Multidimensional scaling and regression.

Cajo J. F. Ter Braak
Agricultural Mathematics Group-DLO and Research Institute for Nature Management

Box 100, 6700 AC Wageningen, the Netherlands

Citation: ter Braak, C. J. F. (1992). Multidimensional scaling and regression. *Statistica Applicata (Italian Journal of Applied Statistics)*, **4,** 577-586.

Constrained multidimensional scaling was put on a firm theoretical basis by Jan De Leeuw and Willem Heiser in the 1980's. There is a simple method of fitting, based on distance via innerproducts, and a numerically more complicated one that is truly based on least-squares on distances. The unconstrained forms are known as principal coordinate analysis and nonmetric multidimensional scaling, respectively. Constraining the solution by external variables brings the power of classical regression analysis back into multidimensional data analysis. This idea is developed and illustrated, with emphasis on constrained principal coordinate analysis.

Keywords: constrained ordination, external linear constraints, principal coordinate analysis, redundancy analysis

1 Introduction

SMACOF is a Dutch invention. It is a model for constructing convergent algorithms to solve the multidimensional scaling problem (J. De Leeuw *et al.* 1977, 1980; J. De Leeuw, 1988). SMACOF is an acronym for 'Scaling by Majorizing a COmplicated Function' (W. Heiser, 1987; J. Meulman, 1986). SMACOF is also the name of a series of computer programs that uses this type of algorithm. They attempt to give least-squares solutions by minimizing STRESS (J. Kruskal, 1964), but may get trapped in a local minimum. SMACOF avoids the problems of differentiability and step-size in the usual gradient methods. In my opinion, the prime importance of SMACOF is that it can deal with many types of constraints on the multidimensional scaling solution.

With my background in agronomy and statistical ecology, I am familiar with the power of the analysis of variance and of regression methodology. Faced with a multivariate research problem, I therefore tend to cast it into the framework of regression analysis. What is to be explained by what? Which are response variables (criterion variables) and which are explanatory variables (regressors)? Regression methodology can be integrated in multidimensional scaling by imposing the constraint that the coordinates of the solution are a (preferably) linear combination of external variables. The external variables thus take the role of regressors. The importance and/or statistical significance of particular regressors can be judged by comparing the fit with and without these regressors. Of course, not all problems can be formulated in this way but in applied work many can.

I focus in this paper on constraining principal coordinate analysis by a regression model, because this serves well to illustrate the general idea. The theory is not new - J. De Leeuw *et al.* (1982: pag. 304) provide a brief description, acknowledging a conference contribution by J. Carroll *et al.* (1976). P. Digby *et al.* (1981) considered a special case in which the analysis is constrained by a one-way classification (*i.e.* by a one-way analysis of variance rather than by a regression model). Two examples illustrate the theory. The final section reviews Dutch contributions to constrained multidimensional scaling.

2 Principal coordinate analysis constrained by regression models

Principal coordinate analysis (PCO; J. Gower, 1966), alias classical scaling (W. Torgerson, 1958) is a simple form of metric multidimensional scaling. PCO is based on principal component analysis (PCA), but is more general in that it can deal with other measures of dissimilarity than just the Pythagorean dissimilarity implicit in PCA. The constrained form of PCO is first derived in an informal way and thereafter a more formal definition is given.

The basic assumption of PCO is that there exists, in principle, an exact representation of the dissimilarities $\mathbf{\Delta} = \{\delta_{ij}\}$ [i,j =1, 2, ..., n] in a high-dimensional Euclidean space, *i.e.* there exist a dimension q and an $n \times q$ matrix $\mathbf{Y}$ such that

