RB 001-'77

TUCKALS 2

A principal component analysis of three mode data

Pieter Kroonenberg
DEPARTMENT OF PEDAGOGY

Jan de Leeuw

MARCH 1977
DEPARTMENT OF DATATHEORY/FACULTY OF SOCIAL SCIENCES
UNIVERSITY OF LEYDEN — THE NETHERLANDS

RB 001-'77

TUCKALS2

A principal component analysis of three mode data

Jan de Leeuw
Pieter Kroonenberg

DEPARTMENT OF DATATHEORY/FACULTY OF SOCIAL SIENCES
UNIVERSITY OF LEYDEN - THE NETHERLANDS

ERRATA

(belonging to Kroonenberg, De Leeuw, TUCKALS2)

p. 1-2	1.16	for	vein about	read	vein we will talk about
p. 2-2,2-3		"	minimalisation	"	minimisation
p. 3-9	1.20	"	p are σ	"	p and σ are
p. 3-13		delete all lines upto point b.1.			
p. 3-16	1.21	for	$s > p$ and $t > q$	read	$s \geq p$ and $t \geq q$
p. 3-17	1.30	"	$\sum^{n} Z_k Z_k' = A$	"	$\sum_{k=1}^{n} Z_k Z_k' = A$
	1.18,19	"	expressions are	"	expression is
	1.8	"	q=eank	"	q=rank
p. 3-18	1.5	"	$s > p$ and $t > q$	"	$s \geq p$ and $t \geq q$
p. 5-4	1.22	"	data showed	"	components showed

p. 5-6,5-6[a] ARP-CHU-KVP group should be labeled : christian democrats

CPN -PSP-PvdA group should be labeled : left

p. A1-3 after 1.8 before "D substep" insert:

$$L_{i+1} = V_i (V_i' V_i)^{-\frac{1}{2}}$$

INDEX

Introduction

INTRODUCTION

In this paper a principal component analysis for three mode data is presented. The model used, christened the Tucker 2 model, is an asymmetric variant of the general three mode model of Tucker. The difference between this model and the general Tucker model is that the principal components of only the first two rather than all three modes appear in the model.

In the first chapter we will present some notation and terminology, as well as some results about traces, and the singular value decomposition,
in the second chapter a number of principal component models is discussed,
in the third chapter the Tucker 2 model and its best approximation are presented, and their properties are proved in detail,
in the fourth chapter an algorithm (TUCKALS2) to solve the best approximation problem of the Tucker 2 model is provided and its converged is investigated,
finally in the fifth chapter four examples are presented, two numerically oriented and two 'real world' examples taken from the Dutch political scene of 1968.
In the first appendix an orthonormalisation procedure for rotating the core matrix based on the ALS principle is outlined,
and in the second appendix a procedure is described for making a joint plot of X and Y in the decomposition Z = XY'.

CHAPTER ONE

Notation, terminology and some theorems

1.1. Matrices

All matrices used in this paper are real and in general the number of rows will be larger than or equal to the number of columns.
The following classes of matrices will be used throughout:

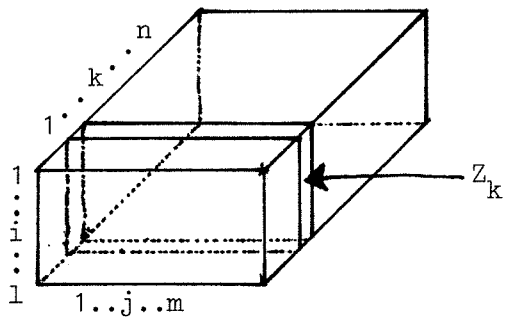
$R^{n\times m}$: the class of all real n×m matrices

$R_p^{n\times m}$: the class of all real n×m matrices of rank p

$K^{n\times m}$: the class of all columnwise orthonormal matrices, i.e. if $U \in K^{n\times m}$ then $U'U = I_m$, but $UU' = I_n$ if and only if n = m; the rank of any $U \in K^{n\times m}$ is m.

