Measuring Multivariate Selection

Bruce Walsh lecture notes
Uppsala EQG 2012 course
version 6 Feb 2012

Change in the mean vector: The Directional Selection Differential S

The multivariate extension of S is to consider the vector

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

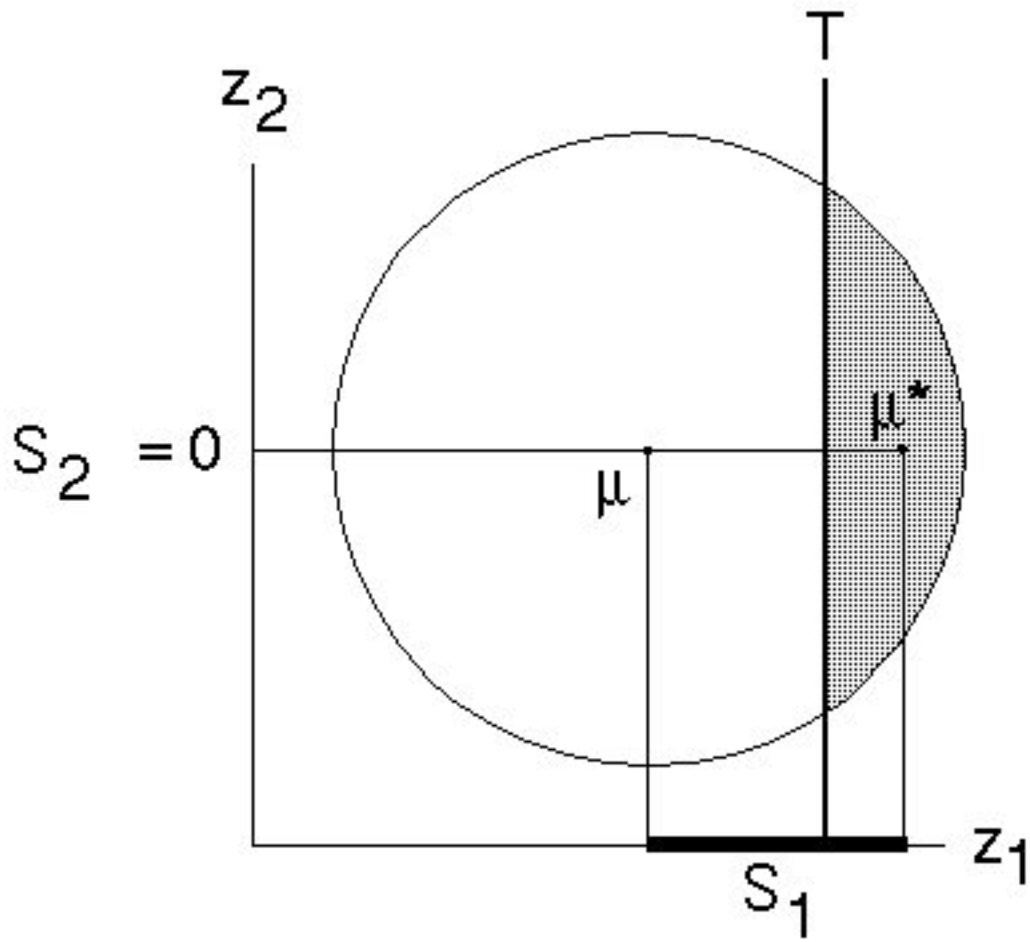
Hence, this is simply the vector of univariate directional Selection differentials

