Appendix 3

Derivatives of Vectors and

Vector-Valued Functions

Please email any comments/corrections to: **jbwalsh@u.arizona.edu**

DERIVATIVES OF VECTORS AND VECTOR-VALUED FUNCTIONS

The **gradient** (or **gradient vector**) of a scalar function of a vector variable is obtained by taking partial derivatives of the function with respect to each variable. In matrix notation, the gradient operator is

$$\nabla_{\mathbf{x}}[\,f\,] = \frac{\partial f}{\partial \mathbf{x}} = \begin{pmatrix} \dfrac{\partial f}{\partial x_1} \\ \dfrac{\partial f}{\partial x_2} \\ \vdots \\ \dfrac{\partial f}{\partial x_n} \end{pmatrix}$$

The gradient at point $\mathbf{x}_o$ corresponds to a vector indicating the direction of steepest ascent of the function at that point (the multivariate slope of f at the point $\mathbf{x}_o$). For example $f(\mathbf{x}) = \mathbf{x}^T\mathbf{x}$ has gradient vector $2\mathbf{x}$. At the point $\mathbf{x}_o$, $\mathbf{x}^T\mathbf{x}$ locally increases most rapidly if we change $\mathbf{x}$ in the same the direction as the vector going from point $\mathbf{x}_o$ to point $\mathbf{x}_o + 2\delta\,\mathbf{x}_o$, where δ is a small positive value.

For a vector $\mathbf{A}$ and matrix $\mathbf{A}$ of constants, it can easily be shown (e.g., Morrison 1976, Graham 1981, Searle 1982) that

$$\nabla_{\mathbf{x}}\left[\,\mathbf{a}^T\mathbf{x}\,\right] = \nabla_{\mathbf{x}}\left[\,\mathbf{x}^T\mathbf{a}\,\right] = \mathbf{A} \tag{A3.1a}$$

$$\nabla_{\mathbf{x}}\left[\,\mathbf{A}\mathbf{x}\,\right] = \mathbf{A}^T \tag{A3.1b}$$

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

$$\nabla_{\mathbf{x}}\left[\,\mathbf{x}^T\mathbf{A}\mathbf{x}\,\right] = 2\cdot\mathbf{A}\mathbf{x} \tag{A3.1c}$$

$$\nabla_{\mathbf{x}}\left[\,(\mathbf{x}-\mathbf{A})^T\mathbf{A}(\mathbf{x}-\mathbf{A})\,\right] = 2\cdot\mathbf{A}(\mathbf{x}-\mathbf{A}) \tag{A3.1 d}$$

$$\nabla_{\mathbf{x}}\left[\,(\mathbf{A}-\mathbf{x})^T\mathbf{A}(\mathbf{A}-\mathbf{x})\,\right] = -2\cdot\mathbf{A}(\mathbf{A}-\mathbf{x}) \tag{A3.1e}$$

Taking $\mathbf{A} = \mathbf{I}$, Equation A3.1c implies

$$\nabla_{\mathbf{x}}\left[\,\mathbf{x}^T\mathbf{x}\,\right] = \nabla_{\mathbf{x}}\left[\,\mathbf{x}^T\,\mathbf{I}\,\mathbf{x}\,\right] = 2\cdot\mathbf{I}\,\mathbf{x} = 2\cdot\mathbf{x} \tag{A3.1f}$$

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

$$\nabla_{\mathbf{x}}\left[\,\exp[\,f(\mathbf{x})\,]\,\right] = \exp[\,f(\mathbf{x})\,]\cdot\nabla_{\mathbf{x}}\left[\,f(\mathbf{x})\,\right] \tag{A3.1g}$$

$$\nabla_{\mathbf{x}}\left[\,\ln[\,f(\mathbf{x})\,]\,\right] = \frac{1}{f(\mathbf{x})}\cdot\nabla_{\mathbf{x}}\left[\,f(\mathbf{x})\,\right] \tag{A3.1h}$$

