 DOI: 10.1086/655218

Appendix A from N. Anthes et al., "Bateman Gradients in Hermaphrodites: An Extended Approach to Quantify Sexual Selection"

(Am. Nat., vol. 176, no. 3, p. 249)

Principal Component Analysis on Male and Female Mating Success

This appendix outlines how Bateman gradients calculated after the replacement of male and female mating success (MS_m and MS_f) with their principal components relate to the original Bateman gradients and affect the decomposition of covariances. Throughout, principal component 1 (PC_1) is considered to represent the positive correlation component between MS_m and MS_f and thus overall mating activity. PC_2 captures the difference between MS_m and MS_f and thus the sexual bias in mating activity. (Note that we stick to these definitions for PC_1 and PC_2, not implying that PC_1 is necessarily the major axis; which of the two is the major axis depends on whether the overall correlation between MS_m and MS_f is positive or negative. Note further that statistical programs return arbitrary signs for PCs. It is thus necessary to check the factor loadings to determine whether PC_1 represents overall mating activity negatively or positively and whether PC_2 represents a male or female bias in mating activity.)

Standardized and Nonstandardized Principal Components

Standardized principal component analysis (PCA) is based on r, the Pearson correlation coefficient between MS_m and MS_f, and the variables standardized as $\tilde{X} = (X - \overline{X})/\sigma(X)$, with $\overline{X}$ the population average of X and $\sigma(X)$ its standard deviation. The two principal components (PC_1 and PC_2) are given as

$$PC_1 = \frac{1}{\sqrt{2}}(\tilde{MS}_m + \tilde{MS}_f),$$

$$PC_2 = \frac{1}{\sqrt{2}}(\tilde{MS}_m - \tilde{MS}_f), \quad (A1)$$

with eigenvalues $\lambda_1 = 1 + r$ and $\lambda_2 = 1 - r$, respectively. Here, MS_m and MS_f are expressed in units of their standard deviation, not in actual numbers of matings, which may not be intuitive for biological interpretation. It may thus often be preferable to express PC_1 and PC_2 in units of actual male and female matings. Exactly this is achieved by using nonstandardized PCA (based on the covariance rather than correlation matrix), even though this leads to more complicated expressions. Assuming positive covariance between MS_m and MS_f, we obtain

$$PC_1 = \frac{x}{\sqrt{1 + x^2}}MS_m + \frac{1}{\sqrt{1 + x^2}}MS_f, \quad (A2a)$$

where

$$x = \frac{1}{2}\frac{\mathrm{Var}(MS_m) - \mathrm{Var}(MS_f) + \sqrt{4\,\mathrm{Cov}(MS_m, MS_f)^2 + [\mathrm{Var}(MS_m) - \mathrm{Var}(MS_f)]^2}}{\mathrm{Cov}(MS_m, MS_f)},$$

and

$$PC_2 = \frac{1}{\sqrt{1 + x^2}}MS_m - \frac{x}{\sqrt{1 + x^2}}MS_f. \quad (A2b)$$

The Link between Original and PCA-Based Regression Coefficients

Using principal components of MS_m and MS_f rather than their original measurements in our analysis of Bateman gradients does not represent an independent type of analysis but primarily shifts our perspective on the connection between MS and RS. In fact, the regressions on the original variables (MS_m and MS_f) are mathematically equivalent to those on PC_1 and PC_2, with (by construction) the same amount of explained and unexplained variance in RS_m and RS_f. The following formulas show how the original Bateman gradients (eqq. [3]) formally link to the regression coefficients obtained after calculating standardized principal components of MS_m and MS_f:

$$
\begin{aligned}
\beta_{mm} &= \frac{1}{\sqrt{2}\sigma(MS_m)}(\beta_{m_{PC_1}} + \beta_{m_{PC_2}}),\\
\beta_{mf} &= \frac{1}{\sqrt{2}\sigma(MS_f)}(\beta_{m_{PC_1}} - \beta_{m_{PC_2}}),\\
\beta_{fm} &= \frac{1}{\sqrt{2}\sigma(MS_m)}(\beta_{f_{PC_1}} + \beta_{f_{PC_2}}),\\
\beta_{ff} &= \frac{1}{\sqrt{2}\sigma(MS_f)}(\beta_{f_{PC_1}} - \beta_{f_{PC_2}}).
\end{aligned} \tag{A3}
$$

Similar expressions apply when one has chosen to calculate nonstandardized principal components of MS_m and MS_f:

$$
\begin{aligned}
\beta_{mm} &= \frac{1}{\sqrt{1+x^2}}(x\beta_{m_{PC_1}} + \beta_{m_{PC_2}}),\\
\beta_{mf} &= \frac{1}{\sqrt{1+x^2}}(\beta_{m_{PC_1}} - x\beta_{m_{PC_2}}),\\
\beta_{fm} &= \frac{1}{\sqrt{1+x^2}}(x\beta_{f_{PC_1}} + \beta_{f_{PC_2}}),\\
\beta_{ff} &= \frac{1}{\sqrt{1+x^2}}(\beta_{f_{PC_1}} - x\beta_{f_{PC_2}}).
\end{aligned} \tag{A4}
$$

Rearrangements in the Covariance Decomposition

Using PCA-based coefficients also affects the decomposition of covariance (eq. [5]). Given the lack of covariance between PC_1 and PC_2, the equation simplifies to

$$
\begin{aligned}
&\mathrm{Cov}(RS_m, RS_f) = \\
&\beta_{m_{PC_1}}\beta_{f_{PC_1}}\mathrm{Var}(PC_1) + \beta_{m_{PC_2}}\beta_{f_{PC_2}}\mathrm{Var}(PC_2) + \mathrm{Cov}(\varepsilon_f, \varepsilon_m).
\end{aligned} \tag{A5}
$$

In the case of a standardized PCA, the two variance terms in this expression are $\mathrm{Var}(PC_1) = 1 + r$ and $\mathrm{Var}(PC_2) = 1 - r$. (Following analysis using nonstandardized PCA, eq. [A5] remains valid and numerically identical.) The first term of equation (A5) is the covariance due to the common dependence of RS_m and RS_f on overall mating activity. The second term is the covariance due to variation in the sexual bias in mating activity. The last term (residual covariance) is identical to that in equation (5). Although the first two terms in equation (A5) are just a rearrangement of the first three terms in equation (5), they do facilitate the understanding of correlated selection on MS_m and MS_f by separately focusing on total mating activity and sexual bias.