Chapter 3

An important eigenvalue property encountered in correspondence analysis

3.1 Introduction

In the previous chapter we introduced correspondence analysis as two-mode component analysis of the weighted matrix of frequencies in deviation from independence, i.e. the analysis of $\tilde{F} = D_r^{-\frac{1}{2}} (F - \frac{1}{2}rr') D_c^{-\frac{1}{2}}$. As an alternative to this approach we could simply consider the weighted frequencies matrix $Q \equiv D_r^{-\frac{1}{2}}FD_c^{-\frac{1}{2}}$. The results of such an approach are, as we will show, equivalent. The reason for this equivalence is closely related to an important result concerning the singular values of the nonnegative matrix $Q$. This particular result is, despite the individual differences between the various mathematically equivalent methods mentioned in the introduction of Chapter 2, encountered in all of them.

Although the property is frequently cited, it is often introduced without proof, or accompanied by a 'statistical' proof. In this chapter some of these 'statistical' proofs will be translated into mathematical proofs. This will show how the several approaches lead to different proofs of the same property. In addition, a proof due to Tenenhaus and Young (1985) treating a special kind of data, namely data in the format of a so-called
Chapter 3. An important eigenvalue property encountered in correspondence analysis

indicator matrix, will be extended so that it also applies to data occurring in the format of a contingency matrix. Finally, four general proofs that tackle the problem directly through the matrix $Q$ will be provided.

From the expositions of the different proofs it will become clear that they rely heavily on the related statistical methods. This means that, in order to understand the proofs, one needs to have some knowledge of the method at hand. For example, in order to comprehend the proof in Greenacre's work one needs to understand some aspects of the geometrical framework in which correspondence analysis is being introduced. Similarly, the dual scaling proof due to Nishisato as described in section 3.3.1 requires knowledge of the decomposition of variance which is essential in dual scaling.

In addition to the treatment of these method-related proofs a new general proof using elementary matrix algebra will be provided. This proof has as great advantage over the other proofs that it directly solves the problem, without using any method-specific features in an amazingly straightforward and compact way. Finally, we will briefly discuss three other general proofs which also do not require familiarity with any of the methods. These proofs, however, involve rather advanced matrix algebra such as the Frobenius theorem (Gower and Hand (1996), Puntanen and Styan(1998)), row stochastic matrices (Puntanen and Styan(1998)) and Geršgorin's discs theorem (Graffelman (1998)). Moreover, the proof of Gower and Hand (1996) implicitly imposes a restriction on the matrix at hand.

Notation

The notation in this chapter will be consistent with the notation introduced in the previous chapter. For convenience let us restate some of the fundamental definitions:

$$r = F1_p \quad \text{and} \quad c = F'1_n$$  \hspace{1cm} (3.1)

are the vectors of row and column totals of a nonnegative $n \times p$ data matrix $F$. Furthermore we have scaling matrices

$$D_r \equiv diag\{r\} \quad \text{and} \quad D_c \equiv diag\{c\}.$$  \hspace{1cm} (3.2)

Finally, assuming that the matrix $F$ has no rows or columns that are completely zero,
3.2. The problem

the matrix of so-called row profiles was defined as

$$R \equiv D_r^{-1}F,$$  \hspace{1cm} (3.3)

and the corresponding column profile matrix as

$$C \equiv D_c^{-1}F'.$$

3.2 The problem

It is known that the singular values of a matrix $A$ are the positive roots of the non-trivial eigenvalues of $AA'$ (or equivalently of $A'A$). In correspondence analysis as well as in related methods one obtains singular values for a matrix $Q = D_r^{-\frac{1}{2}}FD_c^{-\frac{1}{2}}$, and these singular values are understood to lie in the $[0,1]$ interval.

One can write

$$Q'Qv = D_c^{-\frac{1}{2}}F'D_r^{-1}FD_c^{-\frac{1}{2}}v = \lambda v,$$  \hspace{1cm} (3.4)

and

$$QQ'u = D_r^{-\frac{1}{2}}FD_c^{-1}F'D_r^{-\frac{1}{2}}u = \lambda u,$$  \hspace{1cm} (3.5)

where $v$ is a $p \times 1$ eigenvector of $Q'Q$ corresponding to an eigenvalue $\lambda$ and $u$ is an $n \times 1$ eigenvector of $QQ'$ corresponding to the same eigenvalue. Proving that the singular values of $Q$ lie in the $[0,1]$ interval is the same as proving that the eigenvalues in (3.4) or equivalently in (3.5) are in that interval. Hence, we have to prove that

$$\lambda \in [0,1].$$  \hspace{1cm} (3.6)

Since $\lambda$ is an eigenvalue of a positive semi-definite matrix only the upper bound of the interval has to be considered, i.e. we have to prove that $\lambda$ is smaller than or equal to one. Before proceeding let us first clarify the use of $Q$ instead of $\tilde{F}$ as defined in (2.22).