$\delta_{ij}^2 = (\mathbf{y}_i - \mathbf{y}_j)'(\mathbf{y}_i - \mathbf{y}_j)$, or shortly, $\boldsymbol{\Delta}^2 = D^2(\mathbf{Y})$, (1)

where $\mathbf{y}_i'$ denotes the *i*-th row of **Y** (J. Gower *et al.*, 1986). Such a matrix **Y** can be obtained with PCO. If the primary data were not $\boldsymbol{\Delta}$ but **Y**, PCO is equivalent with a PCA of **Y**. A constrained form of PCO can thus be obtained from the constrained form of PCA, *i.e.* by a reduced-rank regression/redundancy analysis (RDA) (*e.g.* R. Jongman *et al.*, 1987) of **Y** with respect to the $n \times p$ matrix **X**. In regression terminology, the columns of **X** and **Y** contain *p* regressors and *q* response variables, respectively. To obtain a constrained PCO, the full *q*-dimensional PCO-solution should therefore be subjected to redundancy analysis. Constrained PCO so defined requires two eigen analyses, a PCO of dissimilarities $\boldsymbol{\Delta}$, resulting in a high-dimensional coordinate matrix **Y,** followed by a RDA of **Y** with respect to the regressor matrix **X**. An interesting question is whether the same result can be obtained from a single eigen analysis. To show that it can, note that the final redundancy analysis is equivalent to a PCA of the fitted values, $= \mathbf{X}^{\circ}\mathbf{Y}$ with $\mathbf{X}^{\circ} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ (P. Davies *et al.*, 1982). The solution can thus be obtained from the eigen analysis of $\mathbf{X}^{\circ}\mathbf{Y}\mathbf{Y}'\mathbf{X}^{\circ}$. The desired single eigen analysis now follows from the basic relation between squared distance and innerproduct that underlies PCO,

$\mathbf{Y}\mathbf{Y}' = -\frac{1}{2}\,\mathbf{J}\,\boldsymbol{\Delta}^2\,\mathbf{J}$, (2)

where **J** is the matrix which centres each *n*-vector, *i.e.* $\mathbf{J} = \mathbf{I} - n^{-1}\,\mathbf{1}_n\mathbf{1}_n'$. This relation, which originated in the 1930's (J. De Leeuw *et al.*, 1982), can be verified by expanding the inner product in (1) and taking the centroid of the points as the origin, *e.g.* $\mathbf{1}_n'\mathbf{Y} = \mathbf{0}_q$. If the regressors are centred, *i.e.* $\mathbf{1}_n'\mathbf{X} = \mathbf{0}_p$, then $\mathbf{J}\mathbf{X}^{\circ}\mathbf{J} = \mathbf{X}^{\circ}$, so that constrained PCO reduces to the eigen analysis

$-\frac{1}{2}\,\mathbf{X}^{\circ}\,\boldsymbol{\Delta}^2\,\mathbf{X}^{\circ}\,\mathbf{P} = \mathbf{P}\,\boldsymbol{\Lambda}$, (3)

where **P** contains the eigenvectors and $\boldsymbol{\Lambda}$ the eigenvalues. For an *r*-dimensional ordination, the required coordinate matrix **F** is equal to the first *r* columns of $\mathbf{P}\,\boldsymbol{\Lambda}^{\frac{1}{2}}$ (typically r=2). Finally, **C** is obtained by regressing **F** on **X**, *i.e.* $\mathbf{C} = (\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{X}'\mathbf{F}$.

More formally, PCO can be defined as finding an $n \times r$ matrix **F** that minimizes (J. Meulman, 1986)

$L(\mathbf{F}) = \| -\frac{1}{2}\,\mathbf{J}\{\boldsymbol{\Delta}^2 - D^2(\mathbf{F})\}\,\mathbf{J} \|^2$, (4)

where $\|\mathbf{A}\|^2 = \text{trace}(\mathbf{A}'\mathbf{A})$. Constrained PCO is now defined as the technique that minimizes L(**F**), subject to the constraint $\mathbf{F} = \mathbf{X}\mathbf{C}$. In the Appendix it is shown that (3) gives the solution to this constrained minimization problem and that the solution can also be obtained from an eigen analysis of order *p*, rather than *n*.