Example 11. Writing the MVN distribution as

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

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

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

Applying Equation A3.1d gives

$$\nabla_{\mathbf{x}}\left[\,\varphi(\mathbf{x})\,\right] = -\varphi(\mathbf{x})\cdot\mathbf{V}_{\mathbf{X}}^{-1}\,(\mathbf{x}-\boldsymbol{\mu}) \tag{A3.2a}$$

Note here that $\varphi(\mathbf{x})$ is a scalar and hence its order of multiplication does not matter, while the order of the other variables (being matrices) is critical. Similarly, we can consider the MVN as a function of the mean vector $\boldsymbol{\mu}$, in which case Equation A3.1e implies

$$\nabla_{\boldsymbol{\mu}}\left[\,\varphi(\mathbf{x},\boldsymbol{\mu})\,\right] = \varphi(\mathbf{x},\boldsymbol{\mu})\cdot\mathbf{V}_{\mathbf{X}}^{-1}\,(\mathbf{x}-\boldsymbol{\mu}) \tag{A3.2b}$$

Example 2. Lande (1979) showed that $\nabla_{\boldsymbol{\mu}}[\,\ln\overline{W}(\boldsymbol{\mu})\,] = \boldsymbol{\beta}$ when phenotypes are multivariate normal. Hence, the increase in mean population fitness is maximized if mean character values change in the same direction as the vector $\boldsymbol{\beta}$. To see this, from Equation A3.1h we have $\nabla_{\boldsymbol{\mu}}[\,\ln\overline{W}(\boldsymbol{\mu})\,] = \overline{W}^{\,-1}\,\nabla_{\boldsymbol{\mu}}[\,\overline{W}(\boldsymbol{\mu})\,]$. Writing mean fitness as $\overline{W}(\boldsymbol{\mu}) = \int W(\mathbf{z})\,\varphi(\mathbf{z},\boldsymbol{\mu})\,d\mathbf{z}$ and taking the gradient through the integral gives

$$\nabla_{\boldsymbol{\mu}}[\,\ln\overline{W}(\boldsymbol{\mu})\,] = \nabla_{\boldsymbol{\mu}}\left[\int\frac{W(\mathbf{z})}{\overline{W}}\,\varphi(\mathbf{z},\boldsymbol{\mu})\,\mathrm{d}\mathbf{z}\right] = \int w(\mathbf{z})\,\nabla_{\boldsymbol{\mu}}\left[\,\varphi(\mathbf{z},\boldsymbol{\mu})\,\right]\;\mathrm{d}\mathbf{z}$$

If individual fitnesses are themselves functions of the population mean (are frequency-dependent), then a second integral appears as we can no longer assume $\nabla_{\boldsymbol{\mu}} \left[\, w(\mathbf{z}) \,\right] = 0$ (see Equation 15.3b????). Applying Equation 15.40b, we can rewrite this integral as

$$
\begin{aligned}
\int w(\mathbf{z}) \, \nabla_{\boldsymbol{\mu}} \left[\, \varphi(\mathbf{z}, \boldsymbol{\mu}) \,\right] \, d\mathbf{z} &= \int w(\mathbf{z}) \, \varphi(\mathbf{z}) \, \mathbf{P}^{\,-1} \, (\mathbf{z} - \boldsymbol{\mu}) \, d\mathbf{z} \\
&= \mathbf{P}^{\,-1} \left(\int \mathbf{z} \, w(\mathbf{z}) \, \varphi(\mathbf{z}) \, d\mathbf{z} - \boldsymbol{\mu} \int w(\mathbf{z}) \, \varphi(\mathbf{z}) \, d\mathbf{z} \right) \\
&= \mathbf{P}^{\,-1} (\boldsymbol{\mu}^* - \boldsymbol{\mu}) = \mathbf{P}^{\,-1} \mathbf{s} = \boldsymbol{\beta}
\end{aligned}
$$

which follows since the first integral is the mean character value after selection and the second equals one as $E[\, w \,] = 1$ by definition.

