

**Michael W. Palmer, Daniel J. McGlinn, Lars Westerberg, and Per Milberg. 2008. Indices for detecting differences in species composition: some simplifications of RDA and CCA. *Ecology* 89:1769–1771.**

## **Appendix A: Simplification of RDA eigenvalue.**

**Notation (modified after Legendre and Legendre 1998):**

For convenience, we use the terminology of fixed plots sampled twice, although the results pertain to split-plot designs and to balanced independent samples.

Samples are indexed by the letter  $i$ : 1, 2, ... $n$

Pairs of samples or ‘plots’ are indexed by the letter  $k$ : 1, 2, ... $n/2$

Species are indexed by the letter  $j$ : 1, 2, ... $r$

Elements  $y_{ij}$  of the response matrix  $\mathbf{Y}$  ( $n \times r$ ) denote the abundance in sample  $i$  (rows) of species  $j$  (columns).

The explanatory variable  $\mathbf{x}$  is a column vector of length  $n$  and is a dummy variable where  $x_i$  equals 1 or 0 indicating two different times.

The response matrix  $\mathbf{Y}$  is divided into two  $\frac{n}{2} \times r$  submatrices  $\mathbf{G}$  and  $\mathbf{H}$ , where  $\mathbf{G}$  contains the samples corresponding to  $x_i = 0$  or ‘time 1’ and where  $\mathbf{H}$  contains the samples corresponding to  $x_i = 1$  or ‘time 2’ The pairs of samples, ‘plots’, in  $\mathbf{G}$  and  $\mathbf{H}$  are indexed by the letter  $k$ : 1, 2, ...  $\frac{n}{2}$ .

The averages of the variables will be denoted by a bar over the variable, e.g.  $\bar{y}_j = \frac{\sum_i y_{ij}}{n}$

**S** stands for a dispersion matrix, e.g. the matrix  $\mathbf{S}_{\mathbf{Y}\mathbf{x}}$  denotes the covariance between the matrix **Y** and the vector **x** (Legendre and Legendre 1998, p135).

---

**Proof:**

The eigenanalysis equation for RDA (Legendre and Legendre 1998, p. 579) is:

$$(\mathbf{S}_{\mathbf{Y}\mathbf{x}}\mathbf{S}_{\mathbf{x}\mathbf{x}}^{-1}\mathbf{S}_{\mathbf{x}\mathbf{Y}} - \lambda_a \mathbf{I})\mathbf{u}_a = 0, \text{ for } a \text{ canonical eigenvalues and eigenvectors.} \quad (\text{A.1})$$

Because **x** is a vector in our case, the inverse covariance matrix  $\mathbf{S}_{\mathbf{x}\mathbf{x}}^{-1}$  is just a scalar equal to the reciprocal of the variance of **x**:

$$\mathbf{S}_{\mathbf{x}\mathbf{x}}^{-1} = \frac{1}{\sigma_{\mathbf{x}}^2} = \frac{n-1}{\sum_i (x_i - \bar{x})^2} = \frac{4(n-1)}{n} \quad (\text{A.2})$$

Furthermore, in our case,  $\mathbf{S}_{\mathbf{Y}\mathbf{x}}$  and  $\mathbf{S}_{\mathbf{x}\mathbf{Y}}$  are column and row vectors respectively. The elements of  $\mathbf{S}_{\mathbf{Y}\mathbf{x}}$  are equal to the elements of  $\mathbf{S}_{\mathbf{x}\mathbf{Y}}$ , i.e.:

$$(\mathbf{S}_{\mathbf{Y}\mathbf{x}})' = \mathbf{S}_{\mathbf{x}\mathbf{Y}} = \frac{1}{n-1} \sum_i [(y_{ij} - \bar{y}_j)(x_i - \bar{x})], \text{ for species } j = 1, 2, \dots, r \quad (\text{A.3})$$

