Topics in correspondence analysis
van de Velden, M.

Citation for published version (APA):
Chapter 5

Correspondence analysis of more than two categorical variables

5.1 Introduction

Correspondence analysis as described in Chapter 2 is concerned with the analysis of a contingency matrix, i.e. the cross-tabulation of frequencies of co-occurrences between two categorical variables. Often the situation occurs that we have data on more than two categorical variables. For the analysis of such data extensions to the usual correspondence analysis approach exist. In this chapter we will present two of these approaches: multiple correspondence analysis and joint correspondence analysis. Multiple correspondence analysis and joint correspondence analysis are closely related to, respectively, principal component analysis and factor analysis. We will treat the methods in relation to these well-known multivariate analysis methods.

An important issue in the analysis of data on more than two categorical variables is the coding of the data. In Chapter 3, section 3.3.3, we introduced a so-called indicator matrix. In the analysis of more than two categorical variables such an indicator matrix, sometimes also referred to as a response pattern matrix, e.g. Nishisato (1994), is of crucial importance. Typically the rows of the indicator matrix denote the observations whereas the columns represent the categories for each variable. For each observation a one is inserted in the columns representing the observed categories. All remaining elements are
zero. This indicator matrix will be the starting point for both multiple correspondence analysis and joint correspondence analysis.

The name multiple correspondence analysis can be misleading as it suggests to represent a method that is different from correspondence analysis. This, however, is not the case. In fact, multiple correspondence analysis is usually defined as the correspondence analysis of an indicator matrix. The geometrical and statistical concepts encountered in the correspondence analysis of a contingency matrix, i.e. the chi-squared distance and the chi-squared statistic, are difficult to justify in multiple correspondence analysis, see e.g. Greenacre (1990, 1991). Mathematically equivalent methods such as homogeneity analysis, reciprocal averaging and dual scaling are not explicitly concerned with these geometrical concepts. Therefore, the rationales underlying these methods may provide a justification for the use of multiple correspondence analysis. Greenacre (1991) discusses the interpretational problems of multiple correspondence analysis as well as the practical interpretation of homogeneity analysis in more detail.

In joint correspondence analysis the aim is to provide a least-squares approximation of all contingency matrices that can be constructed from the data. If there are \( q \) categorical variables we can make \( \frac{1}{2}q(q-1) \) contingency matrices and these contingency matrices are, simultaneously, approximated in a least-squares sense. If we have two categorical variables there is only one contingency matrix and the two methods are in such a case equivalent. In order to obtain a joint correspondence analysis solution an iterative procedure was suggested by Greenacre (1988). Greenacre also mentioned the relationship between joint correspondence analysis and factor analysis. This relationship was further formalized by Boik (1996) who also provided an alternative algorithm to Greenacre's (1988) original algorithm.

In this chapter the focus will be on some mathematical properties of multiple and joint correspondence analysis. In section 5.2 we introduce notation and summarize some important properties of the indicator matrix. The relationship between multiple correspondence analysis and correspondence analysis when there are only two variables will be formalized in section 5.3.1. In such a case the data are, except for a difference in the format, the same. Therefore, the outcomes should preferably also be similar. In fact, as
was already mentioned in Chapter 3, section 3.3.3, the analysis of the indicator matrix $Z$ and the contingency matrix $F$ are closely related when there are only two variables.

In section 5.3.2 the close relationship between multiple correspondence analysis and principal component analysis will be clarified. This relationship is known and in a recent textbook by Gower and Hand (1996), multiple correspondence analysis is introduced as a form of principal component analysis.

In section 5.4, we will introduce joint correspondence analysis as a specific type of factor analysis. For this purpose we will introduce the typical factor analysis model together with a generalization that is sometimes referred to as multiple battery factor analysis or multigroup factor analysis. In section 5.4.2 we will derive an iterative procedure for obtaining a solution in the multigroup factor analysis model. This iterative procedure is essential in joint correspondence analysis.

Let us first start with the introduction of some specific properties of the data.

5.2 Notation

Let $Z$ be an $n \times p$ ($n > p$) indicator matrix that can be partitioned as

$$Z = (Z_1, Z_2, \ldots, Z_q), \quad (5.1)$$

where $Z_i$ is an $n \times p_i$ ($i = 1 \ldots q$) matrix with exactly one row element equal to one and all other elements zero, and $\sum_{i=1}^q p_i = p$.

We have the following relations,

$$Z 1_p = q 1_n, \quad (5.2)$$

and

$$1_n' Z 1_p = nq. \quad (5.3)$$

Let

$$z \equiv Z 1_n = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_q \end{pmatrix}, \quad (5.4)$$
where $z_i$ is a $p_i \times 1$ vector, for $i = 1 \ldots q$, and introduce a $p \times p$ diagonal matrix $D_z$ satisfying

$$D_z 1_p = z. \quad (5.5)$$

Note that $Z_i'Z_i = diag\{z_i\}$, and $Z_i'Z_j$ ($i \neq j$) is the contingency matrix with as its elements the frequencies of co-occurrences between the $i$th and $j$th variable.

Define an $n \times p$ matrix $Z^*$ as

$$Z^* = \frac{1}{\sqrt{q}} D_z^{-\frac{1}{2}}. \quad (5.6)$$

Thus, $Z^*$ is a rescaled version of the original indicator matrix. The non-zero elements of $Z^*$ indicate the relative (with respect to the category totals) number of subjects falling in a category. This means that categories which have only few observations corresponding to it receive higher weights than categories that have many observations.

The constant $\frac{1}{\sqrt{q}}$ in (5.6) is added for convenience.

We define the sample covariance matrix for $Z^*$ as

$$S_{Z^*} = \frac{1}{n} Z^* M Z^*, \quad (5.7)$$

where $M$ is the usual centering matrix of the appropriate order, i.e.

$$M = I_n - \frac{1}{n} 1_n 1_n'.$$

Inserting (5.6) in (5.7) yields

$$S_{Z^*} = \frac{1}{nq} D_z^{-\frac{1}{2}} Z^* M Z^* D_z^{-\frac{1}{2}} \quad (5.8)$$

$$= \frac{1}{nq} D_z^{-\frac{1}{2}} \left( Z - \frac{1}{n} 1_z' \right)' \left( Z - \frac{1}{n} 1_z' \right) D_z^{-\frac{1}{2}}$$

$$= \frac{1}{nq} D_z^{-\frac{1}{2}} \left( Z Z' - \frac{1}{n} z z' \right) D_z^{-\frac{1}{2}},$$

where we used (5.4).

