Area Biplots

J.C. Gower* P.J.F. Groenen† M. van de Velden‡

November 1, 2007

Econometric Institute Report EI 2007-48

*Statistics Department, The Open University, Milton Keynes, MK7GAA, United Kingdom j.c.gower@open.ac.uk

†Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands groenen@few.eur.nl

‡Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands vandevelden@few.eur.nl

Abstract

Classical multivariate analysis techniques such as principal components analysis and correspondence analysis use inner products to estimate data values. The results of these techniques may be visualized by representing the row and column points jointly in a biplot where the projection of a row point onto a column point vector followed by a multiplication by the length of the column point vector gives the inner-product that estimates the corresponding data element. In this paper, we propose a new visualization: after a 90° rotation of the row points, the areas spanned by a triangle of a row point, a column point and the origin estimate the data values. In contrast to the projection biplot, the areas spanned by different row and column points can be compared directly. Areas can only be produced for two dimensions at a time, but higher dimensional solutions can be represented by summing areas over subsequent pairs of dimensions. Here, the area biplot is developed for principal components analysis, correspondence analysis, and for interaction biplots but has general applicability.

Keywords: Biplot, Correspondence Analysis, Interaction, Principal Component Analysis, Visualization

1 Introduction

In many methods of multivariate analysis, one is faced with evaluating an inner product $\mathbf{c}'\mathbf{d}$. Computationally there could hardly be anything more trivial but because of the strong visual component to some of these multivariate methods there is a need for a simple visual appreciation of inner products. The conventional visualization involves projecting $\mathbf{c}$ onto $\mathbf{d}$ and multiplying the result by the length of the vector $\mathbf{d}$. However, it is difficult to compare the estimates of the inner products that do not have the same $\mathbf{c}$ or $\mathbf{d}$; the comparison is only straightforward when either $\mathbf{c}$ or $\mathbf{d}$ is common to both inner products. This problem may be avoided by calibrating the vector $\mathbf{d}$ with the values of the inner product which may then be read off by projecting $\mathbf{c}$ onto $\mathbf{d}$; this is the basis of calibrated biplot methods. Calibrating the vectors $\mathbf{c}$ and $\mathbf{d}$ amounts to treating them like conventional coordinate axes. In this paper, we discuss a new alternative visualization, the *area biplot*, that uses area to estimate the inner product. By definition areas are two-dimensional, but we show that it is easy to combine more than one two-dimensional visualizations to visualize a higher dimensional solution.

The best known biplot is the principal components (PCA) biplot (Gabriel, 1971) where the relations between the samples and the variables are given by inner products. These relations are

often visualized by a two-dimensional display in which the samples are represented by points and the variables by vectors, although a reverse representation is also possible. Technically, it is not necessary to limit the biplot display to two dimensions, but it is the most convenient and most frequently used.

The remainder of this paper is organized as follows. In the next section, the biplot for PCA is defined explicitly. Section 3 introduces the area biplot for PCA. Then, in Sections 4 and 5, we show how our method can also be applied in correspondence analysis (CA) and biadditive models. Two examples of area biplots, are given in Section 6. We end with a discussion and some conclusions.

2 The PCA Biplot

For the moment, and without loss of generality, assume that $\mathbf{X}$ is column centered, that is, $\mathbf{1}'\mathbf{X} = \mathbf{0}'$, where $\mathbf{1}$ denotes a vector of ones of the appropriate order and, similarly, $\mathbf{0}$ denotes a vector of zeroes. Throughout the paper we shall use this notation. PCA can be expressed as minimizing

$$L(\mathbf{C}, \mathbf{D}) = \|\mathbf{X} - \mathbf{C}\mathbf{D}'\|^2 = \sum_{i=1}^{n}\sum_{j=1}^{p}(x_{ij} - \mathbf{c}_i'\mathbf{d}_j)^2, \tag{1}$$

where $\mathbf{C}$ is an $n \times k$ matrix with k-dimensional coordinates for the n sample points and $\mathbf{D}$ is a $p \times k$ matrix with the p variable vectors. Note that $\mathbf{c}_i'$ is row i of $\mathbf{C}$ and $\mathbf{d}_j'$ is row j of $\mathbf{D}$. The dimensionality k is chosen by the researcher and is typically smaller than the rank of the data matrix. If a visualization of the data is required, k is often set to 2.

It is well known that (1) can be minimized by using the singular value decomposition (SVD) of $\mathbf{X}$, that is, $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}'$ with $\mathbf{\Sigma}$ a diagonal matrix with nonnegative singular values in decreasing order, with $\mathbf{U}$ and $\mathbf{V}$ orthonormal matrices. Let $\mathbf{U}_2$ and $\mathbf{V}_2$ denote the first two columns of $\mathbf{U}$ and $\mathbf{V}$ and $\mathbf{\Sigma}_2$ the diagonal matrix with the two largest singular values. Then, a two-dimensional PCA solution is obtained by choosing $\mathbf{C}_{pca} = \mathbf{U}_2\mathbf{\Sigma}_2$ and $\mathbf{D}_{pca} = \mathbf{V}_2$. Note that, in this standardization, the singular values are distributed over the row points. The estimation of x_{ij} is done by evaluating the inner product $\mathbf{c}_i'\mathbf{d}_j = (\mathbf{u}_{i2}'\mathbf{\Sigma}_2)\mathbf{v}_{j2}$, where $\mathbf{u}_{i2}'$ is row i of $\mathbf{U}_2$ and $\mathbf{v}_{j2}'$ is row j of $\mathbf{V}_2$. Furthermore, by using this standardization, the distances between the sample points, that is the rows of $\mathbf{C}_{pca}$ approximate the inter-sample distances; the distances between the rows of $\mathbf{X}$ (e.g., Gower, 1966). For the estimation of the inner-products, the standardization is not important. For example, if we define $\mathbf{C} = q\mathbf{U}_2\mathbf{\Sigma}_2^{1-\alpha}$ and $\mathbf{D} = q^{-1}\mathbf{V}_2\mathbf{\Sigma}_2^{\alpha}$ the inner product matrix $\mathbf{C}\mathbf{D}'$ still approximates the data matrix $\mathbf{X}$. We can exploit this freedom in the visualization of inner products.