$D^{n\times m}$: the class of all n×m diagonal matrices; $E \in D^{n\times m}$ is a 'truly' diagonal matrix if n=m; if n < m then $E = (F\ O)$ with $F \in D^{n\times n}$ and O is a matrix with zeroes; if n > m then $E = \begin{pmatrix} F \\ O \end{pmatrix}$ with $F \in D^{m\times m}$.

A l×m×n block matrix Z is defined as the collection of elements:
$\{z_{ijk} \mid i=1,\ldots,l;j=1,\ldots,m;k=1,\ldots n\}$.
These elements can be thought to be placed in a three dimensional block with the index i running along the vertical axis, the index j running along the horizontal axis, and the index k along the 'depth' axis.



We will also view a block matrix as a collection of n 'normal' matrices Z_k which have been placed behind one another. A Z_k will be called a "frontal

plane" of the block matrix.Following this interpretation of a block matrix Z we will use also the notation: $Z = (Z_1, Z_2, \ldots Z_n)$. An element z_{ijk} of Z is the (i,j)-th element of the frontal plane Z_k. The class of all real block matrices of size $l \times m \times n$ will be written as $R^{l \times m \times n}$.

1.2. Eigenvalues and eigenvectors

When in the following we speak about an eigenvector matrix A we always assume that the m columns are different, independent and orthonormal, i.e. $A \in K^{n \times m}$. Moreover, the columns will be arranged in such a way that the first column is associated with the largest eigenvalue of A, the second column with the second largest eigenvalue, etc. If we speak about an eigenvalue matrix we will always assume that the eigenvalues are arranged on the (main) diagonal in decreasing size. We will refer to the "eigenproblem" in stead of the "eigenvector - eigenvalue problem", and in a similar vein about the "eigendecomposition", the"eigenfactorisation" and an "eigentechnique".

1.3 Traces

The following results regarding traces will be used:
Let $A \in R^{n \times m}$, $B \in R^{n \times m}$, $C \in R^{m \times n}$, $U \in K^{m \times n}$, then

$$\text{Tr } A = \sum_{i=1}^{m} a_{ii} \qquad \text{Tr } A'A = \sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij}^2$$

$$\text{Tr } AC = \text{Tr } CA \qquad \text{Tr } (A + B) = \text{Tr } A + \text{Tr } B$$

$$\text{Tr } UAU' = \text{Tr } U'UA = \text{Tr } A \qquad ||A||^2 = \text{Tr } A'A \text{ , where } ||\cdot|| \text{ is the Euclidean norm.}$$

1.4. Singular value decomposition

Let $A \in R_p^{m \times n}$. Then there exist matrices $U \in K^{m \times m}$ and $V \in K^{n \times n}$, such that $A = U\Lambda V'$ with $\Lambda \in D^{m \times n}$, and $\lambda_1 \geq \lambda_2 \geq \ldots . \lambda_p > 0 = \lambda_{p+1} = \ldots\ldots = \lambda_n$.

$U\Lambda V'$ is called the singular value decomposition of A, the columns of U are called the left singular vectors of A, the columns of V are called the right singular vectors of A, and the λ_i i=1,...,n are called the singular values of A.

For a proof of the existence theorem see, for instance, Ben Israel and Greville(1974,p.233).

The right singular vectors of A are the eigenvectors of A'A, the left singular vectors of A are the eigenvectors of AA'. The non-zero eigenvalues of AA' and A'A are equal and they are the squares of the corresponding singular values.

CHAPTER TWO

Models in principal component analysis and their best approximations.

2.1. Introduction

The factorisation of a given matrix into two or more matrices, which have to satisfy certain constraints is a frequently occuring problem in multivariate analysis. The best known example of this is the eigen factorisation of a definite symmetric matrix. The models which form the basis of principal component analyses can also be formulated in terms of factorisations.

In most practical circumstances the data matrix Z which has to be factorised is of maximal column or row rank. The aim of principal component analysis is to represent the data in terms of an orthonormal basis of rank r, where r is preferably much smaller than the rank of Z. In such cases no exact factorisations exist and one will have to be satisfied with a 'best approximate factorisation'. A 'best approximate factorisation'

$X = A_1 A_2 \ldots A_k$ of Z can be defined as follows:

Definition 2.1.