Using the partitionings of (5.1) and (5.4) we can write

$$S_{Z^*} = \frac{1}{nq} D_z^{-\frac{1}{2}} \begin{pmatrix}
Z_1'Z_1 - \frac{1}{n} z_1 z_1' & Z_1'Z_2 - \frac{1}{n} z_1 z_2' & \cdots & Z_1'Z_q - \frac{1}{n} z_1 z_q' \\
Z_2'Z_1 - \frac{1}{n} z_2 z_1' & Z_2'Z_2 - \frac{1}{n} z_2 z_2' & \cdots & Z_2'Z_q - \frac{1}{n} z_2 z_q' \\
\vdots & \vdots & \ddots & \vdots \\
Z_q'Z_1 - \frac{1}{n} z_q z_1' & Z_q'Z_2 - \frac{1}{n} z_q z_2' & \cdots & Z_q'Z_q - \frac{1}{n} z_q z_q'
\end{pmatrix} D_z^{-\frac{1}{2}}. \quad (5.9)$$
Now, as $Z'_i Z_j$ $(i \neq j)$ is a contingency matrix with as its elements the frequencies of co-occurrences between the categories of the $i$th and $j$th variable, we see that the off-diagonal blocks of $S_z$ are weighted contingency matrices in deviations from the independence model. In simple correspondence analysis we only have two variables and the off-diagonal block matrix is subjected to two-way component analysis. In the next section we will compare the analysis of a two-variable indicator matrix with that of the contingency matrix.

5.3 Multiple correspondence analysis

As mentioned in the introduction of this chapter, multiple correspondence analysis is in fact correspondence analysis applied to the indicator matrix $Z$. As $Z$ is nonnegative the equations from Chapter 2, in particular those of section 2.3, can be used with $Z$ substituted for $F$.

If we have data on two categorical variables we can either apply correspondence analysis to the contingency matrix $F$ or to the indicator matrix $Z$. As the data are the same, the outcomes of the approaches should preferably also be the same. As we will show below this is only true to a certain extent. Note that this relationship is well-known and alternative descriptions can be found in Greenacre (1984), Lebart et al. (1984) and Gifi (1990).

5.3.1 Multiple correspondence analysis of two categorical variables

In the two variable case we have an $n \times p$ indicator matrix $Z = (Z_1, Z_2)$, where $Z_1$ is of the order $n \times p_1$ and $Z_2$ is of the order $n \times p_2$. Define $F$ as the $p_1 \times p_2$ corresponding contingency matrix, i.e.

$$F = Z'_1 Z_2.$$  \hspace{1cm} (5.10)

Also, in accordance with the usual definitions let

$$r \equiv F 1_{p_2} = Z'_1 Z_2 1_{p_2} = Z'_1 1_n = Z'_1 Z_1 1_{p_1},$$
Chapter 5. Correspondence analysis of more than two categorical variables

and

\[ c \equiv F'1_{p_1} = Z'_2Z1_{p_1} = Z'_21_n = Z'_2Z21_{p_2}. \]

Then, due to the special structure of \( Z \), the corresponding scaling matrices \( D_r \) and \( D_c \) can be expressed as

\[ D_r = Z'_1Z_1, \quad (5.11) \]

and

\[ D_c = Z'_2Z_2. \quad (5.12) \]

Now, consider the correspondence analysis of the indicator matrix \( Z \), where the diagonal scaling matrices, say \( D^Z_r \) and \( D^Z_c \), are defined as

\[ D^Z_r \equiv diag\{Z1_{p_1}\} = 2I_n \quad (5.13) \]

and

\[ D^Z_c \equiv diag\{Z'_11_n\} = diag\{(Z_1, Z_2)'1_n\} = \begin{pmatrix} D_r \\ D_c \end{pmatrix}. \quad (5.14) \]

Instead of considering the deviations from independence we will here, for convenience, consider the analysis of \( F \) (and \( Z \)). (It was shown in Chapter 3 that except for the retrieval of the so-called trivial solution in the analysis of \( F \) the two analyses are equivalent). The eigenequations for the standard coordinates, cf. (2.36) and (2.37), become

\[ D^{-1}_rFD^{-1}_cF'X = X\Lambda, \quad (5.15) \]

\[ D^{-1}_cF'D^{-1}_rFY = Y\Lambda, \quad (5.16) \]

where \( \Lambda \) is a \( k \times k \) (where \( k \) is the chosen rank of the approximation; \( k \leq r \leq \min(p_1, p_2) \)) diagonal matrix of eigenvalues, and \( X \) and \( Y \) are matrices (of orders \( p_1 \times k \) and \( p_2 \times k \) respectively) of eigenvectors standardized as \( X'D_rX = Y'D_cY = nk_k \). One apparent
difference in the two approaches is, of course, that in the analysis of $Z$ we not only obtain a solution for the two-variables but we also obtain a solution for the $n$ subjects. In the analysis of a contingency matrix such a solution does not appear. With regard to the solution for the (two) variables, i.e. the columns of $Z$, we have the eigenequation

$$(D_r^Z)^{-1} Z' (D_c^Z)^{-1} Z Y^Z = Y^Z \Gamma,$$  \hspace{1cm} (5.17)

where $\Gamma$ is a $k \times k$ diagonal matrix of eigenvalues and $Y^Z$ is a $p \times k$ matrix of corresponding eigenvectors standardized as $Y'^Z D_c^Z Y^Z = n I_k$. Like before $k$ is the (chosen) rank of the approximation. Writing $Z = (Z_1, Z_2)$ yields, after insertion of (5.13) and (5.14) in (5.17),

$$\frac{1}{2} \left( \begin{array}{cc} D_r^{-1} & \\
D_c^{-1} & 
\end{array} \right) \left( \begin{array}{c} Z_1 \\
Z_2 
\end{array} \right) (Z_1, Z_2) Y^Z = Y^Z \Gamma$$

$$\frac{1}{2} \left( \begin{array}{cc} D_r^{-1} & \\
D_c^{-1} & 
\end{array} \right) \left( \begin{array}{cc} Z_1' Z_1 & Z_1' Z_2 \\
Z_2' Z_1 & Z_2' Z_2 
\end{array} \right) Y^Z = Y^Z \Gamma$$

$$\frac{1}{2} \left( \begin{array}{cc} D_r^{-1} Z_1' Z_1 & D_r^{-1} Z_1' Z_2 \\
D_c^{-1} Z_2' Z_1 & D_c^{-1} Z_2' Z_2 
\end{array} \right) Y^Z = Y^Z \Gamma$$

$$\left( \begin{array}{cc} I_{p_1} & D_r^{-1} F \\
D_c^{-1} F' & I_{p_2} 
\end{array} \right) Y^Z = 2Y^Z \Gamma,$$  \hspace{1cm} (5.18)

where the relationships (5.11) and (5.12) were used.