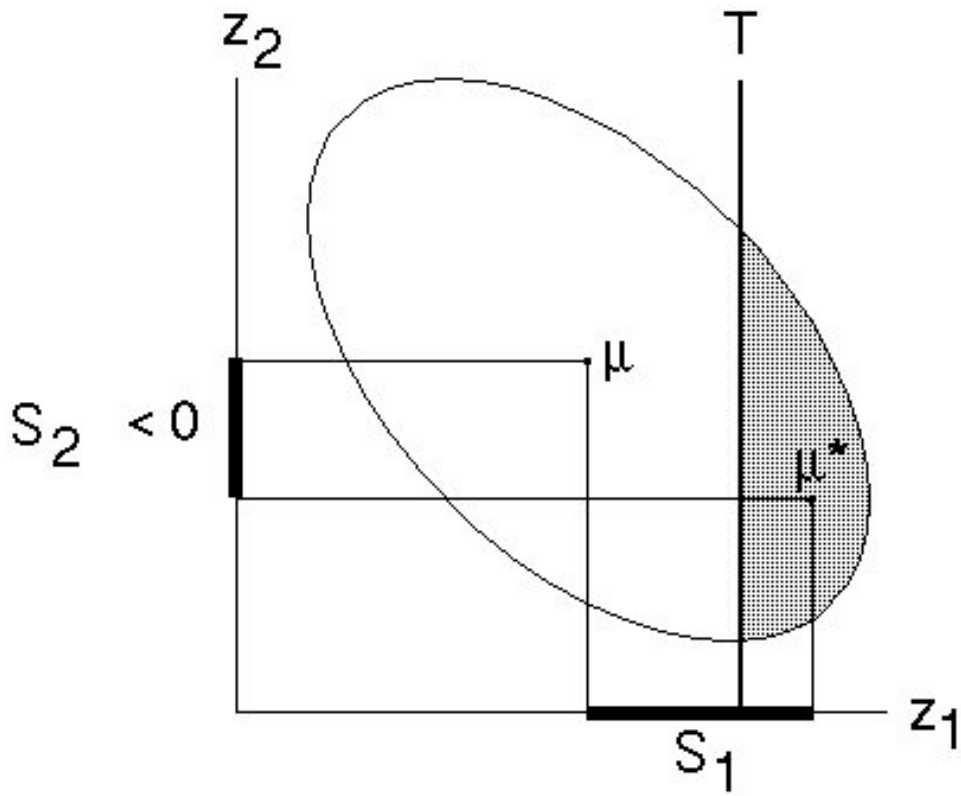
Since S_i is the i-th selection differential, the Robertson-Price identity holds for each element, with $S_i = \sigma(w, z_i)$

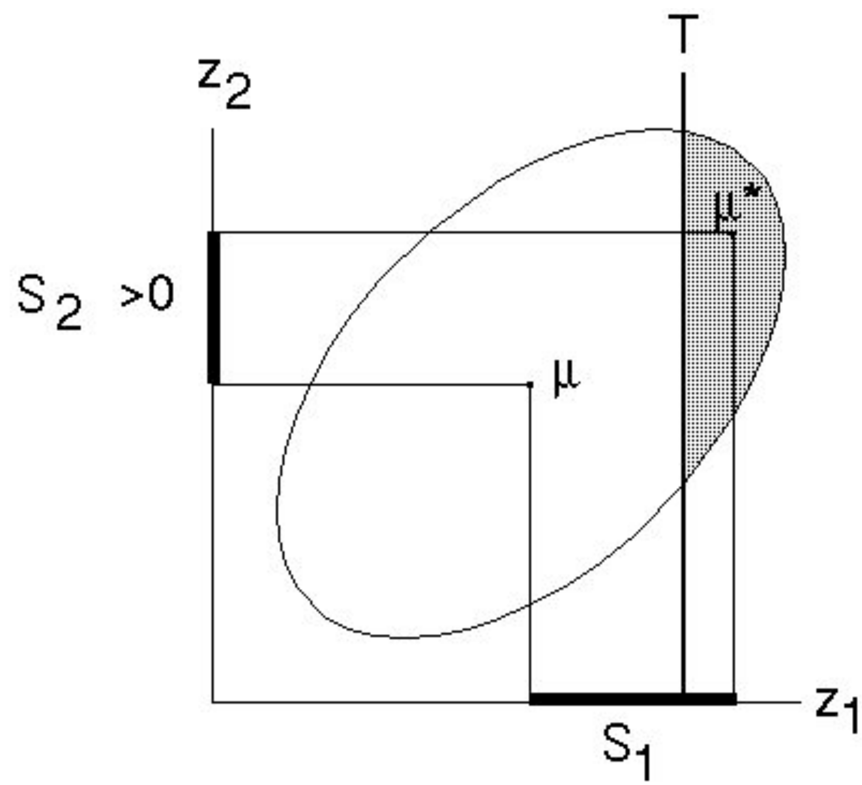
The i-th differential is bounded by I, the opportunity of selection,

$$\frac{|S_i|}{\sigma_{z_i}} \le \sqrt{I}$$

The observed change in a mean, S_i (the directional selection differential) gives a very misleading picture as to which traits are under direct selection. Direct selection on phenotypically-correlated traits also changes the mean (within a generation).




z_2
T
μ
S_2 < 0
μ^*
z_1
S_1


T
z_2
μ^*
S_2 >0
μ
z_1
S_1

The Directional Selection Gradient

The directional selection differential S gives a very misleading picture as to the nature of selection.