Be $Z \varepsilon R_p^{n \times m}$

$X = A_1 A_2 \ldots A_k$ $(X \varepsilon R^{n \times m})$ is a *best approximate factorisation* (or a *best approximation*) of Z if for all $Y \varepsilon R^{n \times m}$

$$||Z - X|| \leq ||Z - Y||$$

and the A_i's fulfill certain constraints.

X is a *best rank r factorisation* of Z if X is a best approximate factorisation of Z and rank X = r ($1 \leq r \leq p$ and r is an integer)

Before entering into a discussion of some principal component models two remarks must be made regarding terminology.

In the sequel the word "mode" will be used to indicate a collection of indices by which the data can be classified. The scores of a number of persons on a battery of tests can be classified by the individuals as well as by the tests. Both determine a mode of the data.

As mentioned above by means of principal component analysis one attempts to represent the data in a lower dimensional space than the one determined by the original data. Or differently: one looks for 'idealised' variables, of which the original variables are linear combinations or the 'best possible' linear combinations. As an example one can think of a battery of achievement tests. The scores on these tests could be considered linear combinations of such idealised variables as intelligence, language ability and computational ability. Commonly the term for idealised variables is "components" and this term will be used in the sequel.

2.2. Two - mode models

A. The classical principal component model (CPCM)

The CPCM can be formulated as $Z = GH'$. G consists of the components of the first mode variables and H consists of the scores of the second mode variables on the same components.
The problem of the best rank r factorisation of the CPCM can be formalised as follows:

Let $Z \in R_p^{n\times m}$ be a data matrix and r an integer s.t. $1 \leq r \leq p$.
The best rank r factorisation for the CPCM is the solution of the minimalisation of:

$$\sigma(G,H) = \mathrm{Tr}\,(Z - GH')'(Z - GH') \text{ over all}$$
$$H \in K^{m\times r} \text{ and all}$$
$$G \in R^{n\times r} \text{ s.t. } G'G = E \text{ with } E \in D^{r\times r}$$
$$\text{and } e_1 \geq e_2 \geq \ldots \geq e_r.$$

The solution of this minimalisation problem is given by Keller (1962). He shows that if $\hat{H}$ is the eigenvector matrix belonging to the largest r eigenvalues of $Z'Z$, $\hat{G} = Z\hat{H}$ and E the eigenvalue matrix of $Z'Z$, then $\hat{G}$ and

$\hat{H}$ are a solution of the minimalisation problem. If r = p an exact solution can be obtained, i.e. $\sigma = 0$. It should be noted that the error of approximation is $||Z - \hat{G}\hat{H}'|| = \left(\sum_{i=r+1}^{p} \lambda_i\right)^{\frac{1}{2}}$, where λ_i (i= r+1, ..., p) is the i-th largest eigenvalue of Z'Z is.

B. The Tucker - Levin two-mode model (TLTM)

Another two-mode model formulated by Tucker (1963) and Levin (1965) was used by Tucker (1963, 1964, 1966) as a basis for the generalisation to three-mode models, viz. the Tucker-Levin two-mode model.
The TLTM has the form $Z = GCH'$ in which G consists of the components of the first mode variables, H those of the second mode variables and finally C the interrelationship of the first and second mode components.

The problem of the best rank r factorisation of the TLTM can be formalised as follows:

Let $Z \in R_p^{n\times m}$ be a data matrix and r an integer

s.t. $1 \leq r \leq p$.

The best rank r factorisation for the TLTM is the solution of the minimalisation of:

$$\sigma(G,H,C) = \mathrm{Tr}\,(Z-GCH')'(Z-GCH') \quad \text{over all}$$
$$G \in K^{n\times r} \quad \text{and all}$$
$$H \in K^{m\times r} \quad \text{and all}$$
$$C \in R^{r\times r}$$

As is shown in for example Ben-Israel & Greville (1974), p. 233 ff. an exact solution exists if r = p, and this solution is the so called singular value decomposition (cf. chapter 1).