Example 3. Compute $\nabla_{\boldsymbol{\mu}} \left[\, \ln \overline{W}(\boldsymbol{\mu}, \mathbf{P}) \,\right]$ for the generalized Gaussian fitness function (Equation 15.33). From Equations A3.1g and 15.36a, we have

$$
\nabla_{\boldsymbol{\mu}} \left[\, \ln \overline{W}(\boldsymbol{\mu}, \mathbf{P}) \,\right] = \nabla_{\boldsymbol{\mu}} \left[\ln \left(\frac{|\, \mathbf{P}^* \,|}{|\, \mathbf{P} \,|} \right) \right] - \frac{1}{2} \cdot \nabla_{\boldsymbol{\mu}} \left[\, f(\boldsymbol{\mu}) \,\right] = -\frac{1}{2} \cdot \nabla_{\boldsymbol{\mu}} \left[\, f(\boldsymbol{\mu}) \,\right]
$$

where $f(\boldsymbol{\mu})$ is given by Equation 15.36b and $\mathbf{P}^*$ by Equation 15.35b (the first term is zero because $\mathbf{P}$ and $\mathbf{P}^*$ are independent of $\boldsymbol{\mu}$). Ignoring terms of f not containing $\boldsymbol{\mu}$ since the gradient of these (with respect to $\boldsymbol{\mu}$) is zero,

$$
\nabla_{\boldsymbol{\mu}} \left[\, f(\boldsymbol{\mu}) \,\right] = \nabla_{\boldsymbol{\mu}} \left[\, \boldsymbol{\mu}^T \mathbf{P}^{-1} \left(\mathbf{I} - \mathbf{P}^* \mathbf{P}^{-1} \right) \boldsymbol{\mu} \,\right] - 2 \cdot \nabla_{\boldsymbol{\mu}} \left[\mathbf{b}^T \mathbf{P}^{-1} \boldsymbol{\mu} \right]
$$

where $\mathbf{b} = \mathbf{W} \boldsymbol{\theta} + \boldsymbol{\alpha}$. Applying Equations A3.1b/c,

$$
\begin{aligned}
\nabla_{\boldsymbol{\mu}} \left[\, \boldsymbol{\mu}^T \mathbf{P}^{-1} \left(\mathbf{I} - \mathbf{P}^* \mathbf{P}^{-1} \right) \boldsymbol{\mu} \,\right] &= 2 \cdot \mathbf{P}^{-1} \left(\mathbf{I} - \mathbf{P}^* \mathbf{P}^{-1} \right) \boldsymbol{\mu} \\
\nabla_{\boldsymbol{\mu}} \left[\mathbf{b}^T \mathbf{P}^{-1} \boldsymbol{\mu} \right] &= \left(\mathbf{b}^T \mathbf{P}^{-1} \right)^T = \mathbf{P}^{-1} \mathbf{b}
\end{aligned}
$$

Hence,

$$
\nabla_{\boldsymbol{\mu}} \left[\, \ln \overline{W}(\boldsymbol{\mu}, \mathbf{P}) \,\right] = \mathbf{P}^{-1} \left[\, \left(\mathbf{P}^* \mathbf{P}^{-1} - \mathbf{I} \right) \boldsymbol{\mu} + \mathbf{b} \,\right] \tag{A3.3a}
$$

Using the definitions of $\mathbf{P}^*$ and $\mathbf{b}$, we can (eventually) express this as

$$
\nabla_{\boldsymbol{\mu}} \left[\, \ln \overline{W}(\boldsymbol{\mu}, \mathbf{P}) \,\right] = \mathbf{P}^{-1} \mathbf{W}^{-1} \left(\mathbf{W}^{-1} + \mathbf{P} \right)^{-1} \mathbf{P} \left(\mathbf{W}(\boldsymbol{\theta} - \boldsymbol{\mu}) + \boldsymbol{\alpha} \right) \tag{A3.3b}
$$

showing (from Equation 15.37a) that when $\mathbf{z} \sim$MVN, this gradient equals $\mathbf{P}^{-1} \mathbf{s} = \boldsymbol{\beta}$.

