Biplots in Practice

MICHAEL GREENACRE

Professor of Statistics at the Pompeu Fabra University

Appendix B Offprint

Bibliography

Bibliography

This bibliography is not intended to be complete but rather gives the main literature and web resources about biplots so that the reader can continue to learn about this method.

The term "biplot" originates in Ruben Gabriel's *Biometrika* paper in 1971:

- Gabriel, K.R. (1971). The biplot graphic display of matrices with application to principal component analysis. *Biometrika* 58, 453–467.

This paper, which at the time of writing has 1008 citations on Google Scholar and 682 on the Science Citation Index (ISI Web of Knowledge), is widely regarded as the origin of the idea. It is worthwhile to repeat its abstract:

> "Any matrix of rank two can be displayed as a biplot which consists of a vector for each row and a vector for each column, chosen so that any element of the matrix is exactly the inner product of the vectors corresponding to its row and its column. If a matrix is of higher rank, one may display it approximately by a biplot of a matrix of rank two which approximates the original matrix. The biplot provides a useful tool of data analysis and allows the visual appraisal of the structure of large data matrices. It is especially revealing in principal component analysis, where the biplot can show inter-unit distances and indicate clustering of units as well as display variances and correlations of the variables."

A less cited paper by Ruben Gabriel, but nevertheless one of my favourite ones on the biplot, appeared the following year in the *Journal of Applied Meteorology* (Ruben was also well-known for his work as a statistician in weather modification projects):

- Gabriel, K.R. (1972). Analysis of meteorological data by means of canonical decompositions and biplots. *Journal of Applied Meteorology* 11, 1071–1077.

In this paper he gives the biplot associated with linear discriminant analysis, also known as canonical variate analysis. He also talks about the vectors linking pairs of variables in a biplot (like the "links" in log-ratio analysis).

Another gem is by Dan Bradu and Ruben Gabriel in *Technometrics* in 1978:

• Bradu, D. and Gabriel, K.R. (1972). The biplot as a diagnostic tool for models of two-way tables. *Technometrics* 20, 47–68.

In this paper they show how certain models lead to points lying in straight lines in the full space of the data, and thus approximately in a biplot that has a good fit to the data. Thus a subset of row points and/or column points lying in a straight line in a biplot suggest models in that submatrix of the data. In addition, orthogonality of the lines suggests a simpler model.

All the above papers are required reading for those interested in the origins of the technique.

Other authors also had the idea of adding variables to an existing configuration of points to make joint displays, although they did not call them biplots. For example, Doug Carroll's vector model for preferences is a biplot:

• Carroll, J.D. (1972). Individual differences and multidimensional scaling. In R.N. Shepard, A.K. Romney, and S.B. Nerlove, eds, *Multidimensional Scaling: Theory and Applications in the Behavioral Sciences (Vol. 1)*, 105–155. Seminar Press, New York.

Only one book exists to date specifically on the topic of biplots, by John Gower and David Hand:

• Gower, J.C. and Hand, D.J (1996). *Biplots*. Chapman & Hall, London, UK.

This book is very complete, both on linear and nonlinear biplots, giving a rigorous theoretical treatment of the subject. Another book by John Gower is with co-authors Sugnet Gardner-Lubbe and Niel le Roux:

• Gower, J.C., Gardner-Lubbe, S. and le Roux, N. (2010). *Understanding Biplots*. Wiley, Chichester, UK.

As far as the vast literature on the singular value decomposition (SVD) is concerned, I mention only two sources, by the author of one of the landmark algorithms for the SVD, Gene Golub in 1971, which seems to be an important year for the biplot:

• Golub, G.H. and Reinsch, C. (1971). The singular value decomposition and least squares solutions. In J.H. Wilkinson and C. Reinsch, eds, *Handbook for Automatic Computation*, 134–151. Springer-Verlag, Berlin.

and the other a classic book by Paul Green and Doug Carroll, originally published in 1976, which was the first time I saw the geometric interpretation of the SVD (called "basic structure" by these authors)—this book is invaluable as a practical introduction to matrix and vector geometry in multivariate analysis:

- Green, P.E. and Carroll, J.D. (1997). *Mathematical Tools for Applied Multivariate Analysis, Revised Edition.* Academic Press, New York.

Most books or articles that treat the methods presented in this book will have a section or chapter on biplots and their interpretation in the context of that method. This is just a tiny selection of some of the literature that can be consulted, and by no means the primary references:

Principal component analysis
- Joliffe, I.T. (2002). *Principal Component Analysis* (2nd edition). Springer, New York.