Correspondence analysis' trivial solution

It is easily verified that $v = \frac{1}{\sqrt{p}}D_c^{\frac{1}{2}}1_p$ and $u = \frac{1}{\sqrt{n}}D_r^{\frac{1}{2}}1_n$ are standardized eigenvectors corresponding to an eigenvalue $\lambda = 1$ in equations (3.4) and (3.5) respectively. Then, the
Chapter 3. An important eigenvalue property encountered in correspondence analysis

Singular value decomposition of $Q$ can be expressed as

$$D_r^{-1} F D_c^{-1} = \left( \frac{1}{\sqrt{s}} D_p^2 1_n, U \right) \left( \begin{array}{c} 1 \\ \Lambda \end{array} \right) \left( \frac{1}{\sqrt{p}} 1_p D_c^2 V' \right) = \frac{1}{s} D_p^2 1_n 1_p D_c^2 + U \Lambda V'.$$

Re-arranging terms yields, after some manipulations,

$$D_r^{-\frac{1}{2}} \left( F - \frac{1}{s} r c' \right) D_c^{-\frac{1}{2}} = U \Lambda V'.$$

Hence, by considering $Q$ instead of the weighted matrix in deviations from independence one always finds the solution with $\lambda = 1$. This solution is usually referred to as the trivial solution. In the next sections we will show that this eigenvalue $\lambda = 1$ is also the largest.

### 3.3 Proofs

In the literature a general proof of the eigenvalue property does not exist. Escoufier (1971) uses the relationship between correspondence analysis and canonical correlation analysis to prove the result. Greenacre (1984) also mentions this proof and in addition he provides another proof. He argues that the optimal one-dimensional approximation to $R$ is the so-called centroid, a weighted average sometimes also referred to as centre of gravity, and this centroid is an eigenvector corresponding to eigenvalue one. Hence the largest eigenvalue equals one. Nishisato (1980, 1994) in his description of dual scaling finds that the eigenvalue $\lambda$ is in fact equal to, what he calls, a squared correlation ratio, hence it has to be between zero and one. Tenenhaus and Young (1985) only treat a special case in correspondence analysis, i.e. the case where the data are in the format of an indicator matrix.

As mentioned before, the one thing in common between these proofs is the fact that some understanding of the underlying statistical method is necessary. One needs to understand the rationale of the statistical method to understand the mathematical proof. We shall streamline these proofs in the following three sections. Each section will open with a short introduction to the method at hand. These introductions are not sufficient to get a deep understanding of the several methods and should not be considered as such. For a complete treatment of the methods the reader is referred to the standard texts mentioned in the previous chapter.
3.3. Proofs

The general proof to be proposed in section 3.3.4 does not need an introduction to any method. It takes the matrix Q as the starting point and the result follows immediately by elementary matrix algebra due to the nonnegativity of F and the scalings imposed by the F-dependent matrices $D_r$ and $D_c$.

3.3.1 Dual scaling proof for a contingency matrix

If one interprets dual scaling, a method closely related to correspondence analysis, as an “analysis-of-variance” approach, see Nishisato (1980, 1994), the result regarding the eigenvalues follows from the decomposition of variance essential to the method. To see this let us start with a brief description of the method.

In dual scaling one assigns scale values, often referred to as "weights", to columns (rows) of the data matrix in such a way that the homogeneity in each row (column) is maximized whilst the homogeneity between the rows (columns) is minimized. Let us consider the "weighting" of columns by a vector $y = (y_1, y_2, \ldots, y_p)'$. Nishisato decomposes the variance in the contingency matrix as follows

$$
\sum_{i=1}^{n} \sum_{j=1}^{p} f_{ij} (y_j - \bar{y})^2 = \sum_{i=1}^{n} \sum_{j=1}^{p} f_{ij} (\bar{y}_i - \bar{y})^2 + \sum_{i=1}^{n} \sum_{j=1}^{p} f_{ij} (y_j - \bar{y}_i)^2,
$$

(3.7)

where $\bar{y}$ denotes the overall average scale value, i.e.