From (5.18) it follows that the eigenvalues $\gamma_i$ are solutions of the equation

$$\left| \begin{array}{cc} (1 - 2\gamma) I_{p_1} & D_r^{-1} F \\
D_c^{-1} F' & (1 - 2\gamma) I_{p_2} 
\end{array} \right| = 0.$$  \hspace{1cm} (5.19)

Assuming that $2\gamma \neq 1$, and applying a well-known rule for determinants of partitioned matrices yields

$$\left| (1 - 2\gamma) I_{p_1} \right| \times \left| (1 - 2\gamma) I_{p_2} - (1 - 2\gamma)^{-1} D_c^{-1} F' D_r^{-1} F \right| = 0$$

$$\rightarrow \left| D_c^{-1} F' D_r^{-1} F - (1 - 2\gamma)^2 I_{p_2} \right| = 0 \rightarrow (1 - 2\gamma)^2 = \lambda.$$
Hence, the eigenvalues \( \lambda_i \) (for \( i = 1 \ldots k \)) obtained after applying correspondence analysis to the contingency matrix \( F \) can be calculated from the eigenvalues \( \gamma_i \) obtained in the correspondence analysis of a two-variable indicator matrix \( Z \).

Furthermore, if we partition \( Y^Z \) as \( \begin{pmatrix} Y_1^Z \\ Y_2^Z \end{pmatrix} \) where \( Y_1^Z \) is a \( p_1 \times k \) and \( Y_2^Z \) is a \( p_2 \times k \) matrix, then (5.18) yields the equations

\[
Y_1^Z + D_r^{-1}F Y_2^Z = 2Y_1^Z \Gamma,
\]

and

\[
Y_2^Z + D_c^{-1}F' Y_1^Z = 2Y_2^Z \Gamma.
\]

We can rewrite these as

\[
D_r^{-1}FY_2^Z = Y_1^Z (2\Gamma - I_k),
\]

and

\[
D_c^{-1}F'Y_1^Z = Y_2^Z (2\Gamma - I_k).
\]

(5.20)

Assuming that, for \( i = 1 \ldots k \), \( 2\gamma_i \neq 1 \), we have

\[
Y_1^Z = D_r^{-1}FY_2^Z (2\Gamma - I_k)^{-1}.
\]

Inserting this expression in (5.20) yields the eigenequation

\[
D_c^{-1}F'D_r^{-1}FY_2^Z = Y_2^Z (2\Gamma - I_k)^2.
\]

(5.21)

In a similar way we obtain as eigenequation for \( Y_1^Z \)

\[
D_r^{-1}FD_c^{-1}F'Y_1^Z = Y_1^Z (2\Gamma - I_k)^2.
\]

(5.22)

Then, as

\[
Y^Z'D_c^2Y^Z = \begin{pmatrix} Y_1^Z'D_rY_1^Z \\ Y_2^Z'D_cY_2^Z \end{pmatrix} = nI_{2k},
\]

and assuming that there are no multiple eigenvalues, comparison of (5.15) and (5.16) with (5.21) and (5.22) immediately shows that the matrix of standard column coordinates \( Y^Z \)
obtained in the analysis of Z, is equivalent to the stacked matrix of row and column standard coordinates obtained in the analysis of F, i.e.

\[ Y^Z = \begin{pmatrix} X \\ Y \end{pmatrix} \]

The principal coordinates obtained in the two analyses are not the same as the matrices of singular values are not the same in the two analyses. Furthermore, the inertias obtained in the analysis of Z are not equal to the inertias obtained in the analysis of F. The quality of the multiple correspondence analysis approximation can be assessed in a similar way as described in Chapter 2, section 2.3, with \( \Gamma \) substituted for \( \Lambda \). An important problem in multiple correspondence analysis concerns the low percentages of explained inertia. Greenacre (1990) describes this problem in more detail and provides some alternative measures for the amount of explained inertia in multiple correspondence analysis.

5.3.2 Multiple correspondence analysis and principal component analysis

In principal component analysis one obtains linear combinations of the original variables such that these linear combinations account for as much variance as possible. In Chapter 2 it was already noted that principal component analysis based on the sample variance is equivalent to one-mode component analysis.

The matrix of \( k \) \((k \leq p)\) principal component loadings in the analysis of \( Z^* \) and corresponding sample covariance matrix \( S_{z^*} \), can be obtained from the eigenequation

\[ nS_{z^*}L = LA, \] (5.23)

where the \( p \times k \) matrix of eigenvectors \( L \) is standardized as

\[ L' L = I_k, \]

and \( \Lambda \) is a \( k \times k \) diagonal matrix with on the diagonal the \( k \) largest eigenvalues of \( nS_{z^*} \).

(For an exposition of principal component analysis see, for example, Joliffe, 1986).
If \( k < \text{rank}(S_{zz}) \) the matrix \( LAL' \) gives a least-squares approximation of \( nS_{zz} \). Hence, applying principal component analysis is equivalent to the least-squares approximation of \( n \) times the sample covariance matrix \( S_{zz} \).

In principal coordinate analysis, as introduced by Gower (1966), the aim is to obtain coordinates for the observations such that the distances, suitably defined, between the observations are best, in a least-squares sense, preserved. Principal coordinate analysis is related to (metric) multidimensional scaling in the following way. The \( n \) observations are considered as points in \( p \)-dimensional space. Upon calculating the Euclidean distances between all pairs of points a distance matrix is constructed. Then, by applying (metric) multidimensional scaling\(^1\) to this distance matrix we obtain a new configuration in a space of low dimensionality. This new configuration is 'closest' to the original configuration in a least-squares sense. It can be shown, e.g. Mardia et al. (1979) and Joliffe (1986), that the coordinates for the \( n \) points in the new configuration, i.e. the principal coordinates, are equal to the centered scores of the \( n \) objects on the first \( k \) principal components. Thus, the matrix of principal coordinates in the analysis of \( Z^* \) can be expressed as

\[
G_{pc} = MZ^*L, \tag{5.24}
\]

where the subscript \( pc \) is added to distinguish between the matrix of principal row coordinates obtained in correspondence analysis.

If the (metric) multidimensional scaling approximation has rank \( p \), i.e. the new configuration has the same rank as the original configuration, the principal coordinates are equivalent to the original coordinates up to an orthogonal rotation (see van de Velden et al., 1999, and Graffelman, 1999).

By employing the usual correspondence analysis equations we will show that the principal row coordinate matrix \( G \) obtained in multiple correspondence analysis is equivalent to the matrix of principal coordinates \( G_{pc} \) multiplied by \( \sqrt{n} \).