Example 4. Consider obtaining the least-squares solution for the general linear model, $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$, where we wish to find the value of $\boldsymbol{\beta}$ that minimizes the residual error given $\mathbf{y}$ and $\mathbf{X}$. In matrix form,

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

where the last step follows from Equation 7.18. To find the vector $\boldsymbol{\beta}$ that minimizes $\mathbf{e}^T\mathbf{e}$, taking the derivative with respect to $\boldsymbol{\beta}$ and using Equations A3.1a/c gives

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

Setting this equal to zero gives $\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} = \mathbf{X}^T\mathbf{y}$ giving

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

The Hessian Matrix, Local Maxima/minima, and Multidimensional Taylor Series

In univariate calculus, local extrema of a function occur when the slope (first derivative) is zero. The multivariate extension of this is that the gradient vector is zero, so that the slope of the function with respect to all variables is zero. A point $\mathbf{x}_e$ where this occurs is called a **stationary** or **equilibrium** point, and corresponds to either a local maximum, minimum, saddle point or inflection point. As with the calculus of single variables, determining which of these is true depends on the second derivative. With n variables, the appropriate generalization is the **hessian** matrix

$$\mathbf{H}_{\mathbf{x}}[\,f\,] = \nabla_{\mathbf{x}}\left[\left(\nabla_{\mathbf{x}}\,[\,f\,]\right)^T\right] = \frac{\partial^2 f}{\partial\mathbf{x}\,\partial\mathbf{x}^T} = \begin{pmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \cdots & \dfrac{\partial^2 f}{\partial x_1\,\partial x_n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f}{\partial x_1\,\partial x_n} & \cdots & \dfrac{\partial^2 f}{\partial x_n^2} \end{pmatrix} \tag{A3.4}$$

This matrix is symmetric, as mixed partials are equal under suitable continuity conditions, and measures the local curvature of the function.

Example 5. Compute $\mathbf{H_x}\left[\,\varphi(\mathbf{x})\,\right]$, the hessian matrix for the multivariate normal distribution. Recalling from Equation 15.40a that $\nabla_{\mathbf{x}}\left[\,\varphi(\mathbf{x})\,\right] = -\varphi(\mathbf{x}) \cdot \mathbf{V}_{\mathbf{x}}^{-1}\,(\mathbf{x}-\boldsymbol{\mu})$, we have

$$\begin{aligned}
\mathbf{H_x}\left[\,\varphi(\mathbf{x})\,\right] &= \nabla_{\mathbf{x}}\left[\left(\nabla_{\mathbf{x}}\left[\,\varphi(\mathbf{x})\,\right]\right)^T\right] \\
&= -\nabla_{\mathbf{x}}\left[\,\varphi(\mathbf{x})\cdot(\mathbf{x}-\boldsymbol{\mu})^T\mathbf{V}_{\mathbf{x}}^{-1}\,\right] \\
&= -\nabla_{\mathbf{x}}\left[\,\varphi(\mathbf{x})\,\right]\cdot(\mathbf{x}-\boldsymbol{\mu})^T\mathbf{V}_{\mathbf{x}}^{-1} - \varphi(\mathbf{x})\cdot\nabla_{\mathbf{x}}\left[\,(\mathbf{x}-\boldsymbol{\mu})^T\mathbf{V}_{\mathbf{x}}^{-1}\,\right] \\
&= \varphi(\mathbf{x})\cdot\left(\mathbf{V}_{\mathbf{x}}^{-1}\,(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^T\mathbf{V}_{\mathbf{x}}^{-1} - \mathbf{V}_{\mathbf{x}}^{-1}\right) \qquad \text{(A3.5a)}
\end{aligned}$$

Likewise,

$$\mathbf{H}_{\boldsymbol{\mu}}\left[\,\varphi(\mathbf{x},\boldsymbol{\mu})\,\right] = \varphi(\mathbf{x},\boldsymbol{\mu})\cdot\left(\mathbf{V}_{\mathbf{x}}^{-1}\,(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^T\mathbf{V}_{\mathbf{x}}^{-1} - \mathbf{V}_{\mathbf{x}}^{-1}\right) \qquad \text{(A3.5b)}$$

