4.11 Visualising models: dbRDA

We may wish to visualise a given model in the multivariate space of our chosen resemblance matrix. The ordination method of PCO was introduced in chapter 3 as a way of obtaining orthogonal (independent) axes in Euclidean space that would represent our data cloud, as defined by the resemblance measure chosen, in order to visualise overall patterns. PCO is achieved by doing an eigenvalue decomposition of Gower's centred matrix **G** (Fig. 3.1). Such an ordination is considered to be *unconstrained*, because we have not used any other variables or hypotheses to draw it. Suppose now that we have a model of the relationship between our data cloud and one or more predictor variables that are contained in matrix **X**. We have constructed the projection matrix **H**, and produced the sums of squares and cross-products matrix of fitted values **HGH** in order to do the partitioning (Fig. 4.2). Distance-based redundancy analysis (dbRDA) is simply an ordination of the *fitted values* from a multivariate regression model; it is achieved by doing an eigenvalue decomposition of matrix **HGH** (Fig. 4.15). That is:

1. Eigenvalue decomposition of matrix **HGH** yields eigenvalues (γ^2_l , $l = 1$, ..., s) and their associated eigenvectors.

2. The dbRDA axes Z (also called dbRDA "scores") are obtained by scaling (multiplying) each of the eigenvectors by the square root of their corresponding eigenvalue.
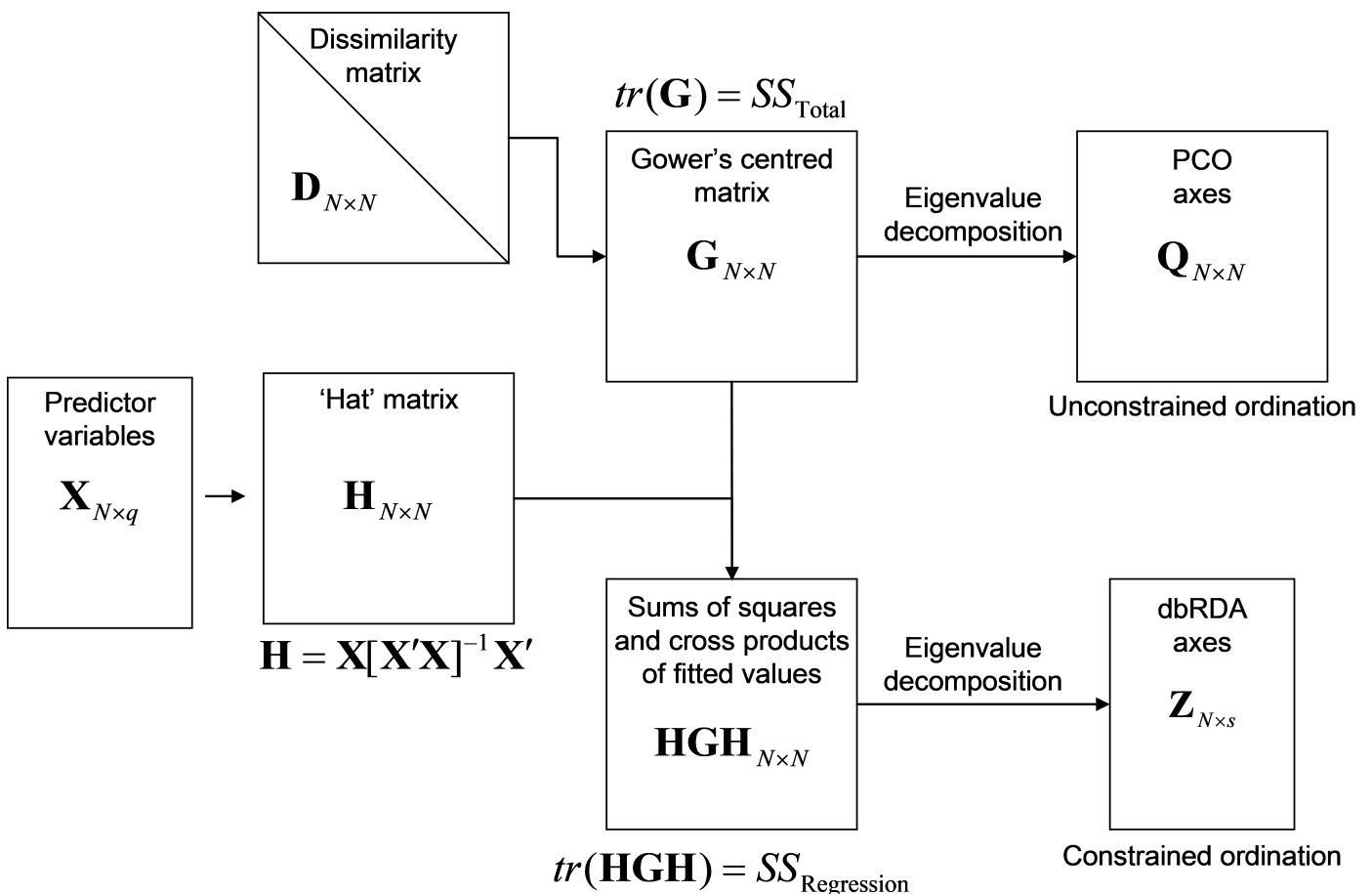

