - \boldsymbol{\mu}) + \frac{1}{2}(\mathbf{z} - \boldsymbol{\mu})^T \boldsymbol{\gamma}(\mathbf{z} - \boldsymbol{\mu})$ $\boldsymbol{\gamma}$ = the quadratic coefficient of the best quadratic fit. $\mathbf{b} = \boldsymbol{\beta}$ when $\mathbf{z} \sim \text{MVN}$
Gradients appear as coefficients in evolutionary equations when $(\mathbf{z}, \mathbf{g}) \sim \text{MVN}$	
$\mathbf{R} = \mathbf{G}\boldsymbol{\beta}$	$\mathbf{G}^* - \mathbf{G} = \mathbf{G}(\boldsymbol{\gamma} - \boldsymbol{\beta}\boldsymbol{\beta}^T)\mathbf{G}$

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

$$w(\mathbf{z}) = 1 + \sum_{i=1}^n b_i z_i + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \gamma_{ij} z_i z_j = 1 + \mathbf{b}^T \mathbf{z} + \frac{1}{2} \mathbf{z}^T \boldsymbol{\gamma} \mathbf{z} \quad (49)$$

If $\mathbf{z} \sim \text{MVN}$, then $\mathbf{b} = \boldsymbol{\beta}$ (the vector of coefficients of the best *linear* fit). As an aside, if we regard Equation 49 as a second-order Taylor series approximation of $w(\mathbf{z})$, \mathbf{b} and $\boldsymbol{\gamma}$ 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 49 is determined by the matrix $\boldsymbol{\gamma}$. Even though a quadratic is the simplest curved surface, its geometry can still be very difficult to visualize.

We start our exploration of this geometry by considering the gradient of this best-fit quadratic fitness surface. Taking the gradient of Equation 49 gives

$$\nabla_{\mathbf{z}}[w(\mathbf{z})] = \mathbf{b} + \gamma \mathbf{z} \quad (50)$$

Thus, 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 $\mathbf{b} + \gamma \mathbf{z}$ (when $\boldsymbol{\mu} = \mathbf{0}$).

Solving $\nabla_{\mathbf{z}}[w(\mathbf{z})] = \mathbf{0}$, a point \mathbf{z}_o is a candidate for a local extremum (stationary point) if $\gamma \mathbf{z}_o = -\mathbf{b}$. When γ is nonsingular,

$$\mathbf{z}_o = -\gamma^{-1} \mathbf{b} \quad (51a)$$

is the unique stationary point of this quadratic surface. Substituting into Equation 49, the expected individual fitness at this point is

$$w_o = a + \frac{1}{2} \mathbf{b}^T \mathbf{z}_o \quad (51b)$$

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 γ . Thus \mathbf{z}_o is a local minimum if γ is positive-definite (all eigenvalues are positive), a local maximum if γ is negative-definite (all eigenvalues are negative), or a saddle point if the eigenvalues differ in sign. If γ is singular (has at least one zero eigenvalue) then there is no unique stationary point. An example of this is seen in Figure 8 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 γ 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 only a single stationary point, which is given by Equation 51a 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 at this point to further our discussion of the geometry of the quadratic fitness surface, which is defined by the geometry of the matrix γ .

Matrix transformations consist of two basic operations, rotations (changes in the direction of a vector) and scalings (changes in its length). We can partition a matrix transformation into these two basic operations by using **orthonormal** matrices. Writing a square matrix \mathbf{U} as $\mathbf{U} = (\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n)$ 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**, or **orthogonal** and satisfy

$$\mathbf{U}^T \mathbf{U} = \mathbf{U} \mathbf{U}^T = \mathbf{I} \quad (52a)$$

Hence, the inverse of \mathbf{U} is its inverse,

$$\mathbf{U}^T = \mathbf{U}^{-1} \quad (52b)$$

The coordinate transformation induced by an orthonormal matrix has a very simple geometric interpretation in that it is a **rigid rotation** of the original coordinate system — all axes of the original coordinate are simply rotated by the same angle to create the new coordinate system. To see this, note first that orthonormal matrices preserve all inner products. Taking $\mathbf{y}_1 = \mathbf{U} \mathbf{x}_1$ and $\mathbf{y}_2 = \mathbf{U} \mathbf{x}_2$,

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

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

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

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

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

$$\mathbf{A} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^T \quad (53a)$$

where $\mathbf{\Lambda}$ 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 i th eigenvalue and its associated unit eigenvector of \mathbf{A} , then

$$\mathbf{A} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n) = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \dots & \dots & \lambda_n \end{pmatrix} \quad (53b)$$

and

$$\mathbf{U} = (\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n) \quad (53c)$$

Geometrically, \mathbf{U} describes a rigid rotation of the original coordinate system while $\mathbf{\Lambda}$ is the amount by which unit lengths in the original coordinate system are scaled in the transformed system. Using Equation 53a, it is easy to show that

$$\mathbf{A}^{-1} = \mathbf{U} \mathbf{\Lambda}^{-1} \mathbf{U}^T \quad (54a)$$

$$\mathbf{A}^{1/2} = \mathbf{U} \mathbf{\Lambda}^{1/2} \mathbf{U}^T \quad (54b)$$

where the **square root matrix** $\mathbf{A}^{1/2}$ (which is also symmetric) satisfies $\mathbf{A}^{1/2} \mathbf{A}^{1/2} = \mathbf{A}$. Since $\mathbf{\Lambda}$ is diagonal, the i th diagonal elements of $\mathbf{\Lambda}^{-1}$ and $\mathbf{\Lambda}^{1/2}$ are λ_i^{-1} and $\lambda_i^{1/2}$ respectively, implying that if λ_i is an eigenvalue of \mathbf{A} , then λ_i^{-1} and $\sqrt{\lambda_i}$ are eigenvalues of \mathbf{A}^{-1} and $\mathbf{A}^{1/2}$. Note that Equations 54a/b imply that \mathbf{A} , \mathbf{A}^{-1} and $\mathbf{A}^{1/2}$ all have the same eigenvectors. Finally, using Equation 53a 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} ,

$$\begin{aligned} \mathbf{U}^T \mathbf{A} \mathbf{U} &= \mathbf{U}^T (\mathbf{U} \mathbf{\Lambda} \mathbf{U}^T) \mathbf{U} = (\mathbf{U}^T \mathbf{U}) \mathbf{\Lambda} (\mathbf{U}^T \mathbf{U}) \\ &= \mathbf{\Lambda} \end{aligned} \quad (54)$$

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.*

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 12. 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 convex (stabilizing) selection on both z_1 and z_2 , with the convex selection surface perhaps rotated due to selection for correlations between z_1 and z_2 . The first caveat is that negative curvature, 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 8 shows, the nature of the fitness surface is very dependent on the amount of selection for correlations between z_1 and z_2 . Figure 8 considers the surfaces associated with three different values of γ_{12} under the assumption that $\mathbf{b} = \mathbf{0}$. 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 $\gamma_{12} = 0.25$ (Figure 8A), the individual fitness surface indeed shows convex selection in both characters. However, when $\gamma_{12} = \sqrt{2} \simeq 1.42$ (Figure 8B), the fitness surface has a ridge in one direction, with convex selection in the other. When $\gamma_{12} = 4$ (Figure 8C), 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 $\gamma_{12} > 0$.

