DISTANCE-BASED TRANSFORMATIONS OF BIPLOTS

JAN DE LEEUW

1. Introduction

In principal component analysis and related techniques we approximate (in the least squares sense) an \( n \times m \) matrix \( F \) by an \( n \times m \) matrix \( G \) which satisfies \( \text{rank}(G) \leq p \), where \( p < \min(n,m) \). Or, equivalently, we want to find an \( n \times p \) matrix \( X \) and an \( m \times p \) matrix \( Y \) such that \( G = XY' \) approximates \( F \) as closely as possible. The rows of \( X \) and \( Y \) are then often used in graphical displays. In particular, biplots [Gower and Hand, 1996] represent \( X \) and \( Y \) jointly as \( n + m \) points in Euclidean \( p \) space.

If formulated in this way, there is an important form of indeterminacy in this approximation problem. If \( R \) of order \( p \) is nonsingular, then we can define \( \tilde{X} = XR \) and \( \tilde{Y} = YR^{-T} \) and we have \( \tilde{X}\tilde{Y}' = XY' \), where \( A^{-T} \) is the transpose of the inverse (or the inverse of the transpose). Thus \( \tilde{X} \) and \( \tilde{Y} \) give exactly the same approximation, but plotting them may give quite different results, depending on \( R \). To give a simple example, we can choose \( R \) scalar, and make \( \tilde{X} \) arbitrarily small and \( \tilde{Y} \) arbitrarily big. In particular for biplots, which are often interpreted in terms of distances between the points, the indeterminacy is a nuisance and can lead to unattractive representations.

In this note we choose \( R \) in such a way that the distances, more specifically the squared Euclidean distances, between selected rows of \( \tilde{X} \) and \( \tilde{Y} \) are small. This takes care of both the relative scaling of the two clouds of points, as well as rotating them to some form of conformance.

Date: April 14, 2006.
2000 Mathematics Subject Classification. 62H25.
Key words and phrases. Multivariate Analysis.
The squared distance between rows \( i \) and \( j \) of the \( n + m \) matrix 
\[
Z = \begin{bmatrix} XR \\ YR^{-T} \end{bmatrix}
\]
can be written as 
\[
d^2_{ij}(R) = (e_i - e_j)'C(e_i - e_j) = \text{tr} \ C A_{ij}.
\]
Here the \( e_i \) are unit vectors (columns of the identity matrix) and we define 
\[
C = \begin{bmatrix} XX' & XY' \\ YX' & YY'^{-1} \end{bmatrix},
\]
as well as \( S = RR' \) and \( A_{ij} = (e_i - e_j)(e_i - e_j)' \).

Thus summing over a selected subset \( I \) of squared distances leads to a loss function of the form 
\[
\lambda(S) = \sum_{(i,j) \in I} d^2_{ij}(S) = \text{tr} \ S X'A_{11}X + \text{tr} \ S^{-1}Y'A_{22}Y
\]
where \( A_{11} \) and \( A_{22} \) are the two principal submatrices of 
\[
A = \sum_{(i,j) \in I} A_{ij}.
\]

If we minimize the sum of squares of all \( nm \) distances between the \( n \) points in \( X \) and the \( m \) points in \( Y \), for example, we find \( A_{11} = mI \) and \( A_{22} = nI \). If \( n = m \) and we want to minimize the sum of the \( n \) squared distances between the corresponding points \( x_i \) and \( y_i \) then \( A_{11} = A_{22} = I \).

3. Problem Solution

Let us minimize \( \lambda(S) = \text{tr} \ S P + \text{tr} \ S^{-1}Q \), where both \( P \) and \( Q \) are positive definite. If \( P \) and/or \( Q \) are singular, the more general results of De Leeuw [1982] must be used, but in most applications we have in mind nonsingularity is guaranteed.

The stationary equations for the problem of minimizing \( \lambda(S) \) are

\[
P = S^{-1}QS^{-1},
\]
which we have to solve for a positive definite $S$. We can use the symmetric square root to rewrite Equation (1) as

$$I = P^{-\frac{1}{2}} S^{-1} P^{-\frac{1}{2}} \left[ P^{-\frac{1}{2}} Q P^{-\frac{1}{2}} \right] P^{-\frac{1}{2}} S^{-1} P^{-\frac{1}{2}},$$

from which

$$P^{-\frac{1}{2}} S^{-1} P^{-\frac{1}{2}} = \left[ P^{-\frac{1}{2}} Q P^{-\frac{1}{2}} \right]^{-\frac{1}{2}},$$

and thus

$$S^{-1} = P^\frac{1}{2} \left[ P^{-\frac{1}{2}} Q P^{-\frac{1}{2}} \right]^{-\frac{1}{2}} P^\frac{1}{2},$$

and

$$S = P^{-\frac{1}{2}} \left[ P^{-\frac{1}{2}} Q P^{-\frac{1}{2}} \right]^\frac{1}{2} P^{-\frac{1}{2}}.$$

If we want to minimize the sum of squares of all distances between the points in $X$ and those in $Y$ we have seen that $A_{11} = mL$ and $A_{22} = nI$. In many forms of principal component analysis $X$ is chosen such that $X'X = I$, and thus $P = mL$. In that case, from (5),

$$S = \sqrt{n} \sqrt{m} (Y'Y)^{\frac{1}{2}}.$$

If $Y = L\Lambda L'$ is an eigen-decomposition of $Y$, we can choose

$$R = \left[ \frac{n}{m} \right]^{\frac{1}{2}} \Lambda L^\frac{1}{2},$$

$$R^{-T} = \left[ \frac{m}{n} \right]^{\frac{1}{2}} \Lambda L^{-\frac{1}{2}}.$$

4. Example

To illustrate the problem, consider the following output from the scalAssoc() program [De Leeuw 2006]. These are 20 votes of 100 US senators. Each vote is presented by a plus ("aye") point and a minus ("nay") point, and the technique jointly scales senators and votes in such a way that senators are closest to the vote points they endorse. Or, equivalently, senators voting “aye” must be separated by a straight line from senators voting “nay”. In Figure 1 all senators are clumped around the origin, and this makes it impossible to read and interpret the plot.
Now let us apply the scaling outlines in this paper. Figure 2 gives the results, which are clearly much more satisfactory.
References


Department of Statistics, University of California, Los Angeles, CA 90095-1554

E-mail address, Jan de Leeuw: deleeuw@stat.ucla.edu

URL, Jan de Leeuw: http://gifi.stat.ucla.edu