Book review

Biplots in Practice

Michael Greenacre

BBVA Foundation, Rubes Editorial

Biplots in Practice is, as the title states, evidently a book about biplots. The book has a very didactic format, with short chapters giving some theory and examples, followed by a summary of the main points of the chapter, a style that is strongly reminiscent of the author's earlier book, *Correspondence Analysis in Practice* (Greenacre, 1993). The book is aimed at applied scientists who have a need to convert large tables of numbers into graphical displays, though will be useful for students in multivariate analysis as well. For a good understanding of the text, some background in matrix algebra and regression is required. Each chapter of the book basically presents a particular type of biplot, related to a specific multivariate technique. The final three chapters concern case studies in biomedicine (gene expression data), socioeconomics (survey research) and ecology (fish morphology and diet). The book has four appendices: a computational appendix with the R code, a bibliography on biplot literature, a glossary of terms and an epilogue by the author. The book is available in electronic format on-line at the website of the BBVA foundation at no cost. On-line books offer the possibility of continued correction and modification which potentially may convert this book into a "living book". The graphics and typesetting of the book are excellent, it is very difficult to find any mistakes in text or formulas.

The book introduces the biplot in a very elegant manner: as a multivariate generalization of the scatterplot, linked to the factorization of a data matrix as the product of two matrices: the biplot points and the biplot vectors. The definition of the scalar product and it associated geometry then follow naturally. Calibration of the biplot vectors is used to illustrate biplot interpretation.

Chapters 2, 3 and 4 link biplots to trivariate regression, using the analogy between the regression factorization $\hat{\mathbf{Y}} = \mathbf{X}\,\mathbf{B}$ and the data matrix factorization $\mathbf{X} = \mathbf{A}\,\mathbf{B}'$ used in biplots. Biplot vectors are presented as gradient vectors of a plane in three dimensions that point towards the direction of steepest ascent. Calibration is again used to equip the gradient vectors with scales, to show that the gradient vector hardly differs from an ordinary scatterplot axis. The idea is extended to regressions with transformed response variables.

The book has a nice strong focus on the close link between biplots and regression. Biplot coordinates (both points and vectors) can always be interpreted as regression coefficients. Once this relationship is understood, it becomes a particularly easy exercise to fit supplementary points and supplementary variables in a biplot, indeed, by just doing regression and plotting the regression coefficients. The extensive terminology (GLM biplot, poisson regression biplot, logistic regression biplot, MDS biplot, etc.) introduced in these chapters seems superfluous. It suggests we have many "different" biplots, but in fact we use the same regression principle all the time, and the single term "regression biplot" would suffice. Another point is the geometric framework in which these chapters are cast. The reader has to imagine a plane in the third dimension, and plot the gradient vector into the horizontal plane below it. It may have appeal to many readers, but I think it is not necessary to go to a third dimension. The biplot vector can also be found by searching for an optimal direction for a variable *within* the two-dimensional scatterplot of the predictors. Least squares minimization of projection errors obtained when projecting scatterplot points onto vectors inside the scatterplot will lead directly to the regression formula for representing the variable (Graffelman & Aluja-Banet (2003)). In fact, the term "supplementary variable" is sorely missing in chapters 2 through 4: the truth is that we are trying to fit supplementary variables in two-dimensional scatterplots all the time.

Chapter 5 tackles what, from a didactical point of view, is probably the most challenging part of biplot theory: the singular value decomposition (SVD). Depending on the public, a lecturer in statistics may wish to explain biplots without the SVD, and precisely the previous chapters of the book have shown that this is very well possible. However, if the audience has a basic understanding of matrix algebra, then the SVD is certainly enlightening as the unifying matrix approximation tool underlying many multivariate methods, and it will pave the way for explaining row and column coordinates, goodness of fit, and differences in scaling. The exposition of the SVD in this chapter is neat and concise, rank and dimensionality are smoothly presented, and the author proceeds from the unweighted to the weighted case, with both weights for cases and variables. The last sections of the chapter treat the approximation of a symmetric distance matrix by the SVD, to show the link between PCA and classical (metric) scaling. I feel that this section will not be understood by readers who do not have a solid background in multidimensional scaling (MDS), as the double-centring of the distance matrix and the multiplication by $-\frac{1}{2}$ are left unexplained.