The directional selection gradient, β, resolves this issue. Let S be the vector of selection differentials and P the phenotypic variance-covariance matrix. Since S is a vector of covariances, it immediately follows that the vector of partial regressions for the traits on relative fitness is given by

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

Vector of covariances between w and z_i

Thus the partial regression of relative fitness on trait value is given by

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

Since we can write $S = P\beta$,

β_i gives the change in relative fitness, holding all other measured traits constant, when we increase trait i by one unit.

Direct selection (β_j non-zero) on traits phenotypically correlated with trait i (P_{ij} non-zero)

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

Direction selection on trait i, β_i non-zero

β is a gradient vector on the mean fitness surface

Recall from vector calculus that the gradient operator is given by

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

β is the gradient (wrt the population mean) of the mean fitness surface,

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

β is a gradient vector on the mean fitness surface

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

Hence, β gives the direction the current mean should change in to maximize the local change in mean fitness

β is also the average gradient of the individual fitness surface over the phenotypic distribution,

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

Changes in the Covariance Matrix: The Quadratic Selection Differential C

When considering a vector of n traits, we follow the change in n means, and n(n-1)/2 variances and covariances.

By analogy with the univariate version of C, for n traits C is now an n x n matrix, with ij-th element

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

Lande and Arnold (1983) showed that such a C is given by

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

Phenotypic covariance matrix after selection

As was true in the univariate case, directional selection alters the variances and covariances. suppose the covariance between quadratic deviations and fitness are zero. In this case,

$$P^*_{ij} - P_{ij} = -S_i S_j$$

The variance of a trait is reduced by directional selection.

The change in the covariance depends on the direction of directional selection on both traits.

If both traits are selected in the same direction, their phenotypic covariance is reduced

If both traits are selected in opposite directions, their phenotypic covariance is increased

The opportunity for selection I bounds the possible change in C_{ij}. Assuming multivariate normality,

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

Phenotypic correlation between i and j

Now consider the multivariate quadratic regression predicting relative fitness from our vector of n trait values,

$$w = a + \sum_{j=1}^{n} b_j \, z_j + \frac{1}{2} \sum_{j=1}^{n} \sum_{k=1}^{n} d_{ij} \, z_i z_j + e$$

We can more compactly write this in matrix notation as $w = a + b^T z + (1/2) z^T D z + e$

The matrix of the best-fitting quadratic coefficients in this regression is given by γ, where the n x n matrix γ is the multivariate version of the quadratic selection gradient,

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

Which also follows from regression theory

Here, γ_{ij} measures the direct selection on the combination of i and j

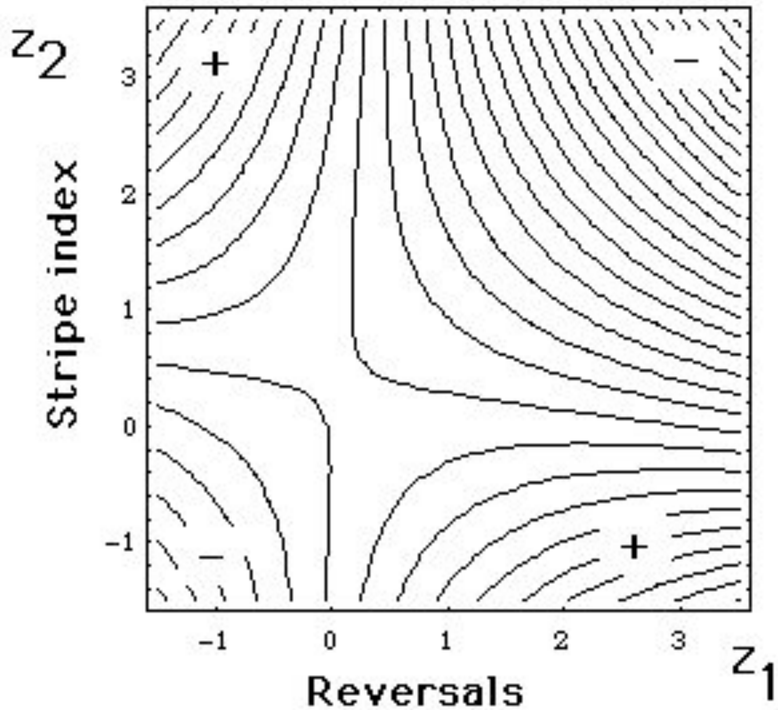
- $\gamma_{ii} < 0$. Convex selection on trait i. Selection to decrease variance
- $\gamma_{ii} > 0$. Concave selection on trait i. Selection to increase variance
- $\gamma_{ij} > 0$. Correlational selection on traits i & j to increase their correlation.
- $\gamma_{ij} < 0$. Correlational selection on traits i & j to decrease their correlation.