A series of regressions allows the construction of an accumulated analysis of variance. In the context of multidimensional scaling, we would rather have an analysis of distance (P. Digby *et al.*, 1981). For constrained PCO, this analysis can be defined analogously to the analysis of variance, because

$\frac{1}{2}\,\Sigma_i\Sigma_j\,\delta_{ij}^2 = \frac{1}{2}\,\mathbf{1}_n'\,\boldsymbol{\Delta}^2\,\mathbf{1}_n = n\,\text{trace}(\mathbf{Y}'\mathbf{Y}) = n\,\Sigma_k\,\{\Sigma_i\,y_{ik}^2\}$ (5)

and a similar relation holds after projection, both in *q*-space and in *r*-space. B. McArdle (1990) noted that the trace of the regression sum of squares matrix can be written in the

form of the Mantel statistic (N. Mantel, 1967; P. Legendre *et al.*, 1989):

$$\| \hat{\mathbf{Y}} \|^2 = \mathrm{tr}(\mathbf{Y}'\mathbf{X}^{\circ}\mathbf{Y}) = \mathrm{tr}(\mathbf{Y}\mathbf{Y}'\mathbf{X}^{\circ}) = \Sigma_{ij} (\mathbf{Y}\mathbf{Y}')_{ij}(\mathbf{X}^{\circ})_{ij} = -\tfrac{1}{2}\Sigma_{ij} \delta_{ij}^2(\mathbf{X}^{\circ})_{ij}. \quad (6)$$

The 'natural' Mantel statistic, wholly based on distances, is

$$\Sigma_i\Sigma_j \delta_{ij}^2 D^2(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-\frac{1}{2}})_{ij} = 4\| \hat{\mathbf{Y}} \|^2 + 2p\mathbf{1}_n\boldsymbol{\alpha} + n\,\mathrm{tr}(\boldsymbol{\alpha}\mathbf{h}'), \quad (7)$$

where $\boldsymbol{\alpha}$ and $\mathbf{h}$ are the vectors taken from the diagonal of $\mathbf{Y}\mathbf{Y}'$ and $\mathbf{X}^{\circ}$, respectively. For many balanced experimental designs, test statistics (6) and (7) are thus equivalent.

Example 1: genetic distances explained by environmental variables

This example concerns the genetic variability in 21 colonies of butterfly *Euphydryas editha* in California and Oregon in relation to 11 quantitative environmental variables. B. Manly (1985: pag. 168-184) applied multiple regression and the Mantel test to these data. C. Ter Braak (1987) analyzed the gene frequencies on three loci by canonical (multiple) correspondence analysis. The implied dissimilarity coefficient herein is, however, not the one preferred by most geneticists. Here this defect is circumvented by applying constrained PCO on Nei's standardized genetic distances (B. Manly 1985: $\boldsymbol{\Delta}^2$ taken from Table 6.8) using the environmental variables as regressors. Summary statistics of both the constrained and unconstrained analysis are given in Tab. I. There are some small negative eigenvalues in both analyses pointing to the fact that the squareroot of Nei's genetic distance is not perfectly embeddable in Euclidean space (J. Gower *et al.*, 1986; perhaps I should have analyzed squared Nei's distances instead). The environmental data explain ca. 88% (155/1.77) of the mean pairwise genetic distance (cf. (5)). The gain of constrained analysis over unconstrained analysis is, however, small, because the first two dimensions of the unconstrained analysis are very well explained by the environmental variables (R^2 = .92 and .89, respectively, compared to R^2 = 1.0, by definition, in the constrained analysis). Fig. 1 shows the constrained configuration which displays 71% ((126+25)/1.77) of the mean pairwise genetic distance and 97% ((126+25)/1.55) of the mean pairwise fitted distances. The weights $\mathbf{C}$ of the linear combination are not interpretable for these data, because of their high multicollinearity. Instead, Fig. 1 displays the environmental variables by their correlation with the axes. (I displayed correlations because they allow interpretation of the strength of the relationship because of their fixed range [-1,1], but theoretically it makes more sense to display regression coefficients $\mathbf{A}$ from the regression of $\mathbf{X}$ on $\mathbf{F}$, cf. R. Jongman *et al.*, 1987: pag. 135). Fig. 1 is remarkably similar to the ordination based on canonical correspondence analysis (C. Ter Braak, 1987: Fig. 6.3).