Log-ratio analysis (unweighted form)
- Aitchison, J. and Greenacre, M. (2002). Biplots of compositional data. *Applied Statistics* 51, 375–392.

Log-ratio analysis (weighted form)
- Greenacre, M. and Lewi, P.J. (2009). Distributional equivalence and subcompositional coherence in the analysis of compositional data, contingency tables and ratio scale measurements. *Journal of Classification* 26, 29–54.

Correspondence analysis
- Greenacre, M. (2007). *Correspondence Analysis in Practice* (2nd edition). Chapman & Hall/CRC, London. Spanish translation may be freely downloaded at: `http://www.fbbva.es` and `http://www.multivariatestatistics.org`

Multiple correspondence analysis
- Greenacre, M. and Blasius, J., eds (2006). *Multiple Correspondence Analysis and Related Methods*, Chapman & Hall/CRC Press, London.
- Michalidis, G. and de Leeuw, J. (1998). The Gifi system for descriptive multivariate analysis. *Statistical Science* 13, 307–336.

Discriminant analysis/centroid biplots
- Hastie, T., Tibshirani, R. and Friedman, J. (2009). *The Elements of Statistical Learning* (2nd edition). Springer, New York. This book may be freely downloaded at `http://www-stat.stanford.edu/~tibs/ElemStatLearn/`

Constrained biplots
- Legendre, P. and Legendre, L. (1998). *Numerical Ecology* (2nd edition). Elsevier, Amsterdam.

Finally we give some resources on the internet, on R packages relevant to this book (in alphabetic order of package names).

- Thioulouse, J. and Dray, S. (2007). Interactive multivariate data analysis in R with the **ade4** and **ade4TkGUI** packages. *Journal of Statistical Software.* Download from `http://www.jstatsoft.org/v22/i05/paper`

- De Leeuw, J. and Mair, P. (2009). Simple and canonical correspondence analysis using the R package **anacor**. *Journal of Statistical Software.* Download from `http://www.jstatsoft.org/v31/i05/paper`

- De Leeuw, J. and Mair, P. (2009). Gifi methods for optimal scaling in R: the package **homals**. *Journal of Statistical Software.* Download from: `http://www.jstatsoft.org/v31/i04/paper`

- La Grange, A., le Roux, N. and Gardner-Lubbe, S. (2000). **BiplotGUI**: Interactive biplots in R. *Journal of Statistical Software.* Download from: `http://www.jstatsoft.org/v30/i12/paper`

- Nenadić, O. and Greenacre, M. (2007). Correspondence analysis in R, with two- and three-dimensional graphics: The **ca** package. *Journal of Statistical Software.* Download from `http://www.jstatsoft.org/v20/a03/paper`

- Markos, A. (2010). **caGUI**: a Tcl/Tk GUI for the functions in the **ca** package. Download from `http://cran.r-project.org/web/packages/caGUI/index.html`

- Graffelman, J. (2010). **calibrate**: Calibration of scatterplot and biplot axes. Download from `http://cran.r-project.org/web/packages/calibrate/index.html`

- Oksanen, J. (2010). **vegan**: Community Ecology Package. Download from `http://cran.r-project.org/web/packages/vegan/index.html`

And some relevant websites:

`http://www.multivariatestatistics.org`
  Supporting website for the series of statistics books published by the BBVA Foundation, including the present book *Biplots in Practice,* with glossary of terms and chapter summaries in Spanish, as well as supplementary material such as animated graphics and links to the data sets and R code.

`http://www.carme-n.org`

Correspondence Analysis and Related Methods Network, with R scripts and data from this book, from *Correspondence Analysis in Practice,* Second Edition, and from *Multiple Correspondence Analysis and Related Methods.*

`http://gifi.stat.ucla.edu`

Jan de Leeuw's website for the Gifi system (centred around multiple correspondence analysis and related methods) and R functions

`http://www.imperial.ac.uk/bio/research/crawley/statistics`

Michael Crawley's material from his book *Statistics: an Introduction Using* R

`http://www.issp.org`

Source of many data sets from the International Social Survey Program

`http://www.r-project.org`

The R project for statistical computing

`http://cc.oulu.fi/~jarioksa/softhelp/vegan.html`

Jari Oksanen's website for the **vegan** package in R, a very complete package which includes PCA, CA, CCA and many more multivariate methods, as well as permutation tests.

`http://people.few.eur.nl/groenen/mmds/datasets`

Website with data sets from book *Modern Multidimensional Scaling* by Ingwer Borg and Patrick Groenen

`http://biplot.usal.es/ClassicalBiplot/index.html`

Website of José Luis Vicente Villardon's biplot software for biplots and simple correspondence analysis