To see how the hessian matrix determines the nature of equilibrium points, a slight digression on the multidimensional Taylor series is needed. Consider the Taylor series of a function of n variables $f(x_1,\cdots,x_n)$ expanded about the point $\mathbf{y}$,

$$f(\mathbf{x}) \simeq f(\mathbf{y}) + \sum_{i=1}^{n}(x_i-y_i)\frac{\partial f}{\partial x_i} + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i-y_i)(x_j-y_j)\frac{\partial^2 f}{\partial x_i\,\partial x_j} + \cdots$$

where all partials are evaluated at $\mathbf{y}$. In matrix form, the second-order Taylor expansion of $f(\mathbf{x})$ about $\mathbf{x}_o$ is

$$f(\mathbf{x}) \simeq f(\mathbf{x}_o) + \nabla^T(\mathbf{x}-\mathbf{x}_o) + \frac{1}{2}(\mathbf{x}-\mathbf{x}_o)^T\mathbf{H}\,(\mathbf{x}-\mathbf{x}_o) \qquad \text{(A3.6)}$$

where ∇ and $\mathbf{H}$ are the gradient and hessian with respect to $\mathbf{x}$ evaluated at $\mathbf{x}_o$, e.g.,

$$\nabla \equiv \nabla_{\mathbf{x}}[\,f\,]\big|_{\mathbf{x}=\mathbf{x}_o} \qquad \text{and} \qquad \mathbf{H} \equiv \mathbf{H_x}[\,f\,]\big|_{\mathbf{x}=\mathbf{x}_o}$$

Example 6. Consider the following demonstration (due to Lande 1979) that mean population fitness increases. Expanding the log of mean fitness in a Taylor series around the current population mean $\boldsymbol{\mu}$ gives the change in mean population fitness as

$$\begin{aligned}\Delta \ln \overline{W}(\boldsymbol{\mu}) &= \ln \overline{W}(\boldsymbol{\mu} + \boldsymbol{\Delta\mu}) - \ln \overline{W}(\boldsymbol{\mu}) \\ &\simeq \left(\nabla_{\boldsymbol{\mu}}[\ln \overline{W}(\boldsymbol{\mu})\,] \right)^T \boldsymbol{\Delta\mu} + \frac{1}{2} \boldsymbol{\Delta\mu}^T\, \mathbf{H}_{\boldsymbol{\mu}}[\ln \overline{W}(\boldsymbol{\mu})\,]\, \boldsymbol{\Delta\mu}\end{aligned}$$

assuming that second and higher-order terms can be neglected (as would occur with weak selection and the population mean away from an equilibrium point), then

$$\Delta \ln \overline{W}(\boldsymbol{\mu}) \simeq \left(\nabla_{\boldsymbol{\mu}}[\ln \overline{W}(\boldsymbol{\mu})\,] \right)^T \boldsymbol{\Delta\mu}$$

Assuming that the joint distribution of phenotypes and additive genetic values is MVN, then $\boldsymbol{\Delta\mu} = \mathbf{G}\boldsymbol{\beta}$, or $\nabla_{\boldsymbol{\mu}}[\ln \overline{W}(\boldsymbol{\mu})\,] = \boldsymbol{\beta} = \mathbf{G}^{-1}\boldsymbol{\Delta\mu}$. Substituting gives

$$\Delta \ln \overline{W}(\boldsymbol{\mu}) \simeq \left(\mathbf{G}^{-1} \boldsymbol{\Delta\mu} \right)^T \boldsymbol{\Delta\mu} = (\boldsymbol{\Delta\mu})^T\, \mathbf{G}^{-1}\, \boldsymbol{\Delta\mu} \geq 0$$

since $\mathbf{G}$ is a variance-covariance matrix and hence is non-negative definite. Thus under these conditions, mean population fitness always increases, although since $\boldsymbol{\Delta\mu} \neq \nabla_{\boldsymbol{\mu}}[\ln \overline{W}(\boldsymbol{\mu})\,]$ fitness does not increase in the fastest possible manner.