Example 2: analysis of an ecological experiment

G. Dirkse *et al.* (1989) studied the effects of experimental fertilization on forest undergrowth in stands of Scots Pine. Here we analyze the data on plant composition from

a 4 × 3 factorial experiment in 4 complete blocks in Lisselbo (Sweden). Such an experiment is traditionally analyzed by analysis of variance. Because of the large number of plant species compared to the number of units, classical multivariate analysis of variance is impossible. Instead I present in Tab. II an analysis of distance based on (5) and (6) using Pythagorean dissimilarity. Therefore, Tab. I is equivalent to an analysis of variance totalled across plant species. The significance values were obtained using Monte Carlo permutation of the residual distances (C. Ter Braak, 1991) with the first eigenvalue of the constrained analysis as test statistic. Apparently, Nitrogen (N) has a major effect but the small effect of Phosphorous (P) is found to be significant also. There is no evidence for an interaction effect. In this example, the 2-d unconstrained PCO shows the effects of blocks and Nitrogen only. The effect of Phosphorous shows up only in a 2-d constrained ordination of N + P with block effects partialled out. This is achieved by requiring that the ordination axes should lie in the subspace spanned by N + P orthogonal to blocks (C. ter Braak, 1987). This example demonstrates how constraints brings the power of the analysis of variance and regression back into multidimensional scaling.

3 Nonmetric multidimensional scaling constrained by regression models

PCO and constrained PCO approximate distances *via* innerproducts. By consequence, distances are approximated from below: the represented distances are all smaller than or equal to the dissimilarities, because the techniques perform a projection in the full *q*-space to a smaller *r*-space (J. Meulman, 1986). This is a serious disadvantage when the fit is poor. The disadvantage is overcome by the scaling techniques that are based on least-squares on distances. For the nonmetric case, the loss-function is

$$L(\mathbf{F}) = \min_{\tilde{\Delta} \in \Gamma(\mathbf{\Delta})} \| \tilde{\Delta} - D(\mathbf{F}) \|^2, \qquad (8)$$

where $\Gamma(\mathbf{\Delta})$ denotes the set of dissimilarity matrices that are normalized monotone transformations of the original dissimilarities. Constrained MDS is obtained with the additional constraint that $\mathbf{F} = \tilde{\mathbf{X}}\,\mathbf{C}$, with $\tilde{\mathbf{X}} \in \Omega(\mathbf{X})$, where Ω is the set of matrices that are admissible transformations of the original regressors. The solution is no longer a simple eigen analysis, but must be searched for with a SMACOF-algorithm. Each iteration consists of an unconstrained update (a so-called Guttman transform of the previous **F**) and a metric projection onto the constrained configuration space. The metric projection problem does not need to be solved completely. Convergence of the algorithm is still retained if a step in the right direction is taken which satisfies the constraints (J. De Leeuw *et al.*, 1980). The steps can often be solved by alternating least-squares. Theory and applications using the SMACOF-II program are given by W. Heiser *et al.* (1983) and J. Meulman *et al.* (1984).