As mentioned above one is usually not interested in a factorisation with r = p, but in a factorisation with $r << p$. Ben-Israel & Greville show that the best rank r factorisation of Z is

$$G_{(r)}\, C_{(r)}\, H'_{(r)}.$$

The columns of $G_{(r)}$ are the eigenvectors belonging to the largest r eigenvalues of ZZ', the columns of $H_{(r)}$ are the eigenvectors belonging to the largest r eigenvalues of Z'Z and $C_{(r)}$ is the diagonalmatrix with the largest r singular values of Z.
If one specifies an additional constraint i.e. that $E_{(r)}$ is a diagonalmatrix, than the factorisation is unique, provided the largest r + 1 singular values are unequal.
The error of approximation is $||Z - G_{(r)}\, C_{(r)}\, H'_{(r)}|| = \left(\sum_{i=r+1}^{p} \mu_i^2\right)^{\frac{1}{2}}$
with μ_i is the i-th singular value of Z (i=r+1, ..., p).

2.3. The Tucker three-mode model (Tucker-model)

(The terminology and the description of the model in this section is largely taken from Tucker (1966)).

The three-mode model was first formulated by Tucker (1963), and it is a direct generalisation of the model described in the previous section.

The model is formulated as follows:

Let $Z = [z_{ijk}]$, $Z \varepsilon R^{l \times n \times m}$, be a three-mode (data) matrix, with

$$z_{ijk} = \sum_{s=1}^{S} \sum_{t=1}^{T} \sum_{u=1}^{U} g_{is}\, h_{jt}\, e_{ku}\, c_{stu}$$

i=1, ...,l ; j=1, ...m, k=1, ..., n.

The coefficients g_{is}, h_{jt}, and e_{ku} are entries in the matrices G, H, and E. The coefficients describe the scores on the original variables in terms of the components. The coefficients c_{stu} are entries in the three-mode matrix, C, which Tucker terms "core matrix". In the original three-mode matrix, Z, each cell represents a particular combination of categories from the original variables and the entry is a measure of the phenomenon whose value depends on the combination of categories.

In the same way each cell in the core matrix, C, represents a unique combination of categories from the components and the entry is a measure of the phenomenon of this combination of categories. The core matrix can be thought of as describing the basic relations existent in the measures of the phenomenon being observed. The matrices G, H, and E transform these relations from applying to the components of the three modes to applying to the original sets of variables. The interrelations among the elements of one of the sets of variables depend, in part, on the similarity of their relations to the components of the three modes, and, in part, to the relations in the core matrix.

In an unpublished manuscript de Leeuw poses the following with regard to the solution of the Tucker model:

$$C_1 = \left[\tilde{c}^{1}_{ss'} \right]; \; s,s' = 1, \ldots\ldots, S; \; C_1 \; \varepsilon \; D^{S \times S} \quad \text{with}$$

$$\tilde{c}^{1}_{ss'} = \sum_{t=1}^{T} \sum_{u=1}^{U} c_{stu} \; c_{s'tu}$$

$$C_2 = \left[\tilde{c}^{2}_{tt'} \right] \; ; \; t,t' = 1, \ldots., T \; ; \; C_2 \; \varepsilon \; D^{T \times T} \; \text{with}$$

$$\tilde{c}^{2}_{tt'} = \sum_{s=1}^{S} \sum_{u=1}^{U} c_{stu} \; c_{st'u}$$

$$C_3 = \left[\tilde{c}^{3}_{uu'} \right] \; ; \; u,u' = 1,\ldots., U \; ; \; C_3 \; \varepsilon \; D^{U \times U} \quad \text{with}$$

$$\tilde{c}^{3}_{uu'} = \sum_{s=1}^{S} \sum_{u=1}^{U} c_{stu} \; c_{stu'}$$

b. - Define $P = \left[p_{ii'}\right]$ with $p_{ii'} = \sum_{j=1}^{m} \sum_{k=1}^{n} z_{ijk}\, z_{i'jk}$

$$Q = \left[q_{jj'}\right] \text{ with } q_{jj'} = \sum_{i=1}^{l} \sum_{k=1}^{n} z_{ijk}\, z_{ij'k}$$

$$R = \left[r_{kk'}\right] \text{ with } r_{kk'} = \sum_{i=1}^{l} \sum_{j=1}^{m} z_{ijk}\, z_{ijk'}$$

- In addition, define $\hat{G}$, $\hat{H}$, and $\hat{E}$ to be the eigenvector matrices of P, Q, and R belonging to the non-zero eigenvalues.
- Finally, define V, W, and X to be:

$$V = \left[v_{ii'}\right] = \hat{G}\,\hat{G}'$$

$$W = \left[w_{jj'}\right] = \hat{H}\,\hat{H}'$$

$$X = \left[x_{kk'}\right] = \hat{E}\,\hat{E}'$$

c. The Tucker model with the above constraints (a) on G, H, E, and C has an exact factorisation, if the two following conditions are fulfilled.