Equation A.3 will simplify further after partitioning the response matrix **Y** into two submatrices **G** and **H**. As noted above we are assuming a balanced design, thus **G** and **H** are both  $\frac{n}{2} \times r$  matrices, with ‘plots’ indexed by *k*. Because  $\bar{x} = 0.5$  the expression A.3 will become:

$$\mathbf{S}_{\mathbf{Y}'\mathbf{x}} = \frac{1}{n-1} \left( \sum_k [(g_{kj} - \bar{y}_j)(-0.5)] + \sum_k [(h_{kj} - \bar{y}_j)(0.5)] \right) \quad (\text{A.4})$$

$$\mathbf{S}_{\mathbf{Y}'\mathbf{x}} = \frac{1}{2(n-1)} \left( \sum_k [(h_{kj} - \bar{y}_j)] - \sum_k [(g_{kj} - \bar{y}_j)] \right) \quad (\text{A.5})$$

$$\mathbf{S}_{\mathbf{Y}'\mathbf{x}} = \frac{1}{2(n-1)} \sum_k (h_{kj} - g_{kj}) \quad (\text{A.6})$$

$$\mathbf{S}_{\mathbf{Y}'\mathbf{x}} = \frac{\sum_k d_{kj}}{2(n-1)} \quad (\text{A.7})$$

where  $d$  represents the difference in species  $j$ 's abundance between the two samples. Note that  $d$  is not necessarily calculated as a pair-wise difference.

Now we digress to prove that in this case, the eigenvalue for the matrix  $\mathbf{S}_{\mathbf{Y}\mathbf{x}}\mathbf{S}_{\mathbf{x}\mathbf{x}}^{-1}\mathbf{S}_{\mathbf{x}\mathbf{Y}}$  is equal to the eigenvalue of the matrix  $\mathbf{S}_{\mathbf{x}\mathbf{Y}}\mathbf{S}_{\mathbf{Y}\mathbf{x}}\mathbf{S}_{\mathbf{x}\mathbf{x}}^{-1}$ . The trace of the  $(r \times r)$  matrix  $\mathbf{S}_{\mathbf{Y}\mathbf{x}}\mathbf{S}_{\mathbf{x}\mathbf{x}}^{-1}\mathbf{S}_{\mathbf{x}\mathbf{Y}}$ , from the RDA eigenanalysis equation (A.1), is equivalent to the sum of all the canonical eigenvalues (McArdle and Anderson 2001), and in our context there is only one canonical eigenvalue because we only have one explanatory variable  $\mathbf{x}$ . Furthermore, for any two hypothetical matrices, e.g.  $\mathbf{A}$  ( $n \times r$ ) and  $\mathbf{B}$  ( $r \times n$ ) the

$\text{trace}(\mathbf{AB}) = \text{trace}(\mathbf{BA})$ , therefore, the trace of  $\mathbf{S}_{\mathbf{Y}\mathbf{x}}\mathbf{S}_{\mathbf{x}\mathbf{x}}^{-1}\mathbf{S}_{\mathbf{x}\mathbf{Y}}$  is equal to the trace of

$\mathbf{S}_{\mathbf{x}\mathbf{Y}}\mathbf{S}_{\mathbf{Y}\mathbf{x}}\mathbf{S}_{\mathbf{x}\mathbf{x}}^{-1}$  which is actually just a scalar. Thus we rewrite A.1 as

$(\mathbf{S}_{\mathbf{x}\mathbf{Y}}\mathbf{S}_{\mathbf{Y}\mathbf{x}}\mathbf{S}_{\mathbf{x}\mathbf{x}}^{-1} - \lambda_a \mathbf{I})\mathbf{u}_a = 0$ . Further simplification is possible because  $\mathbf{I}$ ,  $\mathbf{u}$  and  $a$  are unity:

$$\lambda = \mathbf{S}_{\mathbf{x}\mathbf{Y}}\mathbf{S}_{\mathbf{Y}\mathbf{x}}\mathbf{S}_{\mathbf{x}\mathbf{x}}^{-1} = \frac{\sum_j \left( \sum_k d_{kj} \right)^2}{4(n-1)^2} \cdot \frac{4(n-1)}{n} \quad (\text{A.8})$$

$$\lambda = \frac{\sum_j \left( \sum_k d_{kj} \right)^2}{n(n-1)} \quad (\text{A.9})$$

#### LITERATURE CITED

Legendre, P., and L. Legendre. 1998. Numerical Ecology. Second English edition.

Elsevier, Amsterdam, The Netherlands.