Classical multivariate analysis gives an optimal fit of innerproducts, correlations and primary data, but, like PCO, not of distances. This led J. Meulman (1986) to propose

alternative methods that are truly based on least-squares on distances. For example, distance-based PCA of an $n \times q$ matrix $\mathbf{Y}$ minimizes (8) with $\Gamma(\mathbf{\Delta})$ consisting of one element, namely $D(\mathbf{Y})$. In the nonlinear case, also optimal transformations on $\mathbf{Y}$ are admitted. Distance-based redundancy analysis could be defined likewise with an additional external constraint on $\mathbf{F}$ as below (8). J. Meulman (1988) proposes, however, a definition which stems from another way of imposing constraints (I. Borg *et al.*, 1980). Constrained PCA is in this definition

$$L(\mathbf{F}) = \min_{\mathbf{B}} \min_{\mathbf{C}} \{ \| \mathbf{Y} - \mathbf{FB} \|^2 + b^2 \| \mathbf{F} - \mathbf{XC} \|^2 \} \qquad (9)$$

with b an extra parameter that governs the strength of the constraint. For $b \rightarrow \infty$, (9) yields redundancy analysis (C. Bijleveld, 1989, pag. 38-42). However, J. Meulman (1988) bases her distance-based redundancy analysis on $b=1$ to allow for errors in the regressors. This idea works for regression ($q=1$, J. Meulman *et al.*, 1988), but does it work generally? The solution of (9) can be obtained *via* a PCA of the $n \times (q+p)$ matrix $(\mathbf{Y} \mid \mathbf{X}(\mathbf{X}'\mathbf{X})^{-½})$. If $p=1$ and $q \gg 1$, the solution is therefore very similar to an unconstrained PCA.

Unfolding is a special case of MDS (W. Heiser, 1981). As with square symmetric matrices, the problem can be solved heuristically using a loss-function which has some disadvantages but results in simple eigen analysis, or it can be solved more rigorously using a direct least-squares approach. For unfolding, the heuristic approach leads to correspondence analysis (W. Heiser, 1981, 1987) with as constraint form canonical correspondence analysis (C. Ter Braak, 1986, 1988; Y. Takane *et al.*, 1991). The direct least-squares approach with constraints can be solved using SMACOF-III (W. Heiser, 1981, 1987). The program PROXSCAL, announced in 1988 to contain all previous SMACOF-programs, has not yet been released, unfortunately. I am looking forward to having something like SMACOF 1-2-3 on my desktop computer!

References

Bijleveld C. J. H., 1989, Exploratory linear dynamic systems analysis, DSWO Press, Leiden.

Borg I. & Lingoes J. C., 1980, A model and algorithm for multidimensional scaling with external constraints on the distances, Psychometrika, 45, 25-38.

Carroll J. D., Green P. E. & Carmone F. J., 1976, CANDELINC: a new method for multidimensional analysis with constrained solutions, Paper presented at The International Congress of Psychology, Paris.

Davies P. T. & Tso M. K. -S., 1982, Procedures for reduced-rank regression, Applied Statistics, 31, 244-255.

De Leeuw J., 1988, Convergence of the majorization method for multidimensional scaling, Journal of Classification, 5, 163-180.

De Leeuw J. & Heiser W. J., 1977, Convergence of correction matrix algorithms for

multidimensional scaling, In: Geometric representations of relational data, ed. J. Lingoes, 735-752, Mathesis Press, Ann Harbor.

De Leeuw J. & Heiser W., 1980, Multidimensional scaling with restrictions on the configuration, In: Multivariate Analysis V, ed. P. R. Krishnaiah, 501-522, North-Holland, Amsterdam.

De Leeuw J. & Heiser W. J., 1982, The theory of multidimensional scaling, In: Handbook of Statistics, vol. 2, ed. P. R. Krishnaiah & L. N. Kanal, 285-316, North-Holland, Amsterdam.

Digby P. G. N. & Gower J. C., 1981, Ordination between- and within-groups applied to soil classification, In: Down to Earth Statistics: Solutions looking for Geological Problems, ed. D. F. Merriam, 63-75, Syracuse University Geology Contributions, Syracuse.

Dirkse G. M. & Van Dobben H. F., 1989, Effects of experimental fertilization on forest undergrowth in young Scots Pine in Sweden, Studies in Plant Ecology, 18, 62-64.