$$
\bar{y} = \frac{1}{s} \sum_{i=1}^{n} \sum_{j=1}^{p} f_{ij}y_j = \frac{1}{s} y'Fy,
$$

where

$$
s \equiv \sum_{i=1}^{n} \sum_{j=1}^{p} f_{ij} = 1_n'F1_p,
$$

and $\bar{y}_i$ denotes the ith row average, that is

$$
\bar{y}_i = \frac{\sum_{j=1}^{p} f_{ij}y_j}{\sum_{j=1}^{p} f_{ij}}.
$$

(3.8)

In words (3.7) can be expressed as: Total sum of squared deviations (SS$t$) equals the sum of squared deviations between rows (SS$b$) plus the sum of squared deviations within rows (SS$w$):

$$
SS_t = SS_b + SS_w.
$$

(3.9)
Chapter 3. An important eigenvalue property encountered in correspondence analysis

(Note that (3.7) can be seen as a decomposition of variance where each entry $f_{ij}$ in the original matrix is replaced by $f_{ij}$ repetitions of $y_j$).

Nishisato's proof requires the maximization of the so-called squared correlation ratio $\eta^2$ which he defines as

$$\eta^2 = \frac{SSb}{SSt}.$$

Clearly

$$SSt \geq SSb \geq 0,$$

so that $\eta^2$ is smaller than or equal to one. Nishisato (1980, 1994) restricts $y$ in such a way that the overall average $\bar{y}$ equals zero, i.e.

$$1^n F y = 0.$$

For our purposes, however, we do not need this restriction. Instead, define

$$y_j^* \equiv y_j - \bar{y} \rightarrow y^* \equiv y - \bar{y} 1,$$

and

$$\bar{y_i}^* \equiv \bar{y_i} - \bar{y}.$$

Then, if one defines an $n \times 1$ mean vector $m$ as

$$m \equiv [y_1^*, \ldots, y_n^*],$$

it follows from (3.8) and the definitions of $D_r$, $y^*$ and $\bar{y_i}$ that

$$m = D_r^{-1} F y^*.$$

Rewrite the first two terms of (3.7) using the definitions of $y^*$, $c$, $r$ and $m$ as

$$SSt = \sum_{i=1}^{n} \sum_{j=1}^{p} f_{ij} y_j^{*2} = y^* D_c y^*,$$

and

$$SSb = \sum_{i=1}^{n} \sum_{j=1}^{p} f_{ij} \bar{y_i}^{*2} = m' D_r m = y^* F' D_r^{-1} F y^*.$$
3.3. Proofs

Define

\[ z \equiv D_c^{\frac{1}{2}} y^* , \]

and write the dual-scaling objective as

\[
\max_z \eta^2 = \frac{SSb}{SSt} = \frac{z' D_c^{\frac{1}{2}} F' D_r^{-1} D_c^{\frac{1}{2}} z}{z' z} \tag{3.10}
\]

Differentiating \( \eta^2 \) and equating to zero yields

\[
0 = d\eta^2 = \frac{2z' D_c^{\frac{1}{2}} F' D_r^{-1} D_c^{\frac{1}{2}} dz}{z' z} - \frac{2z' D_c^{\frac{1}{2}} F' D_r^{-1} D_c^{\frac{1}{2}} zz' dz}{(z' z)^2} \rightarrow D_c^{-\frac{1}{2}} F' D_r^{-1} D_c^{-\frac{1}{2}} z = \eta^2 z. \tag{3.11}
\]

Hence, \( \eta^2 \) is the largest eigenvalue of \( D_c^{-\frac{1}{2}} F' D_r^{-1} D_c^{-\frac{1}{2}} \).

A new alternative proof that does not require any maximization can be obtained by noting that, algebraically, equality (3.7) holds for any \( y \) and we may therefore take \( y \) to be zero. Then,

\[
\sum_{i=1}^n \sum_{j=1}^p f_{ij} y_j^2 = \sum_{i=1}^n \sum_{j=1}^p f_{ij} \tilde{y}_i^2 + \sum_{i=1}^n \sum_{j=1}^p f_{ij} (y_j - \tilde{y}_i)^2, \tag{3.12}
\]

where \( y \) can be any vector, and all three terms are, due to the nonnegativity of \( F \), nonnegative.