In the two-dimensional setting, the inner product $\mathbf{c}_i'\mathbf{d}_j$ may be written as

$$\mathbf{c}_i'\mathbf{d}_j = \|\mathbf{c}_i\| \ \|\mathbf{d}_j\| \cos(\theta_{ij}) \tag{2}$$

where $\|\mathbf{c}_i\|$ denotes the Euclidean length of vector $\mathbf{c}_i$, that is, $\|\mathbf{c}_i\| = (c_{i1}^2 + c_{i2}^2)^{1/2}$, and θ_{ij} is the angle between the vectors $\mathbf{c}_i$ and $\mathbf{d}_j$. In a biplot, the sample i is typically represented by a point with coordinates (c_{i1}, c_{i2}) and the variable j by a vector originating from the origin and ending in (d_{i1}, d_{i2}). To visualize $\mathbf{c}_i'\mathbf{d}_j$ and thus estimate data value x_{ij}, sample point i is projected onto vector j yielding

$$\|\mathbf{c}_i\| \ \cos(\theta_{ij}). \tag{3}$$

Clearly, to properly model the inner product defined in (2), (3) remains to be multiplied by $\|\mathbf{d}_j\|$. The projection in (3) can be done for all samples i and to estimate x_{ij} they should all be multiplied by the same $\|\mathbf{d}_j\|$. Hence, the biplot interpretation is: (1) project the sample points $\mathbf{c}_i$ onto the vector $\|\mathbf{d}_j\|$ representing variable j and (2) multiply this projection by the length of vector j, that is, by $\|\mathbf{d}_j\|$. The multiplication by the variable's vector length after the projection is a disadvantage of the biplot. The top panel of Figure 1 shows an example of the projection of sample 8 on variable 2. The projected estimate $\mathbf{c}_8'\mathbf{d}_2/\|\mathbf{d}_2\| = 1.89$ which subsequently needs to be multiplied by $\|\mathbf{d}_2\| = .70$ to get the estimate $1.89 \times .70 = 1.32$, which may be compared with the true data value $x_{82} = 1.08$. Of course, this calculation has to be done mentally, using the visual information in Figure 1 giving the lengths of the projection and the length of $\|\mathbf{d}_2\|$.

To eliminate this problem, Gower and Hand (1996) proposed using so called calibrated axes. In this approach, the lines representing the variables get marker points to indicate the predicted values. These calibrated marker points are obtained by dividing projected marker points by the length $\|\mathbf{d}_j\|$. Such a biplot with calibrated marker points is given in the bottom panel of Figure 1. Using calibrated marker points the estimated value can be read directly off the calibrated axes. The estimate of about 1.32 for x_{82} is found by projection and interpolation between the markers 1 and 1.5. The arrows and edge scales are redundant and, therefore, not shown in the bottom panel of Figure 1.

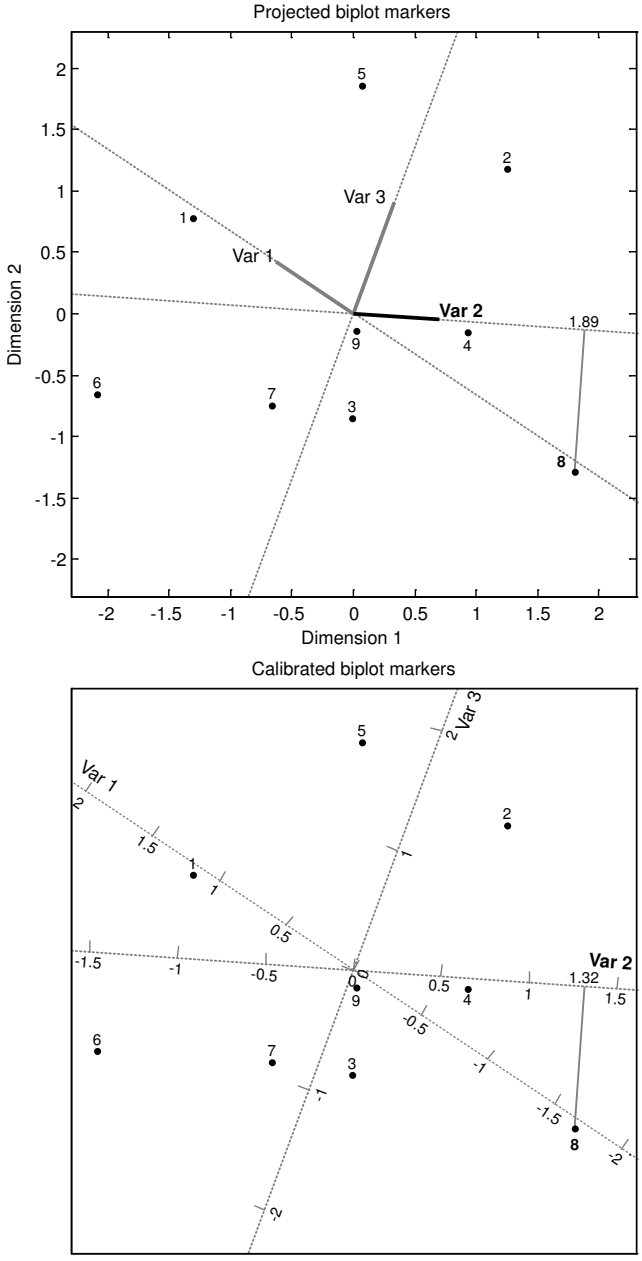

Figure 1: Example of projection in a biplot. Top panel shows only projection ignoring the length of the variable. In the bottom panel the length of the variables is taken into account by using calibrated biplot markers. In this case the value of x_{82} is 1.08.