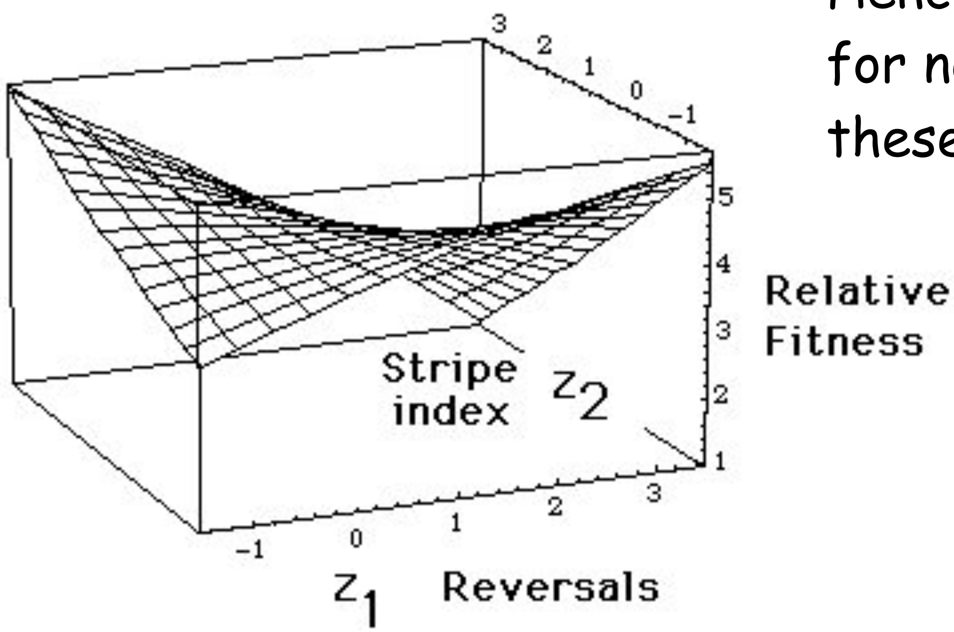
While it has been very popular to infer the nature of quadratic selection directly from the individual γ_{ij} values, as we will shortly see, this can be very misleading!

Finally, note that we can write $C = P\gamma P$, or

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

Thus, as was the case for directional differentials, the quadratic differential is caused by direct selection on a trait plus any selection on all phenotypically correlated traits.





An example of the fitness surface from a quadratic regression.

Brodie (1992) examined two anti-predator traits in garter snakes. Stripe index (pattern) and number of reversals when moving

None of the β_i or γ_{ii} were significant, but $\gamma_{ij} = -0.268 \pm 0.097$

Hence, apparent strong selection for negative covariances between these traits.

The multivariate Lande-Arnold regression

Provided the distribution of phenotypes is MVN,

$$w = 1 + \beta^T z + (1/2) z^T \gamma z + e$$

If phenotypes are *not* MVN, can still estimate the matrix γ (the quadratic selection graident) from the quadratic regression.

However, the vector of regression coefficients for the linear terms is NOT the selection gradient β.

In such cases, β is estimated from a separate linear regression,

$$w = 1 + \beta^T z + e$$

γ and the geometry of individual and mean fitness surfaces

When phenotypes are MVN, γ is the average curvature over the individual fitness surface,

Hessian matrix of the individual fitness surface. $H_{ij}(F) = \partial F / \partial z_i \partial z_j$

$$\gamma = \int \mathbf{H_z}[\, W(\mathbf{z}) \,]\, \phi(\mathbf{z})\, d\mathbf{z}$$

Phenotypic distribution of the vector z

γ and the geometry of individual and mean fitness surfaces

Under MVN, γ also describes the mean fitness surface,

$$\mathbf{H}_{\boldsymbol{\mu}}[\,\ln \overline{W}(\boldsymbol{\mu})\,] = \boldsymbol{\gamma} - \boldsymbol{\beta}\boldsymbol{\beta}^T \quad \frac{\partial \ln \overline{W}(\boldsymbol{\mu})}{\partial \mu_i \, \partial \mu_j} = \gamma_{ij} - \beta_i \beta_j$$

Curvature of the mean fitness surface around the means of i and j depends on the curvature and the product of gradients