- P, Q, and R have the rank S, T, and U respectively

$$-\sum_{i=1}^{l} \sum_{j=1}^{m} \sum_{k=1}^{n} z_{ijk}\, v_{ii'}\, w_{jj'}\, x_{kk'} = z_{i'j'k'}$$

d. An exact factorisation is determined by $\hat{G}$, $\hat{H}$, $\hat{E}$, and

and $\hat{C}$ with $\hat{C} = \left[\hat{c}_{stu}\right]$ and

$$\hat{c}_{stu} = \sum_{i=1}^{l} \sum_{j=1}^{m} \sum_{k=1}^{n} z_{ijk}\, \hat{g}_{is}\, \hat{h}_{jt}\, \hat{e}_{ku}$$

e. The factorisation is unique, if, in addition, the eigenvalues of both P, and Q, and R are different.

A best approximate factorisation for the three-mode model has not been worked out yet, but in principle analoguous proofs to those presented in the following chapters should be applicable.

2.4. The Tucker 2 model

Various authors (e.g. Carroll & Chang (1970), (1972), Harshman (1970), Israelsson (1969), Jennrich (1972) have proposed other three-mode models, all of which can be seen as special cases of the Tucker model.
In the sequel we will consider one such model, proposed by Israelsson (1969), Jennrich (1972) and Carrol & Chang (1970).
We will refer to this model as the Tucker 2 model.

The formulation of the model is as follows:

Let $Z = [z_{ijk}]$, $Z \in R^{l \times m \times n}$, be a three-mode matrix

with

$$\hat{z}_{ijk} = \sum_{s=1}^{S} \sum_{t=1}^{T} g_{is}\, h_{jt}\, c_{stk} \qquad i=1,\ldots,l \; ; \; j=1,\ldots,m; \; k=1,\ldots,n$$

In the notation introduced in chapter 1 this can be written as:

$$Z = (Z_1 , Z_2 , \ldots\ldots , Z_n); \; Z_k \in R^{l \times m} \quad k = 1,\ldots, n \text{ and}$$

$$Z_k = G\, C_k\, H'$$

in which g_{is} and h_{jt} are the entries in G and H respectively

and c_{stk} the entries of C_k $k = 1,\ldots\ldots, n$.

The interpretation of the core matrix $C = (C_1, C_2, \ldots\ldots, C_n)$, G, and H is identical to that in the general three-mode model.
It is readily seen that the Tucker model reduces to the Tucker 2 model if E is the identity matrix.

In the Tucker 2 model only components are computed over two of the three modes. In that sense the Tucker 2 model is an asymmetric variant of the general Tucker model.

In the following chapter we will give a solution of the Tucker 2 model and proofs of its correctness, as well as present an algorithm to find such a solution.

CHAPTER THREE

The best approximate and exact factorisation of the Tucker 2 model.

In this chapter the exact factorisation and the best approximate factorisation of the Tucker 2 model will be discussed.

Minimization problem (I)

Be $Z = (Z_1, Z_2, \ldots, Z_n)$ with $Z_k \in R^{l \times m}$ $(k=1,\ldots,n)$ and be s and t integers s.t. $1 \leq s \leq l$ and $1 \leq t \leq m$. The best approximate factorisation of the Tucker 2 model is the solution of the minimization of:

$$\sigma(G, H, C) = \sum_{k=1}^{n} \mathrm{Tr}\,(Z_k - GC_kH')'(Z_k - GC_kH') \quad (3.1)$$

over all

$$G \in K^{l \times s}\ ,\ H \in K^{m \times t} \text{ and } C = (C_1, C_2, \ldots, C_n) \text{ with}