At an equilibrium point $\widehat{\mathbf{x}}$, all first partials are zero, so that $(\nabla_{\mathbf{x}}[\,f\,])^T$ at this point is a vector of length zero. Whether this point is a maximum or minimum is then determined by the quadratic product involving the hessian evaluated at $\widehat{\mathbf{x}}$. Considering vector $\mathbf{d}$ of a small change from the equilibrium point,

$$f(\widehat{\mathbf{x}} + \mathbf{d}) - f(\widehat{\mathbf{x}}) \simeq \frac{1}{2} \cdot \mathbf{d}^T\, \mathbf{H}\, \mathbf{d} \tag{A3.7a}$$

Applying Equation 15.16, the canonical transformation of $\mathbf{H}$, simplifies the quadratic form to give

$$f(\widehat{\mathbf{x}} + \mathbf{d}) - f(\widehat{\mathbf{x}}) \simeq \frac{1}{2} \sum_{i=1}^{n} \lambda_i y_i^2 \tag{A3.7b}$$

where $y_i = \mathbf{e}_i{}^T \mathbf{d}$, $\mathbf{e}_i$ and λ_i being the eigenvectors and eigenvalues of the hessian evaluated at $\widehat{\mathbf{x}}$. Thus, if $\mathbf{H}$ is positive-definite (all eigenvalues of $\mathbf{H}$ are positive), f increases in all directions around $\widehat{\mathbf{x}}$ and hence $\widehat{\mathbf{x}}$ is a local minimum of f. If $\mathbf{H}$ is negative-definite (all eigenvalues are negative), f decreases in all directions around $\widehat{\mathbf{x}}$ and $\widehat{\mathbf{x}}$ is a local maximum of f. If the eigenvalues differ in sign ($\mathbf{H}$ is indefinite), $\widehat{\mathbf{x}}$ corresponds to a saddle point (to see this, suppose $\lambda_1 > 0$ and

$\lambda_2 < 0$; any change along the vector $\mathbf{e}_1$ results in an increases in f, while any change along $\mathbf{e}_2$ results in a decrease in f).

Example 7. Consider again the generalized Gaussian fitness function (Equation 15.33). From Equations 15.37b and 15.41a, we found that $\boldsymbol{\beta} = \nabla_{\boldsymbol{\mu}} \left[\ln \overline{W}(\boldsymbol{\mu}, \mathbf{P}) \right] = \mathbf{0}$ when $\widehat{\boldsymbol{\mu}} = \boldsymbol{\theta} + \mathbf{W}^{-1}\boldsymbol{\alpha}$. Is this a local maximum or a minimum? Recalling Equation 15.41a, we have

$$\begin{aligned}
\mathbf{H}_{\boldsymbol{\mu}} \left[\overline{W}(\boldsymbol{\mu}) \right] &= \nabla_{\boldsymbol{\mu}} \left[\left(\nabla_{\boldsymbol{\mu}} \left[\ln \overline{W}(\boldsymbol{\mu}) \right] \right)^T \right] \\
&= \nabla_{\boldsymbol{\mu}} \left[\left(\mathbf{P}^{-1} \left(\mathbf{P}^* \mathbf{P}^{-1} - \mathbf{I} \right) \boldsymbol{\mu} + \mathbf{P}^{-1} \mathbf{b} \right)^T \right] \\
&= \mathbf{P}^{-1} \left(\mathbf{P}^* \mathbf{P}^{-1} - \mathbf{I} \right)
\end{aligned}$$

From Equation 15.13a, $\mathbf{P}^* = \mathbf{P} - \mathbf{P} \left(\mathbf{P} + \mathbf{W}^{-1} \right)^{-1} \mathbf{P}$, giving

$$\begin{aligned}
\mathbf{H}_{\boldsymbol{\mu}} \left[\overline{W}(\boldsymbol{\mu}) \right] &= \mathbf{P}^{-1} \left[\left(\mathbf{I} - \mathbf{P} \left(\mathbf{P} + \mathbf{W}^{-1} \right)^{-1} \right) \mathbf{P} \mathbf{P}^{-1} - \mathbf{I} \right] \\
&= \mathbf{P}^{-1} \left[\mathbf{I} - \mathbf{I} - \mathbf{P} \left(\mathbf{P} + \mathbf{W}^{-1} \right)^{-1} \right] \\
&= - \left(\mathbf{P} + \mathbf{W}^{-1} \right)^{-1} \qquad \text{(A3.8)}
\end{aligned}$$