3 The PCA Area Biplot

The idea of the two-dimensional area biplot is very simple. One set of points (for example, the sample points) is rotated by 90°. The estimate of the data value x_{ij} is obtained from the area determined by the origin, sample point i, and the vector end point of variable j. Indeed, the vector

representation is superfluous and it suffices to represent variables as well as samples by points, each with a distinguishing plotting symbol. In each area biplot, no calibrated axes are needed nor post multiplication by the vector length of the variable. One relies on the ability of the eye to compare areas of different shaped triangles.

To develop the area biplot, let $\mathbf{T}$ be a 90° rotation matrix. The idea is to rotate the sample points through 90°, which does not change the length of a vector from the origin. The choice of rotation by $90° = \pi/2$ follows from the equality

$$\cos(\theta_{ij}) = \sin(\theta_{ij} + \pi/2) = \sin(\phi_{ij}),$$

where $\phi_{ij} = \theta_{ij} + \pi/2$ is the angle between the rotated sample point $\mathbf{Tc}_i$ and the variable vector $\mathbf{d}_j$. Therefore, (2) can be expressed as

$$\mathbf{c}_i'\mathbf{d}_j = \|\mathbf{c}_i\| \, \|\mathbf{d}_j\| \sin(\phi_{ij}).$$

This may be immediately identified as twice the area of a triangle with sides $\|\mathbf{c}_i\| \, \|\mathbf{d}_j\|$ and angle ϕ_{ij} but the following derives the result ab initio.

Figure 2 shows $\mathbf{c}_i$, $\mathbf{d}_j$, and the rotated version $\mathbf{Tc}_i$. The part $\|\mathbf{Tc}_i\| \sin(\phi_{ij})$ is exactly that part of $\mathbf{Tc}_i$ orthogonal to $\mathbf{d}_j$, that is, the height of the shaded triangle. Therefore, the area of the triangle with height $\|\mathbf{Tc}_i\| \sin(\phi_{ij})$ and base $\mathbf{d}_j$ is exactly $\mathbf{c}_i'\mathbf{d}_j/2$ and this area estimates data value $x_{ij}/2$.

As mentioned before, the standardization of $\mathbf{C}$ and $\mathbf{D}$ does not affect the inner-products, or, equivalently, the areas. However, by using the typical PCA standardization where $\mathbf{C}'\mathbf{C} = \mathbf{\Sigma}^2$ and $\mathbf{D}'\mathbf{D} = \mathbf{I}$, the singular values are distributed over the row points. Hence, if there are relatively many rows (a large sample) the row coordinates tend to become rather small compared to the column points. In such a configuration, drawing and comparing triangle areas may not be an easy task. For a visual inspection of triangle areas, it may be desirable that the row and column points exhibit a similar spread. In that way, the construction and comparison of triangles, which may have to be executed mentally, becomes much easier. To attain "similar spread" we set

$$\mathbf{C} = \left(\frac{n}{m}\right)^{1/4} \mathbf{U}_2\mathbf{\Sigma}_2^{1/2} \text{ and } \mathbf{D} = \left(\frac{m}{n}\right)^{1/4} \mathbf{V}_2\mathbf{\Sigma}_2^{1/2}.$$

It is then easily verified that the rows and columns of $\mathbf{C}$ and $\mathbf{D}$ are standardized such that $\mathbf{C}'\mathbf{C} = (n/m)^{1/2}\mathbf{\Sigma}_2$ and $\mathbf{D}'\mathbf{D} = (m/n)^{1/2}\mathbf{\Sigma}_2$. Hence, the average spread for both row and column points

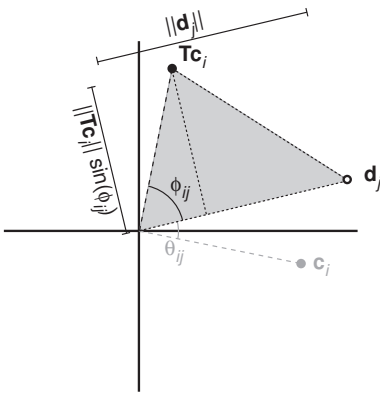

Figure 2: Triangle area interpretation of inner product $\mathbf{c}_i'\mathbf{d}_j/2$ through rotated sample point $\mathbf{Tc}_i$.

is equal to $(nm)^{-1/2}\mathbf{\Sigma}_2$.

The rule for interpreting the data values from the area biplot is to construct a triangle as indicated above. Positive estimates of x_{ij} are obtained by areas that concur with rotation in a anti-clockwise direction starting from $\mathbf{d}_j$ and negative estimates by a clockwise rotation. Zero areas occur whenever $\mathbf{d}_j$ and $\mathbf{Tc}_i$ are on the same line through the origin. We note that all points through $\mathbf{Tc}_i$ parallel to $\mathbf{d}_j$ generate the same area with $\mathbf{d}_j$, a result analogous to points on the circumference of a circle being equidistant from its center.

Figure 3 gives the example of the same random data of $n = 8$ samples and $p = 3$ variables as shown in Figure 1. The solid black points indicate the 90° rotated samples $\mathbf{CT}'$ and open circles denote the variables $\mathbf{D}$. The shaded area belongs to x_{82} as it is the triangle spanned by the origin, variable 2, and sample 8. The area is 1.32 and is positive as the angle ϕ_{82} between $\mathbf{d}_2$ and $\mathbf{Tc}_8$ is between 0 and π. Any black point $\mathbf{Tc}_i$ below the line through $\mathbf{d}_2$ has $-\pi < \phi_{i2} < 0$, hence will have negative estimates of x_{i2}.