The next chapter on biplots in principal component analysis (PCA) is in my eyes the most controversial chapter of the book. First of all, the notation used here for PCA is far from standard. PCA is mainly used for analyzing a quantitative data matrix, and it is fairly standard to refer to the latter as a $n \times p$ matrix (cases times variables) instead of the I times J employed by the author. Then, the centred and scaled data matrix is called $\mathbf{S}$, whereas as $\mathbf{S}$ is the typical notation used to indicate a covariance matrix. Finally,

to indicate row and column coordinates four matrices are used, $\mathbf{F}$, $\mathbf{\Gamma}$ and $\mathbf{G}$, $\mathbf{\Psi}$. Since $\mathbf{F}$ and $\mathbf{\Gamma}$ refer to the same entities (rows), they are better indicated by the same letter, and using a different subscript to indicate the scaling. The same applies to the column markers. One cannot escape from the impression that good old PCA is dressed up and put on stage in a correspondence analysis outfit, using the author's notation from the latter context.

Curiously enough, the term "principal components" seems mainly restricted to the title of the chapter, the components are not mentioned, computed or interpreted. I would recommend computing and plotting the principal components, in order to make these new synthetic variables tangible. Moreover, a scatterplot of the principal components is half a biplot, only the arrows for the variables are missing to complete the latter. Matrix $\mathbf{F}$ in this chapter comes close to the principal components: it contains the components but divided by a factor $\sqrt{p}$. Why so? The fact that we obtained scaled principal components is a direct consequence of scaling the matrix that enters the SVD by $1/\sqrt{p}$. Consequently, the singular values are scaled by $\sqrt{p}$, and the eigenvalues by p. It may be a matter of taste ("cada maestrillo tiene su librillo" as they say in Spain), but I'd rather prefer the SVD of $(1/\sqrt{n})\mathbf{X}_c$, where $\mathbf{X}_c$ contains the centred data. This way the SVD takes "half" of the expression of the covariance matrix, and the squared singular values are the eigenvalues of the covariance matrix and also the variances of the principal components. Most of the things we compute then have a direct interpretation, interpretations that are lost in the rescaling used in the book. The fact that the eigenvalues in the book are the eigenvalues of the covariance matrix but divided by p, provokes that all eigenvalues are smaller, and that the differences between the successive eigenvalues become smaller as well. Consequently, the usual difference in dispersion between the horizontal and vertical axis in the PCA biplot becomes attenuated, more difficult to perceive. If you teach PCA by maximizing the variance of a linear combination of the variables, then it is nice to be able to show plots where the higher variance of the first component is clearly visible. The rescaling used in the book obscures this. A very positive aspect of this chapter is that it presents the full variance decomposition over axes and over points, showing the computation of goodness of fit for each point, and contributions to axes. These additional statistics have always accompanied standard CA output, but were rarely computed in PCA. Another point is that PCA biplots in this book are all based on a PCA of the covariance matrix. A different type of biplot is possible by doing a PCA of the correlation matrix. There are no simple linear relationships that relate the results of covariance based PCA and a correlation based PCA. The latter may actually be the more common form of PCA, because it is often used when the variables have different units. In biplots from a correlation based PCA scalar products between vectors approximate the correlations between the variables. A full treatment of PCA biplots then requires four biplots: two for the covariance based PCA and two for the correlation based PCA, with the singular values to the right or to the left in each case.

Chapter 7 is an interesting contribution, showing how data that have been transformed as log-ratios can be represented in a biplot and interpreted, and how natural laws can be inferred from such plots.