Substituting \( Z \) for \( F \) in the usual correspondence analysis equations yields as eigenequation for the standard column coordinates

\[
\frac{1}{q}D_z^{-1} \left( Z - \frac{1}{n}1z' \right)' \left( Z - \frac{1}{n}1z' \right) Y = Y\Lambda, \tag{5.25}
\]

\(^1\)For an exposition of metric multidimensional scaling see e.g. Mardia et al. (1979).
where
\[ Y'D_2Y = nqI_n. \]

Define
\[ Y^* = \frac{1}{\sqrt{nq}}D_2^{1/2}Y \]
so that
\[ Y^*Y^* = I_n, \]
and (5.25) can be expressed as
\[ \frac{1}{q}D_2^{-1/2}Z'MZD_2^{1/2}Y^* = Y^*\Lambda. \]  
(5.26)
(Note that, for convenience, we used the complete decomposition rather than the \( k \)-dimensional approximation.) Using (5.8) we can rewrite (5.26) as
\[ nS_\tau Y^* = Y^*\Lambda. \]  
(5.27)

Comparing (5.23) and (5.27) immediately shows that the matrix \( Y^* \) is equivalent to the loading matrix \( L \) obtained in the principal component analysis of \( Z^* \). Hence, the matrix of standard coordinates for the columns in the correspondence analysis of \( Z \), i.e. \( Y \), is related to the loading matrix \( L \) through a simple rescaling.

Rewrite transition formula (2.38) in the following fashion,
\[ G = D_\tau^{-1}FY = D_\tau^{-1}\left( F - \frac{1}{s}rc' \right)Y, \]
where we used (2.32) and (2.33). Then, substituting \( Z \) for \( F \) and using the appropriate expressions for \( D_\tau, r \) and \( c \), yields
\[ G = \frac{1}{q} \left( Z - \frac{1}{n}1z' \right)Y = \sqrt{\frac{n}{q}}MZD_2^{-1/2}Y^* = \sqrt{n}MZ^*Y^*. \]  
(5.28)
Comparison of (5.23) and (5.24) with (5.27) and (5.28) immediately shows the relationship between the two approaches, i.e.
\[ G = \sqrt{n}G_{pc}. \]

Note that in multiple correspondence analysis we also obtain a matrix of coordinates for the columns of \( Z \), i.e. the categories of the variables. In principal coordinate analysis this is not the case.
5.4 Joint correspondence analysis

One could argue that only the off-diagonal blocks of $S_{aa}$, which are equal to the matrices of relative frequencies in deviation from independence between all pairs of categorical variables, are of interest. Therefore, Greenacre (1988) proposes an alternative approach which he calls joint correspondence analysis. In joint correspondence analysis all weighted contingency matrices are approximated simultaneously. Greenacre observes that joint correspondence analysis is related to (principal) factor analysis in a similar way as multiple correspondence analysis is related to principal component analysis. In this section we will describe joint correspondence analysis as a form of factor analysis. As a result of this approach some interesting issues, not obvious in the treatment of Greenacre (1988), arise naturally.

Before proceeding, we will briefly introduce the factor analysis model as well as a generalization that is sometimes referred to as multigroup factor analysis, e.g. Basilevsky (1994), or multiple battery factor analysis, e.g. McDonald (1970) and Browne (1980).

5.4.1 Factor Analysis

The factor model can be expressed in the following way:

$$x - \mu = Af + u,$$

where $x$ is an observable $p \times 1$ random vector with mean $\mu$, $A$ a $p \times k$ ($k \leq p$) matrix of factor loadings, $f$ a $k \times 1$ vector of unobservable underlying common factors and $u$ denotes a $p \times 1$ vector of unobserved unique factors. Thus, the $p$ variables can be expressed as a linear combination of $k$ unobservable factors plus a variable specific factor.

In order to obtain estimates in the factor model the following assumptions are made:

$$E(f) = 0_k, \quad Var(f) = I_k,$$  \hspace{1cm} (5.30)

$$E(u) = 0_p, \quad Var(u) = D_\Phi,$$  \hspace{1cm} (5.31)
5.4. Joint correspondence analysis

where $D_\Phi$ is a $p \times p$ diagonal matrix, and

$$Cov(u, f) = 0_{p \times k} \tag{5.32}$$

Furthermore, the loading matrix $A$ can be interpreted as the covariance matrix of the random variable $x$ and the underlying factors $f$, i.e.

$$Cov(x, f) = Cov(Af + u + \mu, f) = ACov(f, f) = A.$$ 

The factor model is well known and there are several methods to find estimates for $A$ and $D_\Phi$, e.g. maximum likelihood (which requires additional assumptions on the distribution of the factors and errors) or least-squares. The least-squares approach, also called principal factor analysis, involves the estimation of the off-diagonal elements of $\Sigma$.

5.4.2 Multigroup factor analysis

If we have data where the $i$th, $i = 1, \ldots, q$, variable is represented by a $p_i \times 1$ vector, and $p = \sum_{i=1}^{q} p_i$, we can construct the following factor model:

$$x - \mu = Af + u,$$

where $x$ is an observable $p \times 1$ random vector with mean vector $\mu$, $A$ is a $p \times k$ (with rank $(A) = k \leq p$) matrix of so-called factor loadings, $f$ is a $k \times 1$ vector of unobservable underlying common factors and $u$ denotes a $p \times 1$ vector of unique factors. Moreover, the vectors $x$ and $u$ and the matrix $A$ can be partitioned in the following way:

$$x = \begin{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_q \end{pmatrix} \end{pmatrix}, \quad u = \begin{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_q \end{pmatrix} \end{pmatrix}, \quad A = \begin{pmatrix} \begin{pmatrix} A_1 \\ A_2 \\ \vdots \\ A_q \end{pmatrix} \end{pmatrix}, \tag{5.33}$$

where $x_i$ and $u_i$ are $p_i \times 1$ vectors, and $A_i$ is a $p_i \times k$ matrix, for $i = 1 \ldots q$. Thus, each variable vector $x_i$, $i = 1 \ldots q$, can be expressed as a linear combination of $k$ underlying
Chapter 5. Correspondence analysis of more than two categorical variables

Factors plus a variable specific vector of residuals. Analogously to assumptions (5.30) through (5.32) we introduce as assumptions for this multigroup factor model:

\[ E(f) = 0_k, \quad \text{Var}(f) = I_k, \]
\[ E(u) = 0_q, \quad \text{Var}(u) = \Phi_d, \]

where \( \Phi_d \) is block diagonal with diagonal blocks of order \( p_i \times p_i, i = 1 \ldots q \), and

\[ \text{Cov}(u, f) = 0_{p\times k}. \]

Using this model we obtain for the variance of \( x \):

\[ \Sigma = \text{Var}(x) = \text{Var}(Af + u) = AA' + \Phi_d. \]

If we partition the variance matrix \( \Sigma \) in a similar way as we partitioned \( x, u \) and \( A \) we see that

\[ \Sigma_{ii} = A_iA_i' + \Phi_{ii} \]

and

\[ \Sigma_{ij} = A_iA_j', \quad (i \neq j). \]

Note that if \( p_i = 1 \), for \( i = 1 \ldots q \), the multigroup factor model reduces to the common factor model described in section 5.4.1.