Within-generation changes in genetic covariance matrix G as a function of γ

$$\mathbf{G}^* - \mathbf{G} = \mathbf{G}\mathbf{P}^{-1}\left(\mathbf{P}^* - \mathbf{P}\right)\mathbf{P}^{-1}\mathbf{G}$$

$$= \mathbf{G}(\boldsymbol{\gamma} - \boldsymbol{\beta}\boldsymbol{\beta}^T)\mathbf{G}$$

$$= -\mathbf{R}\,\mathbf{R}^T + \mathbf{G}\boldsymbol{\gamma}\mathbf{G}$$

$$G^*_{ij} - G_{ij} = -R_i\,R_j + \sum_{k=1}^{n}\sum_{\ell=1}^{n} \gamma_{k\ell}\,G_{ik}\,G_{\ell j}$$

Differentials measure the covariance between relative fitness and phenotypic value

Changes in means (Directional Selection)

$$S_i = \sigma(w, z_i)$$

Changes in covariance (Quadratic Selection)

$$C_{ij} = \sigma[\, w, (z_i - \mu)\,(z_j - \mu)\,]$$

The Opportunity for Selection Bounds the Differential

Changes in means (Directional Selection)

$$\frac{|S_i|}{\sigma\left(z_i\right)} \leq \sqrt{I}$$

No distributional assumptions

$$\frac{|C_{ij}|}{P_{ij}} \leq \sqrt{I}\sqrt{1+\rho_{ij}^{-2}}$$

MVN assumption

Differentials Confound Direct and Indirect Selection

Changes in means (Directional Selection)

$$\mathbf{S} = \mathbf{P}\beta \qquad\qquad S_i = \sum_{j=1}^{n} \beta_j \, P_{ij}$$

Changes in covariance (Quadratic Selection)

$$\begin{aligned} \mathbf{C} &= \mathbf{P}^* - \mathbf{P} + \mathbf{S}\mathbf{S}^T \\ &= \mathbf{P}\gamma\mathbf{P} \end{aligned}$$

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

Gradients measure the amount of Direct Selection (remove confounding effects of phenotypic correlations)

Changes in means (Directional Selection)

$$\beta = P^{-1} S$$

Changes in covariance (Quadratic Selection)

$$\gamma = P^{-1} C P^{-1}$$

$$= P^{-1} (P^{*} - P + SS^{T}) P^{-1}$$

Gradients describe the slope and curvature of the mean Population surface

(when z ~ MVN & frequency-independent fitnesses)

Changes in means (Directional Selection)

$$\beta_i = \frac{\partial \ln \overline{W}(\boldsymbol{\mu})}{\partial \mu_i}$$

Changes in covariance (Quadratic Selection)

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

Gradients describe the average slope and average curvature of the individual fitness surface (when z ~ MVN)

Changes in means (Directional Selection)

$$\beta_i = \int \frac{\partial\, w(\mathbf{z})}{\partial z_i}\, \phi(\mathbf{z})\, \mathrm{d}\mathbf{z}$$

Changes in covariance (Quadratic Selection)

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

Gradients Appear as Coefficients in Fitness regressions

Changes in means (Directional Selection)

$$w = 1 + \beta\, z^T + e$$

Changes in covariance (Quadratic Selection)

$$w = 1 + b^T z + (1/2) z^T \gamma z + e$$

$b = \beta$ when $z \sim$ MVN

Gradients Appear as coefficients in evolutionary Equations (when (z,g) ~ MVN)

Changes in means (Directional Selection)

$$R = G\beta$$

Changes in covariance (Quadratic Selection)

$$G^* - G = G(\gamma - \beta\,\beta^T)G$$

Multidimensional Quadratic Regressions

We wish to explore the geometry of an n- dimensional quadratic regression a bit more carefully.

$$w(\mathbf{z}) = 1 + \sum_{i=1}^{n} b_1 z_i + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \gamma_{ij}\, z_i z_j$$

$$= 1 + \mathbf{b}^T \mathbf{z} + \frac{1}{2} \mathbf{z}^T \boldsymbol{\gamma}\, \mathbf{z}$$

The gradient is $\nabla_{\mathbf{z}}[\, w(\mathbf{z})\,] = \mathbf{b} + \boldsymbol{\gamma}\, \mathbf{z}$

Solving for grad = 0 gives the unique stationary point z_o

$$\mathbf{z}_o = -\boldsymbol{\gamma}^{-1}\mathbf{b} \qquad\qquad w_o = 1 + \frac{1}{2}\, \mathbf{b}^T \mathbf{z}_0$$

Digression: Orthonormal and Diagonalized Matrices

In order to fully explore the geometry implied by γ, we need some additional matrix machinery.

As mentioned, matrix transformation involve rotation and scaling, and we can partition a square matrix into these two operations using orthonormal matrices, matrices whose columns are independent and of unit length (i.e., columns are orthonormal)

Column vectors u_i and u_j are orthonormal if $u_i^T u_j = 0$ for $i \neq j$, while $u_i^T u_i = 1$

The square matrix $U = (u_1, u_2, \dots, u_n)$ is orthonormal provided all of the columns are. Such a matrix is also called a unitary matrix.