Clearly now

\[
\sum_{i=1}^n \sum_{j=1}^p f_{ij} y_j^2 = y' D_c y,
\]

and

\[
\sum_{i=1}^n \sum_{j=1}^p f_{ij} \tilde{y}_i^2 = y' F' D_r^{-1} F y.
\]

Hence

\[
\sum_{i=1}^n \sum_{j=1}^p f_{ij} (y_j - \tilde{y}_i)^2 = y' D_c y - y' F' D_r^{-1} F y = y' D_c^{\frac{1}{2}} (I_p - D_c^{-\frac{1}{2}} F' D_r^{-1} D_c^{\frac{1}{2}}) D_c^{\frac{1}{2}} y.
\]
Due to the nonnegativity of $F$, this expression is nonnegative. Thus, since $y$ can be any vector, the matrix $I_p - D^{-\frac{1}{2}}F'D^{-1}FD^{-\frac{1}{2}}$ is positive semi-definite. This implies that

$$
\rho(I_p - D^{-\frac{1}{2}}F'D^{-1}FD^{-\frac{1}{2}}) \geq 0 \rightarrow 0 \leq \rho(D^{-\frac{1}{2}}F'D^{-1}FD^{-\frac{1}{2}}) \leq 1,
$$

where $\rho(\cdot)$ denotes the eigenvalue function. We see that the result regarding the non-trivial eigenvalues of $D^{-\frac{1}{2}}F'D^{-1}FD^{-\frac{1}{2}}$ is a direct consequence of the variancelike decomposition (3.12).

### 3.3.2 Geometrical proof for a contingency matrix

Greenacre (1984) explains correspondence analysis by employing geometrical concepts such as weighted Euclidean distances. First he standardizes the data by

$$
\mathbf{1}'_n \mathbf{F} \mathbf{1}_p = 1, \quad (3.13)
$$

then row- and column-profile matrices $\mathbf{R}$ and $\mathbf{C}$ are introduced. The aim of correspondence analysis is to simultaneously obtain low-rank approximations of $\mathbf{R}$ and $\mathbf{C}$ by minimizing the so-called weighted chi-squared distances between these approximations and the profiles matrices. As was shown in section (2.4) these weighted chi-squared distances are in fact equivalent to the Mahalanobis distances. For our purposes, it suffices to consider the approximation of the row profiles.

For the approximation of the row profiles matrix $\mathbf{R}$ as defined in (3.3) Greenacre (1984) formulates the following objective

$$
\min_{\mathbf{B}, \mathbf{W}} \text{trace} \left( \mathbf{D}_r (\mathbf{R} - \mathbf{W} \mathbf{B}') \mathbf{D}_c^{-1} (\mathbf{R} - \mathbf{W} \mathbf{B}')' \right) \quad (3.14)
$$

$$
\text{s.t. } \mathbf{B}' \mathbf{D}_c^{-1} \mathbf{B} = \mathbf{I}_k,
$$

where $\mathbf{W}$ is an $n \times k$ matrix of rank $k$ ($k < p$), and $\mathbf{B}$ is of the order $p \times k$ and also of rank $k$. Clearly this objective is similar to (2.49). Instead of the profiles in deviations from the mean profiles however, we consider the profile matrix $\mathbf{R}$. The solution to this minimization problem can be obtained by substituting $\mathbf{R}$ for $\bar{\mathbf{R}}$ in the first order conditions (2.50) and (2.51). Hence, after insertion of (3.3) we have
3.3. Proofs

\[ \mathbf{F}' \mathbf{D}_r^{-1} \mathbf{F} \mathbf{D}_c^{-1} \mathbf{B} = \mathbf{B} \Lambda \quad (3.15) \]

\[ \mathbf{W} = \mathbf{D}_r^{-1} \mathbf{F} \mathbf{D}_c^{-1} \mathbf{B} \]

\[ \mathbf{B}' \mathbf{D}_c^{-1} \mathbf{B} = \mathbf{I}_k, \]

where \( \Lambda \) is a diagonal matrix of eigenvalues, i.e. \( \Lambda = \text{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_k\} \). In order to prove the eigenvalue property (3.6) Greenacre uses the result that an optimal sub-space necessarily contains the so-called centroid, the weighted average, of the original points. This can be shown in the following way.

Suppose we have a set of \( n \) points in \( p \)-dimensional space, say \( y_1, y_2, \ldots, y_n \) where \( y_i \) \((i = 1 \ldots n)\) is a \( p \times 1 \) vector. Define the centroid as

\[ \bar{y} = \frac{1}{n} \sum_{i=1}^{n} w_i y_i, \quad (3.16) \]

where the weights \( w_i \) are nonnegative and

\[ \sum_{i=1}^{n} w_i = 1. \]

Let \( S^* \) denote a \( k \)-dimensional sub-space \((k < p)\), and let \( y_i^*, i = 1 \ldots n \) be the points in this sub-space closest to the points \( y_1, y_2, \ldots, y_n \).