For the multigroup factor model described above our objective is to estimate a loading matrix \( A \) and a matrix \( \Phi_d \) such that the off-diagonal blocks of the variance matrix \( \Sigma \) (or— in practice—those of its sample counterpart \( S \)) are approximated in a least-squares sense. Hence, by applying factor analysis in this fashion we approximate the off-diagonal blocks of \( \Sigma \), rather than the off-diagonal elements of \( \Sigma \), such that the sum of squared deviations is minimal.

The off-diagonal blocks can be seen as matrices of covariances between the several categories of a particular variable with those of another variable, i.e.

\[ \Sigma_{ij} = \text{Cov}(x_i, x_j) \quad i \neq j. \]

The diagonal blocks, on the other hand, give the covariance structures "within" each variable. Thus, approximating the off-diagonal blocks of \( \Sigma \) can be seen as the approximation of the "between" variables covariances.
5.4. Joint correspondence analysis

Derivation of the principal factor analysis solution

Upon defining $E = S - AA'$ we can formulate as objective for the principal factor analysis based on the sample covariance matrix $S$

$$
\min_A \psi = \text{trace} E_u^2,
$$
(5.34)

where for a $p \times p$ matrix $X$, $X_u = X - X_d$ and, analogously to the definition of $\Phi_d$, $X_d$ is a block diagonal matrix with as $q$ diagonal blocks the $p_i \times p_i$ $(i = 1, \ldots, q)$ diagonal blocks of $X$. Hence, $E_u = E - E_d$.

Clearly $E$ is symmetric, i.e. $E = E'$. In order to solve this minimization problem we will proceed in a similar way as described by Satorra and Neudecker (1998). The only difference between their approach and the one described here is that in our analysis the off-diagonal blocks rather than the off-diagonal elements are approximated. It is not difficult to see that this only superficially affects the derivation. As an alternative to the derivation given here, one could obtain a so-called alternating least-squares algorithm, see Ten Berge (1993), by solving

$$
\min_{\Phi_d} \text{trace} \left( S - \left( AA' + \Phi_d \right) \right)^2
$$

for fixed $\Phi_d$, and

$$
\min_A \text{trace} \left( S - \left( AA' + \Phi_d \right) \right)^2
$$

for fixed $A$. However, in line with previous derivations in this thesis we will explicitly derive first-order conditions for the minimization problem (5.34) using matrix derivatives.

We can express the objective as

$$
\min_A \psi = \left( \text{vec} E_u' \right)' \text{vec} E_u = \left( \text{vec} E_u \right)' \text{vec} E_u,
$$

where $\text{vec} E_u$ denotes the vectorization of $E_u$, i.e. $\text{vec} E_u$ is a $p^2 \times 1$ vector obtained by stacking the $p$ columns of $E_u$. Upon defining an appropriate $p^2 \times p^2$ selection matrix $K_d$ we can write

$$
\text{vec} E_u = \text{vec} E - K_d \text{vec} E = (I - K_d) \text{vec} E,
$$
where the matrix \( K_d \) selects the block diagonal elements of \( E \) from \( \text{vec} \ E \). As 
\[
K_d K_d \text{vec} \ E = K_d \text{vec} \ E_d = \text{vec} \ E_d = K_d \text{vec} \ E 
\]
it follows that \( K_d \) is (symmetric) idempotent, hence \( I - K_d \) is symmetric idempotent. Consequently, we have 
\[
\min_{A} \psi = (\text{vec} \ E)' (I - K_d) \text{vec} \ E. 
\]
As 
\[
\text{vec} \ dE = \text{vec} \ d \left( S - AA' \right) 
\]
we can express the differential of \( \psi \) as 
\[
d\psi = 2(\text{vec} \ E)' (I - K_d) \text{vec} \ dE 
\]
\[
= -2(\text{vec} \ E)' (I - K_d) \text{vec} \left[ (dA) A' + A dA' \right]. \tag{5.35}
\]
Let \( K \) denote the commutation matrix, i.e. 
\[
\text{vec} \ X = K \text{vec} \ X'. 
\]
Then 
\[
(I + K) (I - K_d) \text{vec} \ E = (I + K) \text{vec} \ E_u = 2 \text{vec} \ E_u. 
\]
Using this relationship together with a well-known equality relating the Kronecker product and the vec operator, viz. 
\[
\text{vec} \ PQR = \left( R' \otimes P \right) \text{vec} \ Q 
\]
(e.g., Magnus and Neudecker 1999, p. 30), (5.35) becomes 
\[
d\psi = -2(\text{vec} \ E)' (I - K_d) (I + K) \text{vec} \left[ (dA) A' \right] 
\]
\[
= -4(\text{vec} \ E_u)' \text{vec} \left[ (dA) A' \right] 
\]
\[
= -4(\text{vec} \ E_u)' (A \otimes I) \text{vec} \ (dA). 
\]
Thus, the first-order condition is

\[
(A' \otimes I) \text{vec} E_u = \text{vec} E_u A = 0_{p \times 1},
\]

hence

\[
E_u A = 0_{p \times k}. \tag{5.36}
\]

As

\[
E_u = (S - AA')_u = S_u + (AA')_d - AA',
\]

we can write (5.36) as

\[
(S_u + (AA')_d) A = AA'A. \tag{5.37}
\]

Then,

\[
E_u^2 = (S_u + (AA')_d)^2 + AA' AA' - (S_u + (AA')_d) AA' - AA' (S_u + (AA')_d)
= (S_u + (AA')_d)^2 - (AA')^2,
\]

hence \( \psi \) becomes

\[
\psi = \text{trace} \left( S_u + (AA')_d \right)^2 - \text{trace} \left( A'A \right)^2. \tag{5.38}
\]

If we postmultiply \( A \) by an orthogonal matrix \( T \) and insert this rotated matrix, say \( \hat{A} \), in (5.38) it is immediately clear that \( \hat{A} \) is also a solution. Then, without loss of generality we take \( A'A = \Gamma \) where \( \Gamma \) is a positive definite diagonal matrix. Equations (5.37) and (5.38) become

\[
(S_u + (AA')_d) A = A \Gamma \tag{5.39}
\]

and

\[
\psi = \text{trace} \left( S_u + (AA')_d \right)^2 - \text{trace} \Gamma^2. \tag{5.40}
\]

To attain a minimum for \( \psi \) we need to search for a maximum of \( \text{trace} \Gamma^2 \).