Orthonormal matrices satisfy

$$U^T U = U U^T = I \qquad \text{i.e., } U^{-1} = U^T$$

Orthonormal matrices introduce rigid rotations.

The angle between vectors x and y is the same as the angle between Ux and Uy (for all x and y)

A symmetric matrix A can be diagonalized as

$$A = U \Lambda U^T$$

Geometry of diagonalization

Orthonormal matrices (rigid rotation of the original coordinate system)

$$A = U \, \Lambda U^T$$

Diagonal matrix (scaling)

If e_i is an eigenvector of A and λ_i its associated eigenvalue, then the diagonalization of A is given by

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

$$\mathbf{\Lambda} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & \lambda_n \end{pmatrix}$$

With $A = U \, \Lambda \, U^T$

Using diagonalization, it is easy to show that

Hence, if λ_i is an eigenvalue of A, then $1/\lambda_i$ is an eigenvalue of A^{-1}

$$\mathbf{A}^{-1} = \mathbf{U}\mathbf{\Lambda}^{-1}\mathbf{U}^T$$

Likewise, the eigenvectors of A and A^{-1} are identical.

Consider the square root matrix $A^{1/2}$, where
$A^{1/2} A^{1/2} = A$

$$\mathbf{A}^{1/2} = \mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{U}^T$$

A and $A^{1/2}$ have the same eigenvectors

λ_i is an eigenvalue of A,
$\lambda_i^{1/2}$ is an eigenvalue of $A^{1/2}$

Also note that $A^{-1/2}$ and A^n also have the same eigenvectors as A, with eigenvalues $\lambda_i^{-1/2}$ and λ_i^n

Finally, note that

$$U^T A U = U^T (U \Lambda U^T) U = (U^T U)\, \Lambda (U^T U)$$
$$= \Lambda$$

The effect of using such a transformation is to remove correlations on this new scale.