The shaded triangle in Figure 3 is not the only triangle with this area as there is a whole set of triangles with the same area. In Figure 4, two other triangles are added with the same area as the triangle in Figure 3 estimating $x_{82}/2$. All these triangles share the line from the origin to $\mathbf{d}_2$. The third point can be chosen anywhere on the locus of points parallel to the line from the origin to $\mathbf{d}_2$ that goes through point $\mathbf{Tc}_8$. This property is useful for picking up values similar, in this case, to x_{82}. In Figure 4, all other estimated values are below the line and hence are less than 1.32.

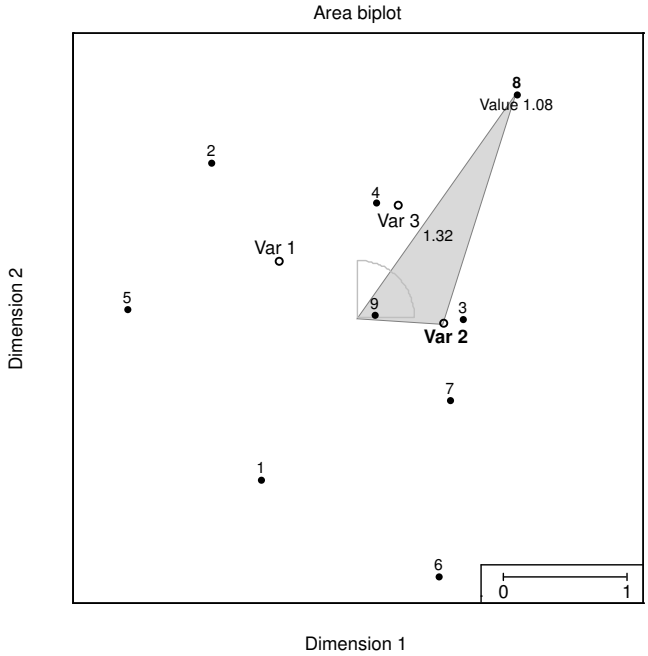

Figure 3: Example of triangle area biplot. Solid circles represent the rotated sample points in $\mathbf{CT}'$ and the open circles the variable points in $\mathbf{D}$. The shaded area estimates $x_{82}/2$.

3.1 The Multidimensional Area Biplot

If one wants to visualize estimated values in more than two dimensions, the area biplot remains valid. To see this, note that the k dimensional inner product may be written as

$$\mathbf{c}'\mathbf{d} = \sum_{j=1}^{k} c_j d_j = \sum_{j=1}^{2} c_j d_j + \sum_{j=3}^{4} c_j d_j + ... + \sum_{j=k-1}^{k} c_j d_j.$$

Hence, for an even number of dimensions, the inner-product can be factorized as the sum of two-dimensional inner-products. Then, by considering area biplots for subsequent pairs of dimensions, or, more precisely, by adding the areas in these plots, the multidimensional inner products are visualized. When plotting pairs of dimensions, care has to be taken to ensure that unit length is the same in all pairs; we refer to linked displays.

If an odd number of dimensions is chosen, a unidimensional area biplot is required to depict the contribution to the inner product for the dimension not contained in the two-dimensional biplots. It is not difficult to see that such a unidimensional area biplot can also be visualized as a two-dimensional area biplot. For the unidimensional case, the inner product simplifies into $c_{ik}d_{jk}$. Hence, in a two dimensional display, these values can be represented by the sample vector $[0\ c_{ik}]'$ and the variable vector $[d_{jk}\ 0]'$. Thus, the d_{jk} only vary along the horizontal axis and the c_{ik} are

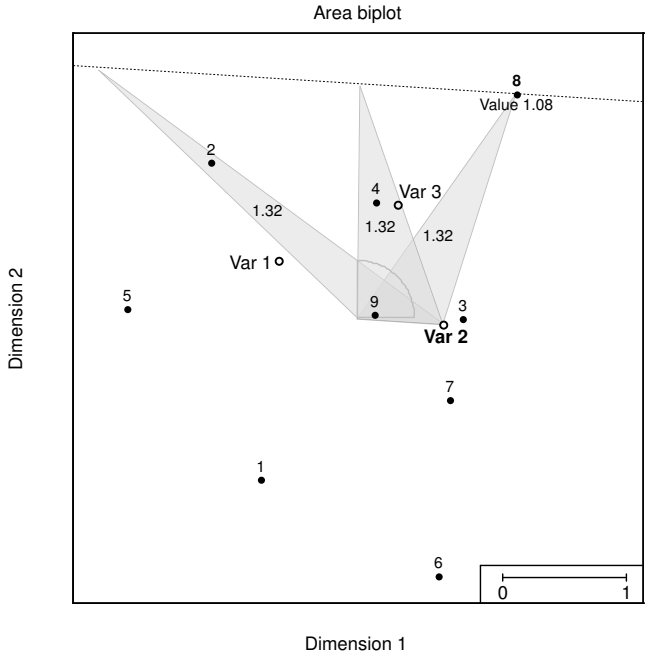

Figure 4: The three triangles have the same area all estimating $x_{82}/2$.

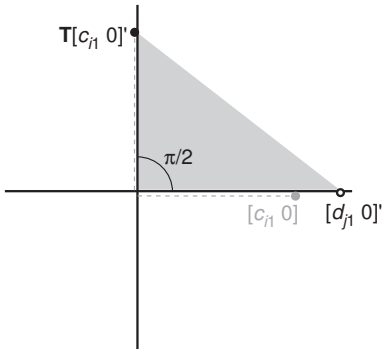

Figure 5: Unidimensional area biplot of inner product $c_{i1}d_{j1}$ through rotated sample point $\mathbf{T}[c_{i1}\ 0]'$ and the variable vector $[d_{j1}\ 0]$.

rotated through 90° to positions along the vertical axis. As before, the product $c_{ik}d_{jk}$ is twice the area of the triangle spanned by the origin, the sample point $(0, c_{ik})$, and the variable point $(d_{jk}, 0)$, see Figure 5. If the area is below the horizontal axis the estimated value is negative, if it is above the horizontal axis it is positive. A consequence of this is that for the same variable all estimated values span areas with the same base that only differ in height.