Fig. 4.15. Schematic diagram of distance-based redundancy analysis as a constrained ordination:

an eigenvalue decomposition of the sums of squares and cross products of the fitted values from a model.

The result is a *constrained* ordination, because it will produce axes that must necessarily be directly and linearly related to the fitted values and, therefore, the predictor variables. If Euclidean distances are used at the outset to produce matrix **D**, then **G** = **YY**′ (where **Y** has been centered) and dbRDA is equivalent to doing a traditional RDA directly on a matrix of *p* response variables **Y**. The number of dbRDA axes that will be produced (denoted by *s* above and in Fig. 4.15) will be the minimum of (*q*, (*N* – 1)). If the resemblance measure used is Euclidean, then *s* will be the minimum of (*p*, *q*, (*N* – 1)). As with most ordination methods, we can draw the first two or three dbRDA axes and examine patterns among the samples seen on the plot.

A dbRDA plot can be obtained within PRIMER in one of two ways using the new PERMANOVA+ add-on: either directly for a given resemblance matrix by choosing **PERMANOVA+ > dbRDA** and then providing the name of a worksheet containing the predictor variables to be used (all of the variables in this worksheet that are selected will be used in this case), or by ticking the option ($\checkmark$Do dbRDA plot) in the DISTLM dialog. For the latter, the dbRDA plot will be done on the predictor variables selected by the DISTLM routine, as obtained using the selection procedure and selection criterion chosen by the user.

Initial interest lies in determining the *adequacy* of the plot. More specifically, how much of the *fitted* model variation is captured by the first two (or three) dbRDA axes? As for other eigenvalue decomposition techniques (such as PCA or PCO), the eigenvalues[88] are ordered from largest to smallest and yield information regarding the variation explained by each orthogonal (independent) dbRDA axis. The sum of the eigenvalues from dbRDA is equal to the total explained variation in the fitted model, that is: $\sum \gamma^2 _l = tr (\text{\bf HGH}) = SS _ \text{Regression}$. So, the percentage of the fitted model's variation that is explained by the *l*th dbRDA axis is (100 × $\gamma^2 _l / \sum \gamma^2 _l$). As with PCO or PCA, we consider that if the percentage of the fitted variation that is explained by the diagram exceeds ~ 70%, then the plot is likely to capture most of the salient patterns in the fitted model.

Another relevant conceptual point regarding dbRDA is that there is an important difference between the percentage of the *fitted* variation explained by a given axis, as opposed to the percentage of the *total* variation explained by that axis. The latter is calculated as (100 × $\gamma^2 _l / tr (\text{\bf G}) $ or (100 × $\gamma^2 _l / \sum \lambda _i $), where $\sum \lambda _i $ is the sum of PCO eigenvalues. The dbRDA axis might be great at showing variation according to the fitted model, but if the model itself only explains a paltry amount of the total variation in the first place, then the dbRDA axis may be of little overall relevance in the multivariate system as a whole. It is therefore quite important to consider the percentage of explained variation for each dbRDA axis out of the fitted *and* out of the total variation. Another way of thinking about this is to compare the patterns seen in the dbRDA plot with the patterns seen in the unconstrained PCO plot. If the patterns are similar, then this indicates that the model is a pretty good one and captures much of what can be seen overall in the multivariate data cloud. Such cases should also generally correspond to situations where the *R*^{2} for the model ($ SS _ \text{ Regression } / SS_ \text{ Total } = \sum \gamma^2 _l / \sum \lambda _i $) is fairly large. In contrast, if the pattern in the dbRDA and the PCO are very different, then there are probably other factors that are not included in the model which are important in driving overall patterns of

variation.

---

[88] The eigenvalues from a redundancy analysis are sometimes called "canonical eigenvalues", as in chapter 11 in Legendre & Legendre (1998) .

Revision #8
Created 13 August 2022 08:27:36 by Arden
Updated 5 September 2022 13:24:37 by Arden