The next three chapters deal with biplots in CA, moving from simple to multiple CA in a natural way: first comes a two-way table, then concatenated tables, and finally the full Burt matrix. The first CA chapter starts with a controversial phrase "CA is the most versatile of the methods based on the SVD for visualizing data". Metric multidimensional scaling (in a weighted form), also known as principal coordinate analysis, underlies CA and many other multivariate methods and may therefore be regarded more versatile. Classical canonical *correlation* analysis (CCO), (not to be confused with canonical *correspondence* analysis (CCA)) also underlies CA, and may also be considered more versatile. Canonical correlation analysis allows the construction of biplots of the between set correlation matrix (Haber and Gabriel, 1976; Ter Braak, 1990; Graffelman, 2005). These biplots are not treated in this book, and that may be considered an omission, since these are tightly related to the CA biplots described in the book.

Simple CA is concisely presented by means of the SVD of the matrix of standardized residuals. For the unfamiliarized, the "standardized residuals" may fall a bit out of the sky, for why would we want to analyze standardized residuals? Some indications that CA studies deviations from an independence model would be welcome in this context. The asymmetric CA biplots are presented with examples. The final section on CA presents the "contribution biplot", a rescaled version of an asymmetric biplot that allows us to easily identify the main contributors to each axis. But is this contribution biplot now really the most interesting way to communicate the results? When interpreting a biplot, we may rather like to focus on those points that have high goodness of fit, so that we are safe about our interpretations. Thus, why don't we scale the standard coordinates in such a way that their vector length equals R^2 of the corresponding regression? This way the longest vectors correspond to the best represented column categories, and they are easily identified as such. It can all be done, and we call the corresponding biplot a "quality biplot", and another biplot scaling is born. It's not my purpose to create new biplot scalings, I raise this issue because in my opinion statisticians have proposed so many ways of scaling biplots that the situation has become chaotic. An inexperienced researcher wishing to make some biplots is confronted with a myriad of scaling possibilities, and will have a hard time just to figure out which scaling is needed, and wondering whether he/she has chosen the "right" scaling for his/her dataset, and be pretty much upset by the fact that the plots resulting from different scalings can look rather different. I feel that for the users of biplot methodology, some simple practical rules are needed, but it is beyond the scope of this review to expose them here in detail. Representing supplementary points in biplots, a classical issue in CA, is treated by using the weighted average relationship between rows and columns. This topic could

be very will linked with the regression approach from the first four chapters of the book, because the coordinates of a supplementary point in a biplot are regression coefficients. The regression approach is unifying, supplementary points in PCA can be obtained by applying the same principle.

The chapter on discriminant analysis biplots is less clear than the other chapters of the book. The topic is initially presented in close relationship with CA and log-ratio analysis, whereas the in the last section classical linear discriminant analysis (LDA) is presented in the form of a SVD. Biplots in LDA are not so well-known as PCA or CA biplots, which makes this chapter interesting. It seems more logical to treat the biplots obtained from classical LDA first. The author states that a CA of a set of concatenated tables is also a discriminant analysis, but this is far from clear, and not further explained.

Chapter 12 is an introduction to constrained biplots. The topic is presented from the perspective of the projection the data matrix of interest onto a subspace spanned by constraining variables.

The final three chapters are case studies demonstrating the use of biplot methodology in biomedicine (gene expression data), socioeconomics (survey research) and ecology (fish morphology and diet). Many of the classical texts in multivariate analysis still suffer from the fact that example data sets are analyzed that often do not even occupy half a printed page. The data sets used in this book, particularly those of the case studies, are of considerable size and come much closer to the large databases often used in modern research. Chapter 13 addresses the topic of reduction of the number of variables in a microarray experiment, with the purpose of identifying those variables (genes) that discriminate different types of cancer. The first section of the chapter tries to accomplish this by PCA, using sequential removal of genes based on the contribution to the PCA solution. This approach is open to a lot of methodological criticism. Why is contribution to the solution taken as a criterion? It is not specified how contribution is measured, is it with respect to a 2, 3, 4 or even higher dimensional solution? Moreover, in PCA there is no guarantee that the first few dimensions do contain the relevant information that separates the cancer types; part of this information may be present in the last principal component. The use of the procrustes statistic to monitor the change in the configuration also requires a choice of dimensionality that is left unspecified here. The final section repeats the analysis of the data, and is a very interesting application of the more natural approach, discriminant analysis, now taking contributions to group differentiation as a criterion for removal. Quadratic discriminant analysis is not considered. The second case study, chapter 14, contains applications of various form of CA to social survey data, with special attention for missing values and middle categories. The last case study investigates the relationships between two sets of variables registered for a sample of fish, morphological and diet variables. The author has chosen for constrained CA, and a constrained log-ratio analysis to analyze the data. The results are interesting, but there is ample margin for discussion of how these data should be analyzed. First of all, the