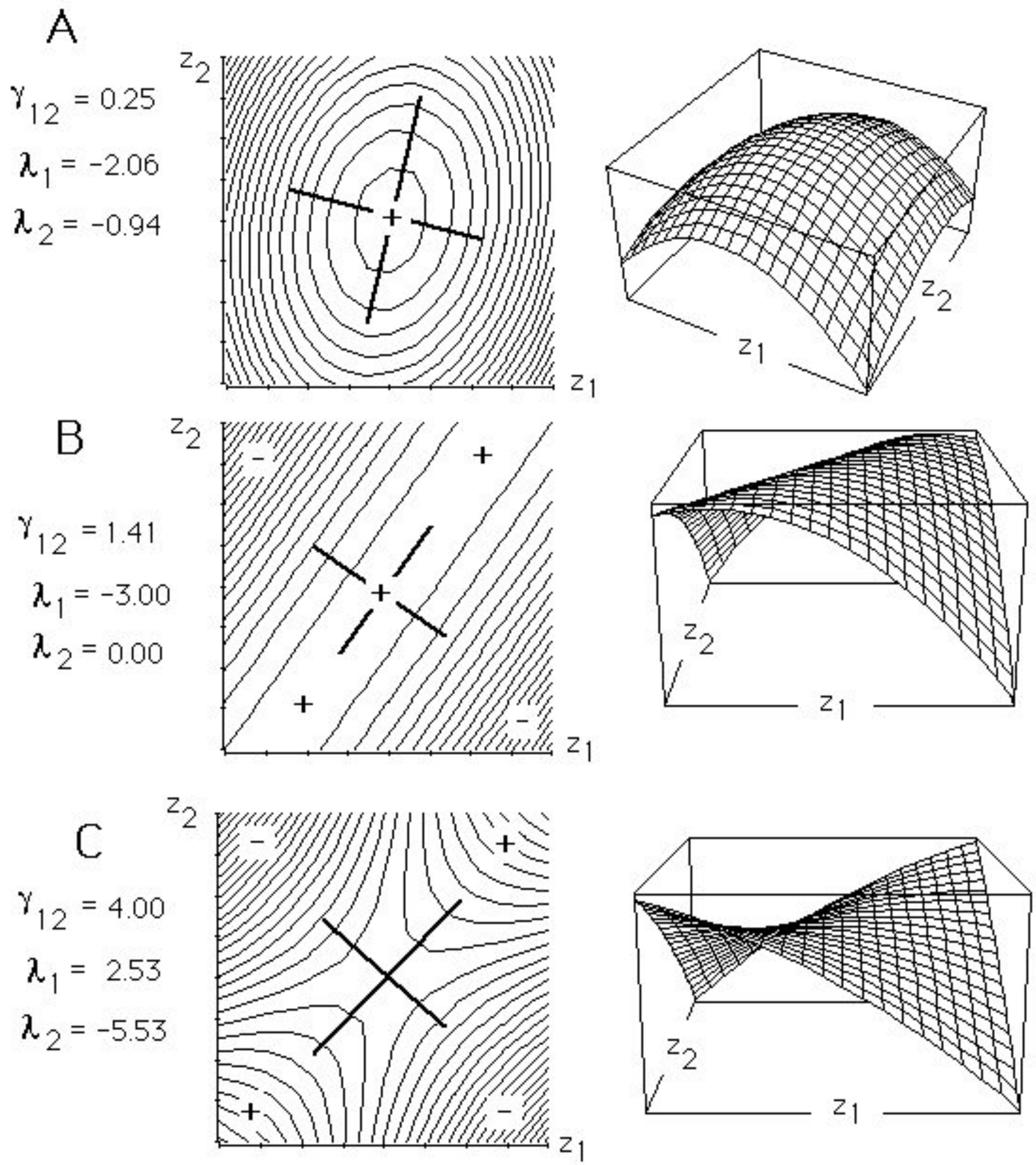
Suppose we have a γ matrix with

$\gamma_{11} = -2$

$\gamma_{22} = -1$

Looking just at these diagonal elements of γ, we might conclude convex selection on both 1 and 2

However, the actual nature of the surface critically depends on γ_{12}



To resolve this issue of inferring the geometry of the quadratic surface, we need to diagonalize γ

For even two traits, visualizing the fitness surface simply the γ_{ij} is tricky at best.

The problem is the cross-product terms γ_{ij}

By creating the appropriate new character y (character axes that are linear index of the current character values), we can remove all of the cross-product terms. In effect, there is no correlational selection among the y_i, only convex, concave, or no quadratic selection

This new vector of characters is given by $y = U^T z$, where U is the matrix of eigenvectors of γ.

Substituting $y = U^T z$, or $z = Uy$ into the Lande-Arnold regression $w = a + b^T z + (1/2) z^T \gamma z$ gives

$$w(\mathbf{z}) = a + \mathbf{b}^T \mathbf{U}\,\mathbf{y} + \frac{1}{2} (\mathbf{U}\mathbf{y})^T \gamma (\mathbf{U}\mathbf{y})$$

$$= a + {}^T \mathbf{U}\,\mathbf{y} + \frac{1}{2} \mathbf{y}^T \left(\mathbf{U}^T \gamma \mathbf{U} \right) \mathbf{y}$$

$$= a + \mathbf{b}^T \mathbf{U}\,\mathbf{y} + \frac{1}{2} \mathbf{y}^T \Lambda\, \mathbf{y}$$

Eigenvalue of γ corresponding to eigenvector e_i

$\mu_i = b^T e_i$.

if z~MVN,

$\mu_i = \beta^T e_i$

$$= a + \sum_{i=1}^{n} \mu_i\, y_i + \frac{1}{2} \sum_{i=1}^{n} \lambda_i\, y_i^2$$

$y_i = e_i^T z$

This is often called the A canonical form of the quadratic surface (Box & Draper 1987)

The fitness change along axis e_i is $\mu_i y_i + (\lambda/2)\, y_i^2$

B canonical form (Box & Draper 1987)

If γ is nonsingular, then a stationary point z_o exists, the transformation $y = U^T(z - z_o)$ removes all linear terms, leading to the so-called B canonical form

$y_i = e_i^T (z - z_o)$

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

Fitness at the equilibrium point z_o

Eigenvalue of γ

The B canonical form shifts the origin to the stationary point. Since the effect on w(z) from b^Tz is a hyperplane (tilting the whole fitness surface), the B canonical form "levels" the fitness surface, focusing entirely on its curvature (quadratic) features.

B canonical form (Box & Draper 1987)

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

- $\lambda_i > 0$: concave selection to increase the variance in y_i
- $\lambda_i < 0$: convex selection to decrease the variance in y_i
- $\lambda_i = 0$: no quadratic selection along the axes given by y_i

Strength of selection: γ_{ii} vs. λ

Blows and Brooks (2003) stress that the eigenvalues of γ, not the diagonal elements, provide a much more accurate description of the strength of selection

In an analysis of 19 studies, they note that $|\gamma_{ii}|_{max} < |\lambda|_{max}$

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

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

Weak convex selection

Strong evidence of concave selection

Weak concave selection

Eigenvalues were 0.132, 0.006, -0.038, -0.064

Subspaces of γ

Blows & Brooks (2003) note that there are several advantages to focusing on estimation of the λ_i vs. the entire matrix of γ_{ij}.