Greenacre defines the weighted squared distance between the original points and the points in \( S^* \) as

\[ \psi(S^*; y_1, y_2, \ldots, y_n) = \sum_{i=1}^{n} w_i (y_i - y_i^*)' \mathbf{D}_q (y_i - y_i^*), \quad (3.17) \]

where \( \mathbf{D}_q \) is a \( p \times p \) diagonal matrix with positive diagonal. Define then

\[ \bar{y}^* = \sum_{i=1}^{n} w_i y_i^*, \]

and let

\[ t = \bar{y} - \bar{y}^*. \]
Chapter 3. An important eigenvalue property encountered in correspondence analysis

Now, let \( \hat{y}_i \) denote a point in another \( k \)-dimensional sub-space, say \( S \), defined as

\[
\hat{y}_i = y_i^* + t, \quad \text{(for } i = 1 \ldots n)\].

Note that the centroid \( \bar{y} \) is also the centroid of the points in \( S \), since

\[
\sum_{i=1}^{n} w_i \hat{y}_i = \sum_{i=1}^{n} w_i (t + y_i^*) = t + \bar{y}^* = \bar{y}. \tag{3.18}
\]

Rewrite (3.17) as

\[
\psi(S^*; y_1, y_2, \ldots, y_n) = \sum_{i=1}^{n} w_i (y_i - \hat{y}_i + \hat{y}_i - y_i^*)' D_q (y_i - \hat{y}_i + \hat{y}_i - y_i^*)
\]

\[
= \sum_{i=1}^{n} w_i (y_i - \hat{y}_i)' D_q (y_i - \hat{y}_i) + \sum_{i=1}^{n} w_i (\hat{y}_i - y_i^*)' D_q (\hat{y}_i - y_i^*)
\]

\[+ 2 \sum_{i=1}^{n} w_i (y_i - \hat{y}_i)' D_q (\hat{y}_i - y_i^*)
\]

\[
= \sum_{i=1}^{n} w_i (y_i - \hat{y}_i)' D_q (y_i - \hat{y}_i) + t' D_q t
\]

\[
= \psi(S; y_1, y_2, \ldots, y_n) + t' D_q t. \tag{3.19}
\]

Here the identities

\[
\sum_{i=1}^{n} w_i (y_i - \hat{y}_i)' D_q (\hat{y}_i - y_i^*) = \sum_{i=1}^{n} w_i (y_i - \hat{y}_i)' D_q t,
\]

and

\[
\sum_{i=1}^{n} w_i (y_i - \hat{y}_i) = \bar{y} - \sum_{i=1}^{n} w_i \hat{y}_i = 0,
\]

were used. From (3.19) it is clear that, as long as \( t \neq 0 \), \( \psi(S^*; y_1, y_2, \ldots, y_n) \) is larger than \( \psi(S; y_1, y_2, \ldots, y_n) \). Thus, \( S^* \) can never be the optimal sub-space unless \( \bar{y}^* = \bar{y} \), i.e. the sub-space closest to the original points must contain their centroid.

Now, let us return to the problem of finding an optimal sub-space for the row profiles as formulated by Greenacre. The \( n \) rows of the profile matrix \( R \) can be seen as \( p \)-dimensional observation vectors. Due to the standardization in (3.13) a centroid can be defined in a similar fashion as was done in (3.16), i.e.

\[
r_c \equiv \sum_{i=1}^{n} r_i R_i,
\]
where \( r_i \) denotes the sum of the \( p \) elements in the \( i \)th row of \( F \), i.e. the \( i \)th element of \( r \) as defined in (3.1), and \( R_i \) is the \( i \)th row of \( R \) written as a column.

From the definitions of \( R \) in (3.3) and \( c \) in (3.1) it follows that

\[
R_c = R'r = F'D_\tau^{-1}r = F'1_n = c.
\]

Hence, \( c \) is the centroid of the row profiles and as such it necessarily lies in the row space of \( R \).