4 The Correspondence Analysis Area Biplot

Correspondence analysis (CA) analyzes association in an $n \times p$ two-way contingency table $\mathbf{F}$. In particular, a bilinear decomposition is done on the deviations from the independence model. Although CA is ideally performed on a contingency table, computationally, the elements of $\mathbf{F}$ may contain any nonnegative values. To introduce the technical details, we need some notation. Let $\mathbf{D}_r$ denote the diagonal matrix with the row sums of $\mathbf{F}$ as elements, $\mathbf{D}_c$ is the diagonal matrix with the column sums of $\mathbf{F}$ as elements, and s is the total sum of $\mathbf{F}$. Then, the independence model is given by $\mathbf{E} = \mathbf{D}_r\mathbf{1}\mathbf{1}'\mathbf{D}_c/s$. CA can be expressed as minimizing

$$L(\mathbf{R}, \mathbf{C}) = \|\mathbf{D}_r^{-1/2}(\mathbf{F} - \mathbf{E} - s^{-1}\mathbf{D}_r\mathbf{R}\mathbf{C}'\mathbf{D}_c)\mathbf{D}_c^{-1/2}\|^2, \tag{4}$$

where $\mathbf{R}$ and $\mathbf{C}$ are $n \times k$ and $p \times k$ matrices of row and column coordinates, respectively. The maximum dimensionality k is $\min(n, p) - 1$. For an exact solution to (4), $\mathbf{R}$ and $\mathbf{C}$ satisfy

$$\begin{aligned} s^{-1}\mathbf{D}_r^{1/2}\mathbf{R}\mathbf{C}'\mathbf{D}_c^{1/2} &= \mathbf{D}_r^{-1/2}(\mathbf{F} - \mathbf{E})\mathbf{D}_c^{-1/2} \\ \text{i.e. } \mathbf{R}\mathbf{C}' &= s\mathbf{D}_r^{-1}(\mathbf{F} - \mathbf{E})\mathbf{D}_c^{-1}. \end{aligned}$$

Then, the inner products in $\mathbf{R}\mathbf{C}'$ estimate the elements

$$\mathbf{r}_i'\mathbf{c}_j = \frac{f_{ij} - e_{ij}}{e_{ij}} = \frac{f_{ij}}{e_{ij}} - 1, \tag{5}$$

where $\mathbf{r}_i'$ is row i of $\mathbf{R}$ and $\mathbf{c}_j'$ row j of $\mathbf{C}$. In other words, the inner products $\mathbf{r}_i'\mathbf{c}_j$ estimates the deviation of f_{ij} from independence relative to the estimated value e_{ij} under independence. The ratio f_{ij}/e_{ij} is sometimes called the contingency ratio.

Approximations to $\mathbf{R}$ and $\mathbf{C}$ may be found from the SVD

$$s^{-1}\mathbf{D}_r^{1/2}\mathbf{R}\mathbf{C}'\mathbf{D}_c^{1/2} = \mathbf{D}_r^{-1/2}(\mathbf{F} - \mathbf{E})\mathbf{D}_c^{-1/2} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}'$$

giving

$$\mathbf{R}\mathbf{C}' = s\mathbf{D}_r^{-1/2}\mathbf{U}\mathbf{\Sigma}\mathbf{V}'\mathbf{D}_c^{-1/2}$$

and

$$\begin{aligned}\mathbf{R} &= s^{1/2}\mathbf{D}_r^{-1/2}\mathbf{U}\mathbf{\Sigma}^{\alpha} \\ \mathbf{C} &= s^{1/2}\mathbf{D}_c^{-1/2}\mathbf{V}\mathbf{\Sigma}^{1-\alpha}\end{aligned}$$

for any scalar α. For $\alpha = 1$ the so-called row principal normalization, for $\alpha = 0$ the column principal normalization, and for $\alpha = 1/2$ a symmetric normalization are obtained. All choices of α simply distribute the singular values $\mathbf{\Sigma}$ differently without changing the inner product $\mathbf{r}_i'\mathbf{c}_j$.

The important part in the derivation above, lies in (5) as it indicates that in a CA biplot the elements estimated by the inner product approximate the relative deviations from independence. Therefore, we can apply the rotation straight away: choose a pair of dimensions of interest, for example, dimensions 1 and 2 and rotate the row points by 90°. Then, the area of the triangle spanned by $\mathbf{Tr}_i$, $\mathbf{c}_j$, and the origin estimates the contingency ratio minus one, that is, $f_{ij}/e_{ij} - 1$.

5 Interaction Biplots

The biadditive model $\mathbf{Y} = \mu\mathbf{1}\mathbf{1}' + \boldsymbol{\alpha}\mathbf{1}' + \mathbf{1}\boldsymbol{\beta}' + \sum_{k=1}^{K}\sigma_k\boldsymbol{\gamma}_k\boldsymbol{\delta}_k'$ is often fitted to data given in an $n \times p$ two-way table $\mathbf{X}$, where $K = \min(n, p)$. Here, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are usually termed main effects while the biadditive terms $\boldsymbol{\gamma}_k\boldsymbol{\delta}_k'$ represent interaction. This summation is parametrised in singular value decomposition form, see below. When fitted by least squares

$$\begin{aligned}\widehat{\boldsymbol{\alpha}} &= \mathbf{X}\mathbf{1}/p - (np)^{-1}\mathbf{1}'\mathbf{X}\mathbf{1} \text{ and} \\ \widehat{\boldsymbol{\beta}} &= \mathbf{X}'\mathbf{1}/n - (np)^{-1}\mathbf{1}'\mathbf{X}\mathbf{1},\end{aligned}$$