This result was obtained for the case of $\boldsymbol{\alpha} = \mathbf{0}$ by Lande (1979). Noting that if λ is an eigenvalue of $\mathbf{A}$, then $-\lambda^{-1}$ is an eigenvalue of $-\mathbf{A}^{-1}$, then if all eigenvalues of the matrix $\mathbf{J} = \left(\mathbf{P} + \mathbf{W}^{-1} \right)$ are positive ($\mathbf{J}$ is positive-definite), all eigenvalues of $\mathbf{H}_{\boldsymbol{\mu}} \left[\overline{W}(\boldsymbol{\mu}) \right]$ are negative and hence $\widehat{\boldsymbol{\mu}}$ corresponds to a local maximum in the mean population fitness surface. If $\mathbf{J}$ is positive-definite, them for all vectors $\mathbf{x}$,

$$\mathbf{x}^T \mathbf{J} \mathbf{x} = \mathbf{x}^T \left(\mathbf{P} + \mathbf{W}^{-1} \right) \mathbf{x} = \mathbf{x}^T \mathbf{P} \mathbf{x} + \mathbf{x}^T \mathbf{W}^{-1} \mathbf{x} > 0$$

Letting λ_i and $\mathbf{e}_i$ be the ith eigenvalue and associated unit eigenvector for $\mathbf{P}$ and likewise γ_i and $\mathbf{f}_i$ be the eigenvalue and associated unit eigenvector for $\mathbf{W}$, then applying the canonical transformation of each matrix (Equation 15.16),

$$\mathbf{x}^T \mathbf{J} \mathbf{x} = \sum_{i=1}^{n} y_i^2 \lambda_i + \sum_{i=1}^{n} z_i^2 \, \gamma_i^{-1}$$

where $y_i = \mathbf{e}_i{}^T \mathbf{x}$ and $z_i = \mathbf{f}_i^T \mathbf{x}$. Since all the eigenvalues of $\mathbf{P}$ are positive, $\mathbf{J}$ is positive-definite if all eigenvalues of $\mathbf{W}$ are positive (implying stabilizing selection on all characters). More generally, using the constraint that

$\mathbf{P}^* = (\mathbf{P}^{-1} + \mathbf{W})^{-1}$ must be positive definite, we can show that if γ_i corresponds to a negative eigenvalue of $\mathbf{W}$ (disruptive selection among the axis given by $\mathbf{f}_i$), then fitness is at a local minimum along this axis.

Optimization under constraints

Occasionally we wish to find the maximum or minimum of a function subject to a constaint. The solution is to use **Lagrange multipliers**: suppose we wish to find the extrema of $f(\mathbf{x})$ subject to the constraint $h(\mathbf{x}) = c$. Construct a new function g by considering

$$g(\mathbf{x}, \lambda) = f(\mathbf{x}) - \lambda(\, h(\mathbf{x}) - c\,)$$

Since $h(\mathbf{x}) - c = 0$, the extrema of $g(\mathbf{x}, \lambda)$ correspond to the extrema of $f(\mathbf{x})$ under the constraint. Local maxima and minima are obtained by solving the series of equations

$$\nabla_{\mathbf{x}}[\, g(\mathbf{x}, \lambda)\,] = \nabla_{\mathbf{x}}[\, f(\mathbf{x})\,] - \lambda \cdot \nabla_{\mathbf{x}}[\, g(\mathbf{x})\,] = \mathbf{0}$$
$$\nabla_{\mathbf{x}}[\, g(\mathbf{x}, \lambda)\,] = h(\mathbf{x}) - c = 0$$

Observe that the second equation is satisfied by the constraint.