Gower J. C., 1966, Some distance properties of latent root and vector methods used in multivariate analysis, Biometrika, 53, 325-338.

Gower J. C. & Legendre P., 1986, Metric and Euclidean properties of dissimilarity coefficients, Journal of Classification, 3, 5-48.

Heiser W. J., 1981, Unfolding analysis of proximity data. Thesis, University of Leiden, Leiden.

Heiser W. J., 1987, Joint ordination of species and sites: the unfolding technique, In: Developments in numerical ecology, ed. P. Legendre & L. Legendre, 189-224, Springer-Verlag, Berlin.

Heiser W. J. & Meulman J., 1983, Analyzing rectangular tables by joint and constrained multidimensional scaling, Journal of Econometrics, 22, 139-167.

Jongman R. H. G., ter Braak C. J. F. & van Tongeren O. F. R., 1987, Data analysis in community and landscape ecology, Pudoc, Wageningen.

Kruskal J. B., 1964, Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis, Psychometrika, 29, 1-27.

Legendre P. & Fortin M-J., 1989, Spatial pattern and ecological analysis, Vegetatio, 80, 107-138.

Manly B. F. J., 1985, The Statistics of Natural Selection on Animal Populations, Chapman and Hall, London.

Mantel N. & Valand R. S., 1970, A technique of nonparametric multivariate analysis, Biometrics, 26, 547-558.

McArdle B., 1990, Detecting and displaying impacts in biological monitoring: spatial problems and partial solutions, In: The XV-th International Biometrics Conference. Proceeding of Invited Papers, ed. Z. Harnos, 249-255, International Biometric Society, Hungarian Region, Budapest.

Meulman J., 1986, A distance approach to nonlinear multivariate analysis, DSWO Press,

Leiden.
Meulman M., 1988, Nonlinear redundancy analysis via distances, In: Classification and related methods of data analysis, ed. H. H. Bock, 515-522, North-Holland, Amsterdam.
Meulman J. & Heiser W. J., 1984, Constrained multidimensional scaling: more directions than dimensions, In: COMPSTAT 1984, ed. A. Anon, 137-142, Physica-Verlag, Vienna.
Meulman J. & Heiser W. J., 1988, Second order regression and distance analysis, In: Data, expert knowledge and decisions, ed. W. Gaul & M. Schader, 368-380, Springer-Verlag, Berlin.
Takane Y., Yanai H. & Mayekawa S., 1991, Relationships among several methods of linearly constrained correspondence analysis, Psychometrika, 56, 00-00.
Ter Braak C. J. F., 1986, Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis, Ecology, 67, 1167-1179.
Ter Braak C. J. F., 1987, CANOCO - a FORTRAN program for canonical community ordination by [partial] [detrended] [canonical] correspondence analysis, principal components analysis and redundancy analysis (version 2.1), Report 87 ITI A 11, Institute of Applied Computer Science, Wageningen.
Ter Braak C. J. F., 1988, Partial canonical correspondence analysis, In: Classification and related methods of data analysis, ed. H. H. Bock, 551-558, North-Holland, Amsterdam.
Ter Braak C. J. F., 1991, Permutation versus bootstrap significance tests in multiple regression and ANOVA, In: Bootstrapping and related resampling techniques, ed. K-H. Jöckel, 000-000, Springer Verlag, Berlin.
Torgerson W. S., 1958, Theory and methods of scaling, Wiley, New York.