where $\mathbf{1}$ is a vector of ones of conformable length. The biadditive term is estimated by first calculating the deviations from the main effects: $\mathbf{Z} = (\mathbf{I} - \mathbf{P})\mathbf{X}(\mathbf{I} - \mathbf{Q})$ where $\mathbf{P} = \mathbf{1}\mathbf{1}'/n$ and $\mathbf{Q} = \mathbf{1}\mathbf{1}'/p$. The biadditive terms are estimated from the singular value decomposition of $\mathbf{Z}$. It is well known that if the singular values σ_k are represented in nonincreasing order then the first k terms in the summation gives the best k-dimensional representation of the interaction. In many (most) practical situations we are concerned with the case of $k = 2$ and with visualization in two dimensions obtained by plotting the n rows of the $n \times 2$ matrix $(\mathbf{c}_1, \mathbf{c}_2)$ for row points and the p rows of the $p \times 2$ matrix $(\mathbf{d}_1, \mathbf{d}_2)$ for column points.

This gives us two sets of points, n and p representing the number of rows and columns, re-

spectively. By rotating one set of points by 90°, the new biplot may be used to visualise the approximation to the interactions.

6 Examples

To show how the area biplot can be applied in practice we give two examples. In the first example, we use a PCA area biplot whereas the second example concerns a CA area biplot.

6.1 PCA area biplot: Alcohol consumption in Japan

We apply the area biplot to 1994 data on alcohol consumption in Japan. The data are obtained from the website `http://mo161.soci.ous.ac.jp/@d/DoDStat/DataList/index.html`. For each of the 46 prefectures, the total consumption, in Japanese Yen, for 5 popular alcoholic beverages is recorded. The beverages are: beer, sake (a traditional Japanese rice wine), shouchuu (a traditional Japanese distilled alcoholic beverage), whiskey and wine. To eliminate the effect of the size of prefectures, the original figures are divided by the number of inhabitants per prefecture. These data are reported in Table 1 together with their z-scores.

To analyze the variations of alcohol we apply PCA to the z-scores and obtain a two-dimensional biplot. This two-dimensional solution accounts for approximately 68% of the total variation.

Figure 6 shows the area biplot where the prefectures are represented by solid dots and the alcoholic beverages by open dots. At the left bottom of the plot we find the prefectures Miyazaki and Kagoshima. These are prefectures located in the Southwestern part of Japan on the Kyushu island. It is clear that the triangles corresponding to the wine consumption have a surface close to zero for these prefectures. Also, the triangles formed with shouchuu and sake will be relatively large and comparable in size. For Kagoshima, the area corresponding to shouchuu consumption is 2.46 and the area for sake consumption is 2.75. However, the sign corresponding to the estimated inner products differs for these two beverages: it is positive for shouchuu and negative for sake. Hence, the estimated z-scores for shouchuu and sake consumption in Kagoshima, are 2.46 and -2.75 respectively. Similarly, we can obtain the estimated values for Miyazaki which are 2.53 and -2.56. Clearly, in these prefectures the estimated consumption of shouchuu is considerably larger than average whereas the estimated consumption of sake is considerably smaller than the average sake consumption. Note that these estimated values are in accordance with the z-scores in Table 1. The Kyushu island is particularly famous for its shouchuu production which explains the higher consumption for these prefectures. Moreover, according to the entry for Kagoshima in

Table 1: Average alcohol consumption per Japanese prefecture for five types of alcohol.

