Appendix S2

COMPUTING THE TOTAL SUM OF SQUARES FROM A DISSIMILARITY MATRIX

Equation 8 in the main paper shows how to compute the total sum of squares, SS_{Total}, for a matrix of Euclidean distances. The equivalence of eq. 2 (computed on raw data) and eq. 8 (computed on distances) was demonstrated in Appendix 1 of Legendre & Fortin (2010). The profile, chord, Hellinger, and chi-square distances are obtained by first transforming the raw abundance data as described in Appendix S1, then computing the Euclidean distance formula on the transformed data. As a consequence, for these distances computed using the Euclidean distance formula, it is clear that SS_{Total} can be computed either from the transformed data through eq. 2 or from the distances through eq. 8.

For the other distances that have the Euclidean property as either **D** or $\mathbf{D}^{(0.5)} = [D_{hi}^{0.5}]$ (Table 2, column P13, codes 1 or 2), eq. 8 also applies. That point is demonstrated as follows: any dissimilarity matrix that has the Euclidean property can be decomposed into principal coordinates by principal coordinate analysis (PCoA, Gower 1966), obtaining a fully Euclidean representation of the data (Gower & Legendre 1986). Calculation of dissimilarity matrix **D** followed by PCoA of **D** or $\mathbf{D}^{(0.5)}$ acts as a data transformation. The total sum of squares of the matrix of principal coordinates, computed through eq. 2, is equal to the total sum of squares computed from dissimilarity matrix **D** or $\mathbf{D}^{(0.5)}$ through eq. 8.

Finally, for dissimilarities that do not lead to a fully Euclidean representation (i.e. those that do not have the Euclidean property; Table 2, column P13, code 0), the same equivalence still exists although these matrices produce negative eigenvalues in PCoA. The demonstration involves two steps. On the one hand, the trace of the Gower-centred matrix on which eigen-decomposition is computed, which is equal to the sum of all eigenvalues (positive and negative), is equal to SS_{Total} computed from the dissimilarity matrix through eq. 8. On the other hand, McArdle & Anderson (2001) and Anderson (2006) have shown how to compute SS_{Total} of the principal coordinate representation using the real and complex principal coordinates, and this is equal again to the trace of the Gower-centred matrix on which eigen-decomposition is computed in PCoA. Their method has three steps: (1) square all values in the matrix of eigenvectors (which were produced by eigen-decomposition with a norm of 1), (2) multiply each squared eigenvector by its eigenvalue, and (3) sum all the resulting values. The calculation is demonstrated in the R function pcoa.short() in R, below.

Illustration: calculation of SS_{Total} for a non-Euclidean dissimilarity matrix

For the example, we will use the first 10 rows of the mite data available in package **vegan** and compute the percentage difference dissimilarity. The calculations are done in the R language.

```
require(vegan)
data(mite)

### Compute the percentage difference dissimilarity for the mite data (first 10 rows)
mite.D <- vegdist(mite[1:10,], "bray")
# Is the dissimilarity matrix Euclidean?
require(ade4)
is.euclid(mite.D)                               # Available in R package ade4
# [1] FALSE

### Compute SSTotal from the dissimilarities

SS.D <- function(D, n) sum(D^2) / (n)           # Equation 8
res.SS.D <- SS.D(mite.D, 10)
res.SS.D
# [1] 0.9626073                                 # Result: SSTotal

### A short function for principal coordinate analysis (PCoA)

##############################################################################
pcoa.short <- function(D, include.zero=FALSE, only.values=FALSE)
#
# Compute PCoA for a Euclidean or non-Euclidean dissimilarity matrix.
# The eigenvectors are not scaled to sqrt(eigenvalues),
# hence they are not principal coordinates in the PCoA sense.
#
# When 'n' is very large, users may choose not to compute the eigenvectors.
# This is obtained by selecting the option only.values=TRUE.
# The statistic VarTotal=SS will not be computed and printed in that case.
#
# License: GPL-2
# Author:: Pierre Legendre
{
        D <- as.matrix(D)
        n <- nrow(D)
        epsilon <- sqrt(.Machine$double.eps)
#
# Gower centring, matrix formula
        One <- matrix(1,n,n)
        mat <- diag(n) - One/n
        G <- -0.5 * mat %*% (D^2) %*% mat
        trace <- sum(diag(G))
        SSi <- diag(G)
#
# Eigenvalue decomposition
        eig <- eigen(G, symmetric=TRUE, only.values=only.values)
# Exclude the null eigenvalue/s if include.zero is FALSE
        select <- 1:n
        exclude <- which(abs(eig$values) < epsilon)
        cat("Note - Eigenvalue/s", exclude, "is/are null\n")
        if(!include.zero) {
                cat("Note - Eigenvalue/s and eigenvector/s", exclude, "was/were excluded\n")
                select <- select[-exclude]
                }
        values <- eig$values[select]
#
if(!only.values) {                              # Compute SS from the eigenvectors
        vectors <- eig$vectors[,select]
        vectors.sq <- vectors^2 %*% diag(values)
        SS <- sum(vectors.sq)
        } else {
        vectors <- NA
        SS <- NA
        }
#
list(values=values, vectors=vectors, trace=trace, SS.total=SS, SSi=SSi, site.names=rownames(D),
select=select)
}
################################################################################
```

```
res <- pcoa.short(mite.D)
# Note - Eigenvalue/s 9 is/are null
# Note - Eigenvalue/s and eigenvector/s 9 was/were excluded

### Compute SS_Total as the trace of the Gower-centred matrix

res$trace
# [1] 0.9626073                                    # Result: SS_Total

### Compute SS_Total as the sum of the PCoA eigenvalues

sum(res$values)
# [1] 0.9626073                                    # Result: SS_Total

### Compute SS_Total as the sum of squares of the principal coordinates scaled to lengths equal to
the square roots of the eigenvalues, including the one with a negative eigenvalue. The result is the
same with options include.zero=FALSE or include.zero=TRUE.

res$SS.total
# [1] 0.9626073                                    # Result: SS_Total
```

REFERENCES

Anderson, M.J. (2006). Distance-based tests for homogeneity of multivariate dispersions. *Biometrics*, 62, 245–253.

Gower, J.C. & Legendre, P. (1986). Metric and Euclidean properties of dissimilarity coefficients. *J. Classif.*, 3, 5–48.

Legendre, P. & Fortin, M.-J. (2010). Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. *Mol. Ecol. Resour.*, 10, 831–844.

McArdle, B.H. & Anderson, M.J. (2001). Fitting multivariate models to community data: a comment on distance-based redundancy analysis. *Ecology*, 82, 290–297.