Appendix [Patrick Groenen pointed out that the result holds true but that the proof is not. The proof contains an error [the second equality in (A.1)]. For a correct proof see Borg, I., Groenen, P J. F. (2005) Modern Multidimensional Scaling, Springer Springer Series in Statistics.]
http://www.springer.com/west/home/statistics/social?SGWID=4-10137-22-52088824-0

By setting $\mathbf{B} = -\frac{1}{2}\,\mathbf{J}\,\boldsymbol{\Delta}^2\,\mathbf{J}$ and expanding $-\frac{1}{2}\,\mathbf{J}\,D^2(\mathbf{F})\mathbf{J}$ we obtain, successively,

$$L(\mathbf{F}) = L(\mathbf{XC}) = \left\|\mathbf{B} - \mathbf{X}^{\circ}\,\mathbf{B}\,\mathbf{X}^{\circ} + \mathbf{X}^{\circ}\,\mathbf{B}\,\mathbf{X}^{\circ} - \mathbf{XCC'X'}\right\|^2 =$$
$$\left\|(\mathbf{I} - \mathbf{X}^{\circ})\mathbf{B}(\mathbf{I} - \mathbf{X}^{\circ}) + \mathbf{X}^{\circ}(\mathbf{B} - \mathbf{XCC'X'})\mathbf{X}^{\circ}\right\|^2 =$$
$$\left\|(\mathbf{I} - \mathbf{X}^{\circ})\mathbf{B}(\mathbf{I} - \mathbf{X}^{\circ})\right\|^2 + \left\|\mathbf{X}^{\circ}(\mathbf{B} - \mathbf{XCC'X'})\mathbf{X}^{\circ}\right\|^2 \qquad \text{(A.1)}$$

so the minimum is obtained by minimizing the second term on the last line of (A.1), which is achieved by the eigen analysis in (3). Note that **C** can also be obtained from the

spectral decomposition of the $p \times p$ matrix

$$\mathbf{A} \equiv (\mathbf{X}'\mathbf{X})^{-½}\,\mathbf{X}\,\mathbf{B}\,\mathbf{X}'(\mathbf{X}'\mathbf{X})^{-½} = \mathbf{Q}\,\boldsymbol{\Lambda}^{½}\,\mathbf{Q}', \tag{A.2}$$

by setting $\mathbf{C} = (\mathbf{X}'\mathbf{X})^{-½}\,\mathbf{Q}$. This can verified by expanding

$$\mathbf{X}'(\mathbf{X}'\mathbf{X})^{-½}\,\mathbf{A}\,(\mathbf{X}'\mathbf{X})^{-½}\,\mathbf{X} = \mathbf{X}^{\circ}\,\mathbf{B}\,\mathbf{X}^{\circ} = \mathbf{P}\,\boldsymbol{\Lambda}\,\mathbf{P}' \tag{A.3}$$

where $\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-½}\,\mathbf{Q} = \mathbf{X}\mathbf{C}$. Because $\mathbf{P}'\mathbf{P} = \mathbf{I}$ and the unicity of the spectral decomposition, $\mathbf{P}$ is also the solution of (3).

Table I. Eigenvalues of unconstrained and constrained principal coordinate analysis (PCO) on Nei's genetic distance beteen butterfly colonies with respect to environmental variables (see text and Figure 1).

eigenvalues	PCO	constrained PCO
λ_1	1.37	1.26
λ_2	0.28	0.25
λ_3	0.12	0.04
sum of λ's	1.77	1.55
sum of negative λ's	-0.09	-0.03

Table II. Analysis of distance of a 4×2 factorial experiment in 4 complete blocks (see text).

Source	df	mean pairwise distance (%)	P-value based on λ_1
Blocks	3	10	
+ N	3	43	.01
+ P	1	4	.01
+ N.P	3	5	.36
residual	21	38	
total	31	100	

Fig. 1. Ordination diagram based on constrained principal coordinate analysis of Nei's genetic distances between 21 colonies with respect to 11 environmental variables (data from B. Manly, 1985). The environmental variables are (arrows): altitude, latitude, annual precipitation (ANNUPREC), annual maximum and minimum temperature (MAX TEMP, MIN TEMP), highest and lowest temperature in the post-diapause (H-POSDIA, L-POSDIA), the adult phase (H-ADUL, L-ADUL). The two short arrows without names are the highest and lowest temperature in the pre-diapause. The two- and three-letter abbreviations of colony names follow B. Manly (1985). The scale-marks apply to the arrows.