	Average per inhabitant					z-scores of average per inhabitant				
Prefecture	Sake	Shochu	Bear	Wine	Whisky	Sake		Bear	Wine	Whisky
Aichi	8.63	2.67	56.89	1.08	1.11	-0.63	-0.75	0.76	-0.06	-0.28
Akita	20.31	5.25	57.09	1.04	1.77	2.28	-0.07	0.79	-0.10	1.31
Aomori	11.71	7.80	56.36	1.20	2.12	0.14	0.60	0.69	0.11	2.17
Chiba	7.15	5.56	40.52	1.10	1.02	-1.00	0.01	-1.52	-0.03	-0.50
Ehime	10.99	4.08	52.73	0.67	0.94	-0.04	-0.38	0.18	-0.60	-0.69
Fukui	13.31	1.57	56.51	0.62	1.07	0.54	-1.03	0.71	-0.66	-0.39
Fukuoka	8.55	7.10	52.87	1.05	1.21	-0.65	0.41	0.20	-0.09	-0.03
Fukushima	14.58	5.25	47.07	0.89	1.79	0.85	-0.07	-0.61	-0.31	1.37
Gifu	11.65	2.09	49.13	0.73	0.87	0.12	-0.90	-0.32	-0.52	-0.86
Gunma	9.63	7.80	44.02	1.07	0.92	-0.38	0.60	-1.03	-0.06	-0.76
Hiroshima	10.97	4.68	59.14	1.09	1.25	-0.05	-0.22	1.08	-0.03	0.07
Hokkaido	8.18	8.91	55.48	1.85	1.72	-0.74	0.89	0.57	0.95	1.19
Hyogo	9.23	2.35	51.86	0.99	1.21	-0.48	-0.83	0.06	-0.17	-0.05
Ibaragi	9.88	5.06	42.50	0.81	1.01	-0.32	-0.12	-1.25	-0.40	-0.53
Ishikawa	15.62	1.74	58.01	0.89	1.19	1.11	-0.99	0.92	-0.31	-0.10
Iwate	12.09	7.22	47.88	1.03	1.32	0.23	0.44	-0.49	-0.12	0.23
Kagawa	10.91	2.92	51.04	0.81	1.05	-0.06	-0.68	-0.05	-0.41	-0.44
Kagoshima	1.36	18.42	44.68	0.54	0.47	-2.44	3.38	-0.94	-0.76	-1.85
Kanagawa	7.55	5.96	46.92	1.51	1.65	-0.90	0.12	-0.63	0.51	1.04
Kochi	13.95	4.02	66.53	0.76	1.37	0.70	-0.39	2.11	-0.47	0.35
Kumamoto	5.01	11.76	50.60	0.74	0.68	-1.53	1.63	-0.11	-0.50	-1.34
Kyoto	10.24	2.01	56.96	1.33	1.27	-0.23	-0.92	0.77	0.27	0.10
Mie	10.79	2.63	45.65	0.68	1.00	-0.09	-0.76	-0.80	-0.58	-0.56
Miyagi	11.78	4.98	46.44	1.19	2.13	0.16	-0.14	-0.69	0.10	2.21
Miyazaki	2.79	19.93	49.66	0.81	0.64	-2.09	3.78	-0.24	-0.41	-1.42
Nagano	14.30	5.58	49.29	1.80	1.12	0.78	0.01	-0.30	0.89	-0.27
Nagasaki	9.25	7.93	50.44	0.77	1.09	-0.47	0.63	-0.14	-0.46	-0.32
Nara	7.81	1.65	43.69	0.63	0.71	-0.83	-1.02	-1.08	-0.64	-1.27
Niigata	24.43	2.83	60.14	1.17	1.50	3.31	-0.70	1.22	0.07	0.67
Oita	8.30	11.76	48.49	0.86	1.61	-0.71	1.64	-0.41	-0.34	0.93
Okayama	11.31	3.73	46.02	0.77	1.03	0.04	-0.47	-0.75	-0.46	-0.48
Osaka	9.49	2.92	71.21	1.52	1.69	-0.42	-0.68	2.76	0.53	1.14
Saga	12.89	5.07	49.46	0.60	0.81	0.43	-0.12	-0.27	-0.68	-1.02
Saitama	7.20	5.94	41.09	1.08	1.04	-0.99	0.11	-1.44	-0.05	-0.45
Shiga	10.84	1.96	42.15	0.71	0.87	-0.08	-0.93	-1.29	-0.54	-0.87
Shimane	16.96	5.46	51.41	1.64	0.99	1.45	-0.02	-0.00	0.68	-0.58
Shizuoka	10.25	6.00	47.66	1.16	1.14	-0.23	0.13	-0.52	0.06	-0.22
Tochigi	9.64	5.33	42.29	0.78	1.00	-0.38	-0.05	-1.27	-0.45	-0.56
Tokushima	10.88	2.76	47.94	0.61	0.92	-0.07	-0.72	-0.48	-0.67	-0.74
Tokyo	10.59	7.16	71.14	3.49	2.39	-0.14	0.43	2.75	3.10	2.83
Tottori	15.32	3.19	53.31	0.78	1.20	1.04	-0.61	0.26	-0.45	-0.06
Toyama	16.20	1.81	54.47	0.80	1.15	1.26	-0.97	0.43	-0.43	-0.18
Wakayama	13.45	2.17	59.04	0.69	1.22	0.57	-0.88	1.06	-0.56	-0.03
Yamagata	16.53	4.88	49.14	1.37	1.98	1.34	-0.17	-0.32	0.33	1.84
Yamaguchi	10.92	6.50	57.23	0.75	1.25	-0.06	0.26	0.81	-0.48	0.06
Yamanashi	9.73	7.50	42.97	5.08	0.93	-0.35	0.52	-1.18	5.19	-0.71

the Wikipedia, Kagoshima is the only prefecture in Japan that does not produce sake. Opposite Kagoshima and Miyazaki, we find Niigata and Akita. These prefectures are famous for their sake. This reflects itself in the plot as the triangle with the origin and sake is relatively large, and the sign of the estimated inner product is positive.

On the top left of the plot we find Tokyo prefecture which comprises the Tokyo metropolitan area. Making the triangles we immediately see that Tokyo is the only prefecture where the estimated consumption is higher than average for all beverages. In particular, whiskey (with an

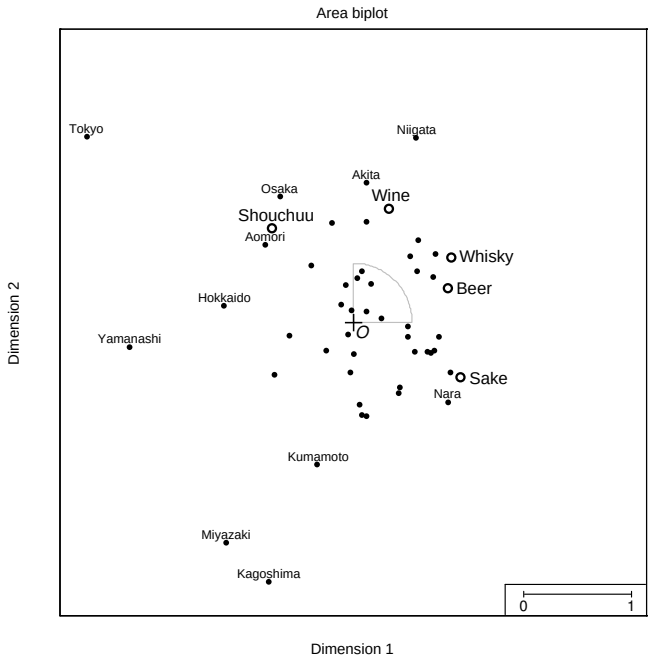

Figure 6: Area biplot for average expenditure on alcohol consumption in 46 Japanese prefectures (solid dots, some with label) on 5 types of alcohol (open dots and labels).

estimated z-score equal to 3.01), wine (3.09) and beer (2.27) show much higher levels of estimated consumption than in the rest of the country. Again, comparing the estimated values to the data values in Table 1, we see that the approximation is quite reasonable.