First, there are n eigenvalues vs. n(n-1) elements of γ.

Further, many of the eigenvalues are likely close to 0, so that a subspace of γ (much like a subspace of G) describes most of the variation.

Blow & Brooks suggest obtaining the eigenvalues of γ, and using these to generate transformed variables $y = U^T z$ for those eigenvalues accounting for most of the variation.

Note that the quadratic terms in the transformed regression on the y correspond to the eigenvalues, and hence GLM machinery can estimate their standard errors and confidence intervals.

Finally, Blows et al. (2004) suggest that one should consider the projection of γ into G, just like we examined the projection of β into subspaces of G.

Here we are projecting a matrix instead of a vector, but the basic ideal holds.

First, construct a matrix B formed by a subset of γ, namely the eigenvectors corresponding to k (< n/2) leading eigenvalues, $B = (e_1, \ldots, e_k)$

Form a similar matrix with k eigenvectors of G

Using results of Krzanowski (1979), the γ and G subspaces can be compared by the matrix

$$S = A^T B B^T A$$

The eigenvalues of S describe the angles between the orthogonal axes of the matrices A and B

Specifically, the smallest angle is given by $\cos^{-1}(\lambda_1^{1/2})$, where λ_1 is the leading eigenvalue of S.

Path Analysis models

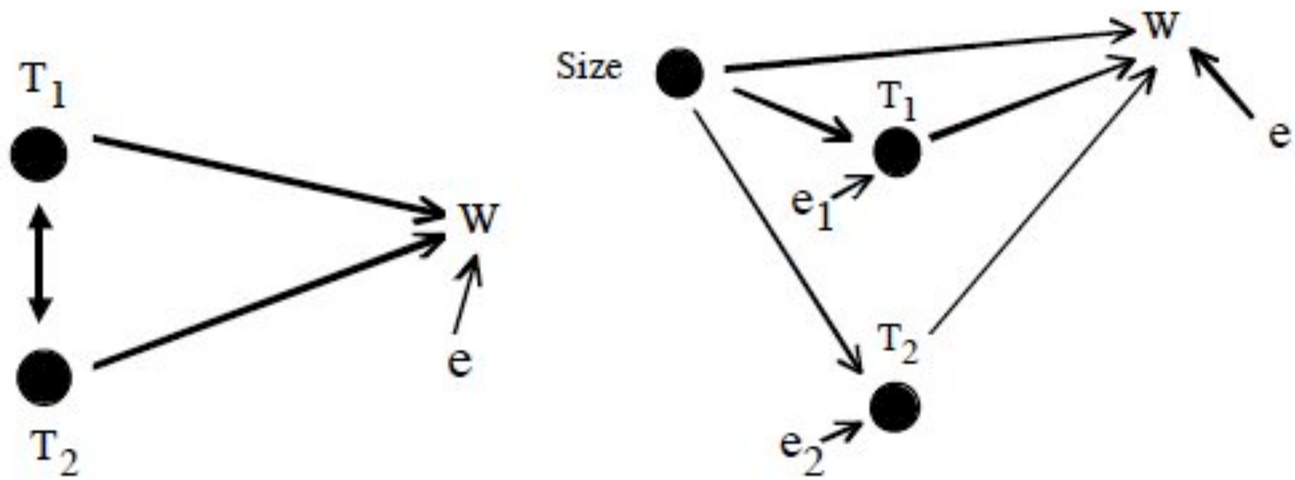


Figure 29.13. Morphological traits T_1 and T_2 are being examined for their effects on fitness w. **Left:** Under a standard Lande-Arnold regression, we allow for direct effects of both traits on fitness w as well as for indirect effects from correlations between these two traits (double-headed arrow). Here e represents the additional variance in fitness not accounted for by the paths through T_1 and T_2. **Right:** Suppose that size influences both traits as well as fitness. When we ignore size, it is possible that all of the correlations between trait value and fitness arise because each trait is serving as a surrogate for size. Since this is a missing trait (size is not included in the analysis), this effect would not be picked up by a Lande-Arnold regression. By extracting a surrogate measure for size (such as PC 1) from the data, we can include this in the analysis. The path diagram allows for direct paths between size and fitness, and between both trait and fitness, after size effects have been removed.

Contextual Analysis: Incorporating levels of selection

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

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

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

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

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