layout of the data, two sets of different, quantitative variables is the classical layout for canonical correlation analysis and for multivariate regression. So why not use these tools? Moreover, one of the main reasons for using the constrained approach (CCA or redundancy analysis (RDA)) in ecology is that there are typically more variables in one set than there are observations. This leads to singularity of the within set covariance matrix of one set of variables, and this inhibits the use of CCO, because it needs to invert these. But for the fish data, there are more fish than variables, and singularity is not a problem. CA is used for diet data with the argument that there are many zeros in the data set. However, the data come in percentage form. Some amalgamation of food categories may greatly reduce the number of zeros, and the CODA (compositional data analysis) approach of log-ratio transformations of the diet variables may become feasible. CCO or multivariate regression with two sets of log-ratio transformed variables then may be an alternative. A permutation test is used as a practical criterion for variable selection. However, if there are strong correlations between the predictors, the results may overstate the importance of the selected variables.

The computational appendix gives website references for downloading the data sets and R scripts. The scripts are documented in this appendix and show how to construct most of the biplots in the book. This will be of great practical value for the readers, enabling them to repeat or modify any analysis in the book, as well as for analyzing their own data.

Most of the literature on biplots is available in the form of research articles. The author has chosen not to include any references in the chapters, but to give some references with comments in an appendix. The same appendix also contains references to R software and R packages, and to some relevant websites. The bibliography does not pretend to be complete, though a few important references that are tightly related with some topics addressed in the book are missing: the seminal paper of Ter Braak (1986) on canonical (constrained) correspondence analysis, Gabriel and Odoroff (1990) and Graffelman and van Eeuwijk (2005) for the topic of biplot calibration. The LDA biplot in the book concerns an analysis of group means, and this is closely related to Gabriel's MANOVA biplot (Gabriel, 1995).

Finally, the epilogue gives additional reflections of the author about biplots and their future, and contains many useful recommendations on good biplot design beyond setting the aspect ratio to 1. *Biplots in Practice* is, in short, a very welcome text in the field that will certainly help to disseminate biplot theory and help many researchers to make nice pictures of their data.

Jan Graffelman
jan.graffelman@upc.edu
Department of Statistics and Operations Research
Universitat Politècnica de Catalunya

References

Gabriel, K. R. (1995). MANOVA biplots for two-way contingency tables. In Krzanowski, W. J., editor, *Recent Advances in Descriptive Multivariate Analysis*, 227-268.

Gabriel, K. R. and Odoroff, C. L. (1990). Biplots in biomedical research. *Statistics in Medicine*, 9, 469-485.

Graffelman, J. (2005). Enriched biplots for canonical correlation analysis. *Journal of Applied Statistics*, 32, 173-188.

Graffelman, J. and Aluja-Banet, T. (2003). Optimal representation of supplementary variables in biplots from principal component analysis and correspondence analysis. *Biometrical Journal*, 45, 491-509.

Graffelman, J. and van Eeuwijk, F. A. (2005). Calibration of multivariate scatter plots for exploratory analysis of relations within and between sets of variables in genomic research. *Biometrical Journal*, 47, 863-879.

Greenacre, M. J. (1993). *Correspondence Analysis in Practice*. Academic Press.

Haber, M. and Gabriel, K. R. (1976). Weighted least squares approximation of matrices and its application to canonical correlations and biplot display. Technical report, University of Rochester, Department of Statistics.

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. (1990). Interpreting canonical correlation analysis through biplots of structure correlations and weights. *Psychometrika*, 55, 519-531.