Equations (5.39) and (5.40) suggest the following iterative procedure to find \( A \):
Chapter 5. Correspondence analysis of more than two categorical variables

(0) Select an initial block diagonal matrix $\Phi_d^t$, e.g.

$$\Phi_d^t = (W^tW)_d^t,$$

with $W$ a $q \times k$ matrix whose columns are eigenvectors of $S$ associated with the $k$ largest eigenvalues and of squared length equal to the associated eigenvalue.

(1) Compute $S' := S_u + \Phi_d^t$.

(2) Compute the eigenvalues and associated eigenvectors of $S'$ and determine $A$ and $\Gamma$ in such a way that the columns of $A$ are appropriately scaled eigenvectors of $S'$ associated with the $k$ largest eigenvalues of $S'$.

(3) In step 1, replace $\Phi_d^t$ by $(AA')_d$.

(4) Iterate (1) – (3) until stability has been reached.

This algorithm is easily implemented and appears to converge sufficiently fast in practice. However, after convergence we do not know whether the obtained minimum is a global, or a local minimum. In addition to this problem it should be noted that the matrix $S'$ in step (1) is not necessarily positive semi-definite. Moreover, one cannot be sure that after convergence of the iteration procedure the diagonal blocks of $\Phi_d = (S - AA')_d$ are positive (semi)definite. As both $S$ and the diagonal blocks of $\Phi_d$ represent covariance matrices, this situation is cumbersome. Consider for example the case where we find negative or zero eigenvalues among the $k$ largest eigenvalues calculated in step (2).

In ordinary factor analysis the situation where a diagonal element of $D_\phi$ (i.e. a variable-specific variance) becomes negative is referred to as a Heywood case. In the context of joint correspondence analysis, which will be explained in the next section, Boik (1996) introduced an alternative algorithm aimed to deal with this multigroup factor analysis analogue to a Heywood case. If the situation should occur, i.e. if the diagonal blocks of $\Phi_d$ are not positive semi-definite, his algorithm can be used instead.

Finally, it should be noted that the solutions of principal factor analysis are not nested, i.e. if instead of a $k$–dimensional solution one is interested in a $k^*$–dimensional solution (with $k^* < k$) the iteration process must be carried out again with $k^*$ substituted for $k$ in steps (0) and (2).
5.4.3 Joint correspondence analysis

In section 5.3.2 we showed that multiple correspondence analysis is closely related to principal coordinate analysis. In a similar way joint correspondence analysis is related to principal factor analysis. In fact, joint correspondence analysis is mathematically equivalent to multigroup factor analysis. The only difference between the two methods is the explicit formulation of a factor model underlying the data. In joint correspondence analysis there is no such model. In this section, however, we derive the joint correspondence analysis solution using the multigroup factor model.

Consider the rows of the indicator matrix $Z$ as observations on a random vector, say

$$
\zeta = \begin{pmatrix}
\zeta_1 \\
\zeta_2 \\
\vdots \\
\zeta_q
\end{pmatrix},
$$

and assume that these $q$ variables can be described using the multigroup factor model:

$$
\zeta - \mu = Af + u.
$$

where $\mu$ is the mean vector. The sample covariance matrix of $\zeta$ is

$$
S_{\zeta} = \frac{1}{n}Z'MZ.
$$

Hence, to estimate a loading matrix $A$ we use the algorithm as described in the previous section with $S_{\zeta}$ substituted for $S$. Then, $AA'$ provides a least-squares approximation for the off-diagonal blocks of $S_{\zeta}$.

In accordance with (5.6) we introduce the "standardized" random variable $\zeta^* = \frac{1}{\sqrt{q}} D_z^{-\frac{1}{2}} \zeta$. The factor model for this standardized variable can be expressed as

$$
\zeta^* - \mu^* = \frac{1}{\sqrt{q}} D_z^{-\frac{1}{2}} Af + \frac{1}{\sqrt{q}} D_z^{-\frac{1}{2}} u = \tilde{A}f + \tilde{u},
$$

where

$$
\tilde{A} \equiv \frac{1}{\sqrt{q}} D_z^{-\frac{1}{2}} A \quad \text{and} \quad \tilde{u} \equiv \frac{1}{\sqrt{q}} D_z^{-\frac{1}{2}} u.
$$

Then, we have as expression for the variance matrix of $\zeta^*$:

$$
Var(\zeta^*) = \frac{1}{q} D_z^{-\frac{1}{2}} AA'D_z^{-\frac{1}{2}} + \frac{1}{q} D_z^{-\frac{1}{2}} \Phi D_z^{-\frac{1}{2}} = \tilde{A}\tilde{A}' + \tilde{\Phi}_d.
$$
Chapter 5. Correspondence analysis of more than two categorical variables

Hence, the loading matrix $\tilde{A}$ can be obtained immediately using $A$, i.e.

$$\tilde{A} = \frac{1}{\sqrt{q}} D_z^{-\frac{1}{2}} A$$

and

$$\frac{1}{q} \tilde{A}' D_z \tilde{A} = A'A = \Gamma. \quad (5.41)$$

When applying principal factor analysis based on the sample covariance matrix of $\zeta^*$ we obtain a least-squares approximation for the off-diagonal blocks of $S_z = \frac{1}{nq} D_z^{-\frac{1}{2}} Z'MZD_z^{-\frac{1}{2}}$.

Greenacre (1988) defined joint correspondence analysis as the simultaneous least-squares approximation of the weighted contingency matrices in deviations from independence. From (5.9) it is clear that these weighted contingency matrices in deviations from independence are the off-diagonal blocks of $S_z$. Hence, the multigroup factor analysis approach described above is equivalent to Greenacre's joint correspondence analysis approach.

A joint correspondence analysis solution can be obtained using the algorithm described in the previous section where $S_z$ is inserted for $S$ and $A$ is standardized according to (5.41). It should be noted, however, that Greenacre (1988) proposes another algorithm. Moreover, the solution he obtains is standardized as

$$A'D_z A = I_k.$$

**Quality**

To assess the quality of the joint correspondence analysis solution we compare the sum of squared residuals ($Rss$) with the total variation between the categories of the different variables ($Tss$). We have

$$Rss = \psi = \text{trace } E_u'E_u,$$

and

$$Tss = \text{trace } S_u'S_u. \quad (5.42)$$

Hence, the quality can be expressed as

$$1 - \frac{Rss}{Tss}.$$
5.4. Joint correspondence analysis

Rotation

When describing joint correspondence analysis as a factor analysis approach it is immediately clear that the loading matrix $A$ is not unique. In factor analysis the loadings are usually rotated such that they are of simple structure, i.e. elements are, in absolute value either relatively high, or close to zero. A well-known approach for rotation in factor analysis is Kaiser's varimax rotation. In the next chapter rotation will be discussed more fully.