Finally, Yamanashi prefecture, in the plot located on the far left side, appears to have a somewhat unique position that indicates a relatively high estimated consumption of wine. The estimated value is 2.06. The reason for this may be the fact that Yamanashi, which is located directly to the west of Tokyo and in which mount Fuji is located, is the country's top producer of wine. Note that, the z-score for wine consumption in Yamanashi is 5.19. Hence, the estimated value underestimates this score quite significantly.

We can interpret the levels of consumption of the alcoholic beverages for other prefectures in a similar fashion.

6.2 CA area biplot: Seat distribution in the European parliament (hypothetical data)

In Table 2, a hypothetical data set is presented concerning the distribution of seats in the European parliament (Borg & Groenen, 2005). As this table has only three columns, the maximum dimensionality is 2. Figure 7 shows the corresponding exact area biplot using the symmetric stan-

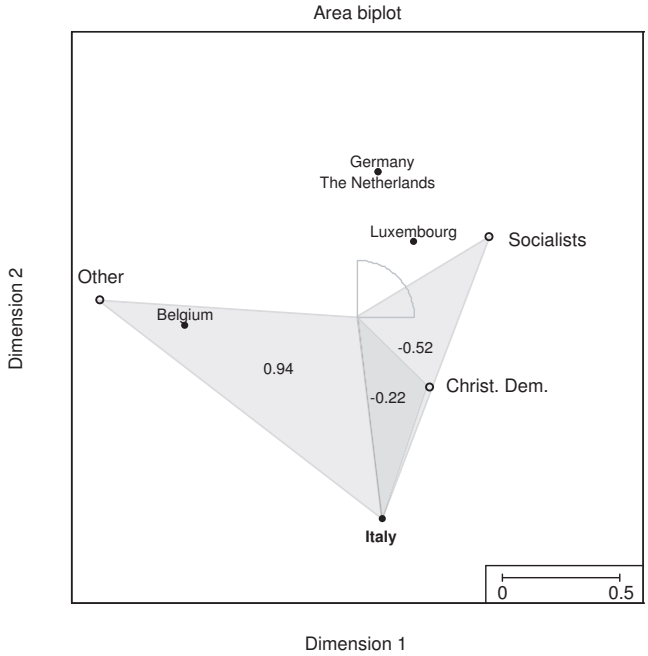

Figure 7: Area biplot for correspondence analysis.

dardization of $\alpha = .5$. Note that the relative distributions of the seats over the political factions are the same for Germany and The Netherlands, so that these points are superimposed in the biplot as is required by CA. The last three columns of Table 2 give the relative deviation from independence that is estimated by the area in the area biplot. The table shows that Italy deviates most from independence. This observation is reflected in the shaded areas of Italy with each of the political factions: the positive area of .94 for the Other faction indicates that Italy is relatively more represented by the Other faction and less by the socialists (a negative area of $-.52$) and the Christian democrats (area is $-.22$).

Table 2: A contingency table of the distributions of seats by country and political faction (hypothetical data) (Borg & Groenen, 2005).

	Frequency f_{ij}				$(f_{ij} - e_{ij})/e_{ij}$		
Country	CD	Soc	Other	Total	CD	Soc	Other
Belgium	8	9	7	24	-0.22	0.24	0.09
Germany	39	30	6	75	0.21	0.32	-0.70
Italy	25	11	39	75	-0.22	-0.52	0.94
Luxembourg	3	2	1	6	0.16	0.10	-0.38
The Netherlands	13	10	2	25	0.21	0.32	-0.70
Total	88	62	55	205			

7 Conclusion and Discussion

The area biplot can be used for any model that has a biadditive component. Here it is shown how the area biplot can be used in principal components analysis, correspondence analysis, and biadditive models. In all three cases, the area spanned by a triangle between the origin, row point i, and column point j estimates the data value ij. In PCA, this happens directly, in CA the estimate approximates the contingency ratio minus one, and in the biadditive model, it estimates the deviation from the main effects.

The idea of using area to estimate data values comes up naturally in the singular value decomposition of skew-symmetric matrices (Gower, 1977; Constantine & Gower, 1978). There, the singular values come in equal pairs with corresponding left and right singular vectors. Instead of plotting both right and left singular vectors and interpreting the inner products, only one pair (hedron) of dimensions of the left (or right) singular vectors is used to plot n points. Similar geometry then applies to that discussed in Section 3. The area of triangles generated by the ith and jth points and the origin estimation the corresponding data value in the skew symmetric matrix.

PCA is often explained as the eigendecomposition of a correlation matrix. The resulting plot of a single et of points representing the component loadings of the variables gives inner product estimates of the correlation between pairs of variables. In this case, too, the area representation in two dimensions is available but to obtain the second set of points, the first set has to be rotated through 90°. Thus, the number of points to be represented is doubled, which could be seen as too high a price to pay.

We have seen that area biplots may be used to represent any number of dimensions, odd or even. Hence, provided some convenient way is available for evaluating the two-dimensional inner-products, as in the area biplot, a visual approximation of several dimensions is available. The areas in successive pairs of dimensions have only to be added, requiring care in the compatible scaling of linked two-dimensional plots. Of course, as a visual process, this loses credibility as the number of dimensions increases. These remarks draw attention to the common practise of plotting all pairs of dimensions which seems unnecessary when inner products are the vehicle of interpretation. When distance interpretations are required, our approach is not available but mentally combining two or more distance maps requires distances in each pair of dimensions to be squared, added and then "square-rooted" which seems to us to be beyond reasonable demands.

References

Borg, I., & Groenen, P. J. F. (2005). *Modern multidimensional scaling: Theory and applications (2nd edition).* New York: Springer.

Constantine, A. G., & Gower, J. C. (1978). Graphic representations of asymmetric matrices. *Applied Statistics*, *27*, 297–304.

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

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

Gower, J. C. (1977). The analysis of asymmetry and orthogonality. In F. Brodeau, G. Romier, & B. Van Cutsem (Eds.), *Recent developments in statistics* (pp. 109–123). Amsterdam: North-Holland.

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