Now, let \( k = 1 \), i.e. approximate \( R \) by \( uv' \). Equation (3.15) can then be rewritten as

\[
F'D_\tau^{-1}FD_\tau^{-1}v = \lambda_1 v.
\] (3.20)

Due to the fact that the centroid \( c \) has to be in the row space of \( uv' \) (the optimal sub-space contains the centroid) we can write

\[
c = vu'g \Rightarrow v = \gamma c,
\] (3.21)

where \( g \) is an \( n \times 1 \) vector, and \( \gamma = \frac{1}{u'g} \). Inserting (3.21) in (3.20) yields

\[
F'D_\tau^{-1}FD_\tau^{-1}v = \lambda_1 v \Rightarrow F'D_\tau^{-1}FD_\tau^{-1}c = \lambda_1 c \Rightarrow c = \lambda_1 c,
\]

hence

\[
\lambda_1 = 1.
\]

Greenacre (1984) also provides another argument which can be formalized in a proof. He argues that the singular values obtained in correspondence analysis of a contingency matrix are equal to the canonical correlations obtained in the canonical correlation analysis of the same categorical data. Then, as the canonical correlations lie in the [0, 1] interval (e.g. Mardia et al., 1979, Escoufier, 1971), the eigenvalue property immediately follows. This relationship between these two methods is well known, see e.g. Escoufier (1971), Greenacre (1984) or Lebart et al. (1984). A complete proof of the relationship between the singular values obtained in correspondence analysis and the canonical correlations, can be found in several texts on correspondence analysis, e.g. Greenacre (1984) or Lebart et al. (1984).
3.3.3 Tenenhaus and Young's proof for an indicator matrix

Tenenhaus and Young (1985) solve the problem for the case where the data matrix is a so-called indicator matrix. Correspondence analysis of such data is usually referred to as multiple correspondence analysis (Greenacre, 1984) or homogeneity analysis (Gifi, 1990). By using the relationship between multiple correspondence analysis and correspondence analysis of a contingency matrix Tenenhaus and Young's result can be extended to the case where $F$ is a contingency matrix.

Suppose one has data for $n$ individuals on $k$ categorical variables with $p_j$ ($j = 1 \ldots k$) categories. An indicator matrix $Z_j$ can be constructed for the $j$th variable by letting the $p_j$ columns of $Z_j$ represent the categories of the variable whereas each row of $Z_j$ represents an individual. The elements of $Z_j$, say $z_{ij}$, are one if an individual $i$ falls in a category $j$, and all remaining elements are zero. Moreover, each individual is restricted to fall in one category only, i.e. each row consists of $p_j - 1$ zero elements and one element equal to one. Hence, $Z_j 1_{p_j} = 1_n$.

Data for $n$ individuals on $k$ categorical variables can be collected in what is also called an indicator matrix, say $Z$, where $Z$ is the partitioned matrix defined as

$$Z = (Z_1, Z_2, \ldots, Z_k).$$

By considering this matrix $Z$, Tenenhaus and Young (1985) prove the eigenvalue property (3.6) in the following way. The scaling matrices $D_r$ and $D_c$ are calculated like before with the indicator matrix $Z$ playing the role of $F$, i.e., $D_r$ and $D_c$ are now implicitly defined as

$$D_r 1_n = Z 1_p \text{ and } D_c 1_p = Z' 1_n,$$  \hspace{1cm} (3.22)

where

$$p \equiv \sum_{j=1}^{k} p_j.$$

We can rewrite the original problem using the indicator matrix $Z$ and the appropriate scaling matrices as defined in (3.22). That is, we now have to prove that the singular values of $D_r^{-\frac{1}{2}} Z D_c^{-\frac{1}{2}}$ lie in the $[0,1]$ interval.
From the definition of an indicator matrix it follows that $Z_j 1_p = 1_n$. Hence,

$$Z_1 p = k 1_n \rightarrow D_r = k I_n.$$  \hspace{1cm} (3.23)

Note that the singular values of $D_r^{-\frac{1}{2}} Z D_c^{-\frac{1}{2}}$ are the same as the positive square roots of the non-trivial eigenvalues of $D_c^{-1} Z' D_r^{-1} Z$. Let $g$ be a $p \times 1$ eigenvector of $D_c^{-1} Z' D_r^{-1} Z$, corresponding to an eigenvalue $\lambda$, i.e.

$$D_c^{-1} Z' D_r^{-1} Z g = \lambda g.$$ 

Substitute for $D_r^{-1}$ and divide through $\lambda$ to get

$$\frac{1}{\lambda k} D_c^{-1} Z' Z g = g.$$  \hspace{1cm} (3.24)