5.4.4 Relationship with correspondence analysis

An important advantage of joint correspondence analysis over multiple correspondence analysis is the fact that joint correspondence analysis has correspondence analysis as a special case. As we showed in section 5.3.1 multiple correspondence analysis of two variables yields equivalent standard coordinate matrices as obtained in the correspondence analysis of the corresponding contingency matrix. However, the principal coordinates cannot be retained without additional steps to calculate the appropriate inertias. In joint correspondence analysis, however, we are only concerned with the least-squares approximation of the off-diagonal blocks. In the case of two variables there will be two such block matrices, each being equal to the other's transpose. The least-squares approximation of these blocks will be equivalent to the least-squares approximation obtained in correspondence analysis.

To obtain this solution in one iteration, we express the covariance matrix $S_{rr}$ for the two-variable case as

$$S_{rr} = \frac{1}{2n} \begin{pmatrix} D_r^{-\frac{1}{2}} & (D_r - \frac{1}{n}rr') \left( F - \frac{1}{n}rc' \right) \left( D_r^{-\frac{1}{2}} \right) \\ D_c^{-\frac{1}{2}} & (F' - \frac{1}{n}cr') \left( D_c^{-\frac{1}{2}} \right) \end{pmatrix}$$

$$= \frac{1}{2n} \begin{pmatrix} I_p - D_r^{\frac{1}{2}} 1_p 1_p' D_r^{\frac{1}{2}} & D_r^{-\frac{1}{2}} \left( F - \frac{1}{n}rc' \right) D_c^{-\frac{1}{2}} \\ D_c^{-\frac{1}{2}} \left( F - \frac{1}{n}rc' \right)' D_r^{-\frac{1}{2}} & I_p - D_c^{\frac{1}{2}} 1_p 1_p' D_c^{\frac{1}{2}} \end{pmatrix}$$

where we used (5.10) through (5.12). For convenience we will drop the constant term $\frac{1}{2n}$. Furthermore, we choose as initial block diagonal matrix $\Phi^i_d$ a matrix consisting solely of zeroes. Then, in step (2) of the algorithm we consider the eigenequation of $S^t = S_u + \Phi^i_d$,
Chapter 5. Correspondence analysis of more than two categorical variables

\[ \left( \begin{array}{cc} 0_{nxn} & D_r^{-\frac{1}{2}} \left( F_{-\frac{1}{n}rc'} \right) D_c^{-\frac{1}{2}} \\ D_c^{-\frac{1}{2}} \left( F_{-\frac{1}{n}rc'} \right) & 0_{pxp} \end{array} \right) \left( \begin{array}{c} U \\ V \end{array} \right) = \left( \begin{array}{c} U \\ V \end{array} \right) \Gamma. \]

Note that \( S' \) is not positive (semi)definite. However, as will become clear, by choosing \( k \leq \kappa \), where \( \kappa \) is the rank of \( D_r^{-\frac{1}{2}} \left( F_{-\frac{1}{n}rc'} \right) D_c^{-\frac{1}{2}} \), this will not pose a problem. Assuming that \( \Gamma^{-1} \) exists we get

\[ \left\{ \begin{array}{l} D_r^{-\frac{1}{2}} \left( F_{-\frac{1}{n}rc'} \right) D_c^{-\frac{1}{2}} V = U \Gamma \\ D_c^{-\frac{1}{2}} \left( F_{-\frac{1}{n}rc'} \right)' D_r^{-\frac{1}{2}} U = V \Gamma \end{array} \right. \rightarrow \left\{ \begin{array}{l} D_r^{-\frac{1}{2}} \left( F_{-\frac{1}{n}rc'} \right) D_c^{-\frac{1}{2}} V \Gamma^{-1} = U \\ D_c^{-\frac{1}{2}} \left( F_{-\frac{1}{n}rc'} \right)' D_r^{-\frac{1}{2}} U \Gamma^{-1} = V \end{array} \right. \]

so that

\[ \left\{ \begin{array}{l} D_c^{-\frac{1}{2}} \left( F_{-\frac{1}{n}rc'} \right)' D_r^{-1} \left( F_{-\frac{1}{n}rc'} \right)' D_r^{-\frac{1}{2}} U = U \Gamma^2 \\ D_c^{-\frac{1}{2}} \left( F_{-\frac{1}{n}rc'} \right)' D_r^{-1} \left( F_{-\frac{1}{n}rc'} \right)' D_c^{-\frac{1}{2}} V = V \Gamma^2 \end{array} \right. \]

and, letting \( U'U = V'V = I_k \), these are exactly the eigen-equations obtained in the correspondence analysis of \( F \) with \( \Gamma^2 \) playing the role of \( \Lambda \). The least-squares approximation for the off-diagonal blocks of \( S' \) becomes \( U \Gamma V' \) for the upper right off-diagonal block, and \( V \Gamma U' \) for the lower left off-diagonal block. Thus, this solution is optimal and the algorithm immediately terminates.

A loading matrix \( \tilde{A} \) can be obtained in the following way,

\[ \tilde{A} = \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} = D_r^{-\frac{1}{2}} \begin{pmatrix} U \\ V \end{pmatrix} \Gamma^{\frac{1}{2}} = \begin{pmatrix} D_r^{-\frac{1}{2}} U \Gamma^{\frac{1}{2}} \\ D_c^{-\frac{1}{2}} V \Gamma^{\frac{1}{2}} \end{pmatrix} \]

so that, in accordance with (5.41),

\[ \frac{1}{2} \tilde{A}'D_r\tilde{A} = \Gamma, \]

and

\[ A_1'D_rA_1 = A_2'D_cA_2 = \Gamma. \]

The two sub-matrices of the matrix of factor loadings, i.e. \( A_1 \) and \( A_2 \), are rescaled versions of the principal coordinate matrices \( H \) and \( G \), cf. (2.27) and (2.28), obtained in the correspondence analysis of \( F \), viz

\[ G = \sqrt{n}D_r^{-\frac{1}{2}}U\Gamma = \sqrt{n}A_1\Gamma^{\frac{1}{2}} \]
5.5. An example

and

\[ H = \sqrt{n}D_c^{-\frac{1}{2}}V\Gamma = \sqrt{n}A_2\Gamma, \]

so that

\[ G'D_rG = H'D_rH = n\Gamma^2. \]

In the context of correspondence analysis, a plot in which the rows of \( \sqrt{n}A_1 \) and \( \sqrt{n}A_2 \) are depicted as points in \( k \)-dimensional space, is sometimes referred to as a *symmetric joint correspondence analysis plot*, see e.g. Tijssen (1989) and Gifi (1990).