Example 8. Consider a new (univariate) character which is a linear combination of n characters $(z_1, z_2, \cdots, z_n)$,

$$i = \mathbf{b}^T\,\mathbf{z} = \sum_{i=1}^{n} b_i\, z_i$$

where we further impose $\mathbf{b}^T\mathbf{b} = 1$. Denote the directional selection differential of the new character i by $s(i)$ and observe that if $\mathbf{s}$ is the vector of directional selection differentials for $\mathbf{z}$, then $s(i) = \mathbf{b}^T\mathbf{s}$. We wish to solve for $\mathbf{b}$ such that for a fixed amount of selection on i ($s(i) = r$) we maximize the response of another linear combination of $\mathbf{z}$, $\mathbf{A}^T\boldsymbol{\mu} = \sum a_i\,\mu_i$. Assuming the conditions leading to the multivariate breeders' equation hold, the function to maximize is

$$f(\mathbf{b}) = \mathbf{a}^T\boldsymbol{\Delta\mu} = \mathbf{a}^T\mathbf{G}\mathbf{P}^{-1}\mathbf{s}$$

under the associated constraint function

$$g(\mathbf{b}) - c = \mathbf{b}^T\mathbf{b} - 1 = 0$$

Since $s(i) = \mathbf{b}^T\mathbf{s}$ and we have the constraint $\mathbf{b}^T\mathbf{b} = 1$ take $\mathbf{s} = r \cdot \mathbf{b}$ so that $s(i) = \mathbf{b}^T\mathbf{s} = r \cdot \mathbf{b}^T\mathbf{b} = r$. Taking derivatives gives

$$\nabla_{\mathbf{x}}[\, g(\mathbf{x}, \lambda)\,] = r \cdot \nabla_{\mathbf{x}}[\,\mathbf{a}^T\mathbf{G}\mathbf{P}^{-1}\mathbf{b}\,] - \lambda \cdot \nabla_{\mathbf{x}}[\,\mathbf{b}^T\mathbf{b}\,] = r \cdot \left(\mathbf{a}^T\mathbf{G}\mathbf{P}^{-1}\right)^T - (2\lambda) \cdot \mathbf{b}$$

which is equal to zero when

$$\mathbf{b} = (2\lambda/r) \cdot \mathbf{P}^{-1}\mathbf{G}\mathbf{a}$$

where we can solve for λ by using the constraint $\mathbf{b}^T\mathbf{b} = 1$. Thus

$$i = c \cdot \left(\mathbf{P}^{-1}\mathbf{G}\mathbf{a}\right)^T \mathbf{z} \tag{A3.9}$$

where the constant c depends on the level of selection r set. This is Smith's **optimal selection index** (Smith 1936, Hazel 1943) who obtained it by a very different approach. Index selection is the subject of Chapter 37.

The example introduces the concept of a **selection index**, which will be more fully developed in Chapter 42.

Example 9. Suppose we wish to maximize the change in mean population fitness. Expanding mean population fitness in a Taylor series gives, to first order,

$$\begin{aligned}
\overline{W}(\boldsymbol{\mu} + \boldsymbol{\Delta\mu}) - \overline{W}(\boldsymbol{\mu}) &\simeq \nabla_{\boldsymbol{\mu}}[\,\overline{W}(\boldsymbol{\mu})\,]^T\,\boldsymbol{\Delta\mu} \\
&= \overline{W} \cdot \left(\overline{W}^{\,-1} \cdot \nabla_{\boldsymbol{\mu}}[\,\overline{W}(\boldsymbol{\mu})\,]^T\,\boldsymbol{\Delta\mu}\right) \\
&= \overline{W} \cdot \left(\nabla_{\boldsymbol{\mu}}[\,\ln\overline{W}(\boldsymbol{\mu})\,]^T\,\boldsymbol{\Delta\mu}\right) \\
&= \overline{W} \cdot \boldsymbol{\beta}^T\,\boldsymbol{\Delta\mu}
\end{aligned}$$

Thus, in taking $\boldsymbol{\beta} = \mathbf{A}$, the selection index that maximinzes the change in mean population fitness is given by

$$i = (\mathbf{P}^{-1}\,\mathbf{G}\boldsymbol{\beta})\mathbf{z}$$