Now, if one writes $Z = (z(1), z(2), \ldots, z(p))$, where $z(l)$ (for $l = 1 \ldots p$) denotes the $l$th column of $Z$, (3.24) becomes

$$\frac{1}{\lambda k} \begin{pmatrix} \frac{1}{c_l} z(1) \\ \frac{1}{c_2} z(2) \\ \vdots \\ \frac{1}{c_p} z(p) \end{pmatrix} \rightarrow g_l = \frac{1}{\lambda k c_l} z(l)^T Z g,$$  \hspace{1cm} (3.25)

where $g_l$ denotes the $l$th element of $g$. Let $g_m$ be the largest element of $g$ in absolute value, and note that, without loss of generality, the choice $g_m > 0$ is legitimate. Define a vector $g_m$ as

$$g_m \equiv g_m 1_p,$$

then, from (3.25) and the definition of $g_m$ it follows that,

$$g_m = \frac{1}{\lambda k c_m} z(m)^T Z g \leq \frac{1}{\lambda k c_m} z(m)^T Z g_m = \frac{g_m}{\lambda k c_m} z(m)^T Z 1_p.$$  \hspace{1cm} (3.26)

Finally, applying (3.23) and the definition for $D_c$ to the right-hand side of (3.26) yields

$$\frac{g_m}{\lambda k c_m} z(m)^T Z 1_p = \frac{g_m}{\lambda c_m} z(m)^T 1_n = \frac{g_m}{\lambda},$$

Thus,

$$g_m \leq \frac{g_m}{\lambda} \rightarrow \lambda \leq 1.$$
Chapter 3. An important eigenvalue property encountered in correspondence analysis

Hence, the result is proven for the special case where $F$ is an indicator matrix. If, however, one has data on two categorical variables comprised in the indicator matrix $Z = (Z_1, Z_2)$, a contingency matrix can be constructed by simply calculating

$$F \equiv Z_1 Z_2.$$  \hfill (3.27)

It is known that the correspondence analysis of this contingency matrix $F$ and the correspondence analysis of the two-variable indicator matrix $Z$ are closely related (see for example Greenacre, 1984, Lebart et al., 1984 and Gifi, 1990). In fact, the eigenvalues are related by

$$\lambda_F = (1 - 2\lambda_Z)^2,$$  \hfill (3.28)

where $\lambda_F$ denotes an eigenvalue obtained in the correspondence analysis of the contingency matrix $F$ and $\lambda_Z$ denotes an eigenvalue obtained in the correspondence analysis of the indicator matrix $Z$. (A complete derivation of (3.28) can be found in Chapter 5, section 5.3.1). Because

$$0 \leq \lambda_Z \leq 1,$$

we also have

$$0 \leq \lambda_F \leq 1.$$

Hence, the non-trivial singular values obtained in a correspondence analysis of a contingency matrix $F$ are smaller than or equal to one.

3.3.4 Direct proof using the norm of a matrix

Define the row sum norm for a $p \times p$ matrix $A$ as

$$\| A \| \equiv \max_i \sum_{j=1}^{p} |a_{ij}|.$$  \hfill (3.29)

Compatible with the matrix norm (3.29) is the vector norm of an $n \times 1$ vector $z$

$$\| z \| = \max_{i \in \{1, \ldots, n\}} |z_i|, \text{ the largest element in absolute value of } z.$$
3.3. Proofs

Note that

$$\|Az\| \leq \|A\|\|z\|,$$

because

$$\|Az\| = \max_i \left| \sum_{j=1}^{n} a_{ij}z_j \right| \leq \max_i \sum_{j=1}^{n} |a_{ij}z_j| \leq \|z\| \max_i \sum_{j=1}^{n} |a_{ij}| = \|z\|\|A\|.$$ 

Now, let t be an eigenvector of A corresponding to eigenvalue $\lambda$, i.e., $At = \lambda t$. Then,

$$\|At\| = \|\lambda t\| = \|\lambda\|\|t\|,$$

and

$$\|At\| \leq \|A\|\|t\| \rightarrow \|\lambda\|\|t\| \leq \|A\|\|t\| \rightarrow |\lambda| \leq \|A\|.$$ 

Thus, no eigenvalue can exceed (in absolute value) the row sum norm. (For the defining properties of norms in general see, e.g. Basilevsky, 1983 or Stewart, 1973).

As $F$ is a nonnegative matrix and $D_c^{-1}F'D_r^{-1}F 1_p = 1_p$, i.e. each row sum equals one, the norm of $D_c^{-1}F'D_r^{-1}F$ becomes

$$\| D_c^{-1}F'D_r^{-1}F \| = 1$$

and it immediately follows that the singular values of the matrix $Q$, being equal to the positive roots of the non-trivial eigenvalues of $D_c^{-1}F'D_r^{-1}F$ lie in the [0,1] interval.

3.3.5 Three other proofs

A proof using the Frobenius theorem

Gower and Hand (1996, p. 261) provide a proof of the property that, like the proof in section 3.3.4, does not require an introduction to any statistical method. Their proof, however, is not valid for a reducible nonnegative matrix $Q'Q$. (For a definition of a reducible matrix see e.g. Gantmacher, 1959 or Horn and Johnson, 1996).