Note that

\[ \text{cov}(\zeta^*, f) = \frac{1}{\sqrt{q}}D^{-\frac{1}{2}}\text{cov}(\zeta, f) = \frac{1}{\sqrt{q}}D^{-\frac{1}{2}}A = A, \]

hence, using the analogy between factor analysis and joint correspondence analysis we can interpret these *symmetrical coordinates* as covariances between the rescaled variables and the underlying factors.

5.5 An example

To illustrate the use of both multiple correspondence analysis and joint correspondence analysis we will analyze data from Jobson (1996). The data set consists of 200 observations from a mail survey concerned with the analysis of women's attitudes pertaining to shopping for clothing. The data can be downloaded from Springer's website at http://www.springer-ny.com/supplements/jobson.

Respondents indicated on a scale from 1 (strongly agree) to 5 (strongly disagree) their attitudes with respect to the following seven statements:

A: I like sales people to leave me alone until I find clothes that I want to buy.

B: I like to pay cash for clothing purchases.

C: Price is a good indicator of the quality of clothes.

D: I usually spend more than I planned when shopping for clothes.
Chapter 5. Correspondence analysis of more than two categorical variables

E: I do not think clothes shops provide enough customer service these days.

F: When clothing is sold at a reduced price there is often something wrong with it.

G: I like to shop where my friends shop for clothes.

These seven items are designed to measure shopping orientation. In addition to these items we also have a variable which indicates whether respondents work outside the home or whether they do not work outside the home. Finally, the age of the respondents is registered by means of a categorical variable categorized in the following manner: < 24, 25-34, 35-44, 45-54, 55-64 and >65.

The data on these nine variables were recoded in the format of an indicator matrix $Z$. This indicator matrix was subjected to multiple and joint correspondence analysis. As rank of the approximation we chose for both approaches $k = 2$. Fig. 5.1 gives a plot of the principal coordinates for the variables obtained in the multiple correspondence analysis and Fig. 5.2 gives a plot obtained using joint correspondence analysis where the coordinates are scaled as in (5.41). Comparison of Fig. 5.1 and Fig. 5.2 shows that the differences between the two displays are very small. Note, however, the difference in the proportion of explained inertia: 24% for the multiple correspondence analysis versus 88% for the joint correspondence analysis.

We can interpret the results, as represented in Fig. 5.1 and Fig. 5.2, in the following way. The younger, typically outside the home working respondents are separated from the older, not outside the home working respondents with respect to their shopping for clothing attitudes. The younger respondents indicate a strong agreement with items A and D, i.e. they like to do their shopping without assistance of sales people and they feel that they usually spend more money than planned. This is contrasted with a strong disagreement on all other items: They do not prefer to pay with cash, the price is not considered to be a good indicator for quality and, in accordance with this, clothes that are sold at a reduced price are not necessarily flawed, customer service is not insufficient and they do not prefer to shop at the same place as friends. For the older respondents exactly the opposite holds.
5.5. An example

Figure 5.1: Multiple correspondence analysis plot (categories in principal coordinates)  
Explained Inertia: 24.44%

Figure 5.2: Joint correspondence analysis plot. Explained Inertia: 88.26%
5.6 Concluding remarks

In this chapter we have restricted ourselves to two approaches for the analysis of more than two categorical variables using correspondence analysis. There are, however, several alternatives to these approaches. For example, Gower and Hand (1996) propose alternatives that are based on different distance functions than the—for correspondence analysis of a contingency matrix typical, and justifiable—chi-squared distance. In addition, Greenacre (1993b) defines multiple correspondence analysis as the correspondence analysis of $Z'Z$, which is usually referred to as the Burt matrix, rather than that of $Z$. This alternative definition is rather confusing as most texts on correspondence analysis, including Greenacre (1984), define multiple correspondence analysis as the correspondence analysis of $Z$. (The procedure also followed in this chapter). It is known, and in fact not difficult to show, e.g. Greenacre (1984), Lebart et al. (1984), that the analysis of $Z'Z$ and $Z$ are tightly connected. In fact, the standard coordinates for the columns of $Z$, i.e. for the categories of the variables, are equivalent to the standard coordinates for the rows (and columns, as $Z'Z$ is symmetric) of $Z'Z$. As the singular values of $Z'Z$ are the squares of the singular values of $Z$ the principal coordinates are standardized in a different fashion in the two approaches. For a derivation of this relationship we refer the interested reader to Lebart et al. (1984). A more elaborate treatment of some interpretational aspects of the use of $Z$ and $Z'Z$ can be found in Greenacre (1994).

As alternative to the iterative joint correspondence analysis approach described in section 5.4.3, Greenacre (1991) proposes a “compromise” between joint correspondence analysis and multiple correspondence analysis. Gower and Hand (1996) refer to this “compromise” as a non-iterative joint correspondence analysis approach. In this non-iterative approach the standard coordinate matrix $Y$ is the same as the one obtained in multiple correspondence analysis. However, these standard coordinates are rescaled such that the off-diagonal blocks of the Burt matrix are, in a least-squares sense, best approximated. For a clear exposition of this approach we refer to Gower and Hand (1996).

The one common feature of the approaches mentioned above, as well as some additional alternatives given by Greenacre (1993c), is that they are all restricted to the
5.6. Concluding remarks

analysis of pairwise associations in the multiway data. A completely different approach
that is also concerned with higher-order interactions is described by Carlier and Kroon-
enberg (1996). Their approach is related to so-called three-mode principal component
analysis, e.g. Kroonenberg and de Leeuw (1980) and Kroonenberg (1983). Kapteyn
et al. (1986) showed that extensions for three-mode principal component analysis to
$n$-mode component analysis are straightforward. These extensions can also be used to
define $n$-mode correspondence analysis in a similar fashion as proposed by Carlier and
Kroonenberg (1996). Finally, using so-called quantification matrices Kiers (1989) showed
that multiple correspondence analysis can in fact be considered as a form of three-mode
component analysis. Moreover, Kiers (1989) showed that rotation of the multiple corre-
spondence analysis approximation can simplify the interpretation. In the next chapter
we will consider rotation of correspondence analysis in a more general fashion.