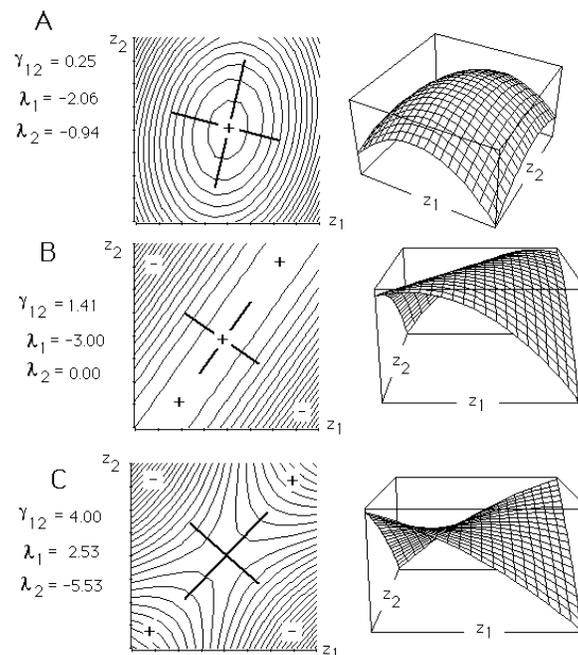


Figure 8. Three quadratic fitness surfaces, all of which have $\gamma_{11} = -2$ and $\gamma_{22} = -1$ and $\mathbf{b} = \mathbf{0}$ (i.e., no directional selection). On the left are curves of equal fitness values, with peaks being represented by a +, and valleys by a -. Axes of symmetry of the surface (the canonical or principal axes of γ) are denoted by the thick lines. These axes correspond to the eigenvectors of γ . 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 convex 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 γ is singular. The resulting fitness surface shows a ridge in one direction with strong convex selection in the other. **Bottom:** When $\gamma_{12} = 4$, the fitness surface now shows a saddle point, with convex selection along one of the canonical axes of the fitness surface and concave selection along the other.

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. From Equation 54, if we consider the matrix \mathbf{U} whose columns are the eigenvectors 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, as

$$\begin{aligned}
 w(\mathbf{z}) &= a + \mathbf{b}^T \mathbf{U} \mathbf{y} + \frac{1}{2} (\mathbf{U} \mathbf{y})^T \gamma (\mathbf{U} \mathbf{y}) \\
 &= a + \mathbf{b}^T \mathbf{U} \mathbf{y} + \frac{1}{2} \mathbf{y}^T (\mathbf{U}^T \gamma \mathbf{U}) \mathbf{y} \\
 &= a + \mathbf{b}^T \mathbf{U} \mathbf{y} + \frac{1}{2} \mathbf{y}^T \mathbf{A} \mathbf{y} \\
 &= a + \sum_{i=1}^n \theta_i y_i + \frac{1}{2} \sum_{i=1}^n \lambda_i y_i^2
 \end{aligned} \tag{55}$$

where $\theta_i = \mathbf{b}^T \mathbf{e}_i$, $y_i = \mathbf{e}_i^T \mathbf{z}$, with λ_i and \mathbf{e}_i the eigenvalues and associated unit eigenvectors of γ . Alternatively, if a stationary point \mathbf{z}_0 exists (e.g., γ is nonsingular), 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

$$w(\mathbf{z}) = w_o + \frac{1}{2} \mathbf{y}^T \mathbf{A} \mathbf{y} = w_o + \frac{1}{2} \sum_{i=1}^n \lambda_i y_i^2 \tag{56}$$

where $y_i = \mathbf{e}_i^T (\mathbf{z} - \mathbf{z}_0)$ and w_o is given by Equation 51b. Equation 55 is called the **A canonical form** and Equation 56 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 of the quadratic surface is determined by the eigenvectors of γ while the eigenvalues of γ determine the nature and amount of curvature of the surface along each canonical axis. Along the axis defined by y_i , the individual fitness function has positive curvature (is concave) if $\lambda_i > 0$, has negative curvature (is convex) if $\lambda_i < 0$, or has no curvature (is 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 is measured by λ_i .

Returning to Figure 8, 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 $\theta_i = \mathbf{b}^T \mathbf{e}_i > 0$ 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 8B, 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 55, the fitness change along a particular axis (here given by \mathbf{e}_i) is $\theta_i y_i + (\lambda_i/2) y_i^2$. If $|\theta_i| \gg |\lambda_i|$, 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.

Strength of Selection: γ_{ii} Versus λ

We have seen that the 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 stress this point, noting that in an analysis of 19 studies that $|\gamma_{ii}|_{max} < |\lambda|_{max}$. Thus, studies (such as Kingsolver et al. 2001) that report weak values for quadratic selection are potentially biased, as they simply used γ_{ii} values, rather than the full geometry of γ , as described by the eigenvectors.

Example 13. 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.