The singular values of $Q$ are equal to the positive roots of the non-trivial eigenvalues of

$$Q'Q = D_c^{-\frac{1}{2}}F'D_r^{-1}FD^{-\frac{1}{2}}_c.$$
Chapter 3. An important eigenvalue property encountered in correspondence analysis

Post-multiplication of $Q'Q$ by $D_c^{-\frac{1}{2}}1_p$ yields

$$D_c^{-\frac{1}{2}}F'D_r^{-1}FD_c^{-\frac{1}{2}}D_c^{-\frac{1}{2}}1_p = D_c^{-\frac{1}{2}}F'D_r^{-1}F1_p = D_c^{-\frac{1}{2}}F'1_n = D_c^{-\frac{1}{2}}c = D_c^\frac{1}{2}1_p,$$

where we have used (3.1). Hence, the vector

$$v = D_c^\frac{1}{2}1_p$$

is a positive (i.e. all its elements are greater than zero) eigenvector of $Q'Q$ corresponding to an eigenvalue, say $\lambda$, equal to one. Now, if $Q'Q$ is a nonnegative irreducible matrix, it follows immediately from the Frobenius theorem (e.g. Wielandt, 1950, Gantmacher, 1959 or Seneta, 1973), that $\lambda$ is the largest eigenvalue.

As Gower and Hand note, proofs of the Frobenius theorem are difficult, e.g. Wielandt (1950), Gantmacher (1959) or Seneta (1973). Moreover, as mentioned above, the Frobenius theorem requires that the matrix $Q'Q$ is nonnegative and irreducible. Clearly, as $F$ is nonnegative, $Q'Q$ is nonnegative, however, the second requirement, i.e. $Q'Q$ is irreducible, which is not mentioned in Gower and Hand (1996), implicitly imposes a restriction on $F$. In the case of correspondence analysis, where the matrix $F$ is a matrix of observed frequencies of co-occurrences, it is unlikely that the restriction does not hold. However, it is not difficult to construct a matrix $F$ with a sufficient number of zero entries such that $Q'Q$ is a reducible matrix causing the proof to break down.

A proof using a row stochastic matrix

Puntanen and Styan (1998) note that the singular values of $Q$ are equal to the positive square roots of the non-trivial eigenvalues of the so-called row stochastic matrix $S$, i.e. a nonnegative matrix with row sums equal to one, defined as

$$S = D_r^{-1}FD_c^{-1}F'.$$

It is then stated, without proof, that the eigenvalues of a row-stochastic matrix have a spectral radius of one, i.e.

$$\max_i |\lambda_i| = 1,$$
whence the result follows immediately.

It is not difficult to see that the proof of section 3.3.4 can immediately be applied to $S$ to prove this property of stochastic matrices. Instead, a reference is given to Marcus and Minc (1992). However, Marcus and Minc do not provide a proof but give a reference to Gantmacher (1959). Of essential importance in Gantmacher's proof is the following inequality:

$$\min_i \sum_{j=1}^{n} s_{ij} \leq \lambda_{\text{max}} \leq \max_i \sum_{j=1}^{n} s_{ij}, \quad (3.30)$$

which is proven for irreducible nonnegative matrices (using the Frobenius theorem) and it is said to hold for any nonnegative matrix. The straightforward proof using matrix norms is apparently not known to Gantmacher (1959). In fact, in Gantmacher (1959) matrix norms do not occur whilst in the second and third editions, e.g. Gantmacher (1967), matrix norms are treated in a subsequent chapter.

A proof using Geršgorin's theorem

Graffelman (1998) provides a proof that is also based on inequality (3.30). For a proof of the inequality Graffelman refers to Barbolla and Sanz (1998). Their proof appears to be for positive rather than for nonnegative matrices.

For the right-hand side inequality, i.e. the upperbound (which is in fact the only relevant bound) Barbolla and Sanz use Geršgorin's theorem. A proof of Geršgorin's theorem can be found in standard texts on matrices such as Horn and Johnson (1996).

Apparentely the proofs by Puntanen and Styan (1998), and Graffelman (1998) employ the same inequality for which they provide different proofs (or, in fact, references to proofs). Both of these proofs involve advanced matrix algebra such as the Frobenius theorem and Geršgorin's theorem. As we saw in section 3.3.4 the relevant inequality, i.e. right-hand side of (3.30), can be proven in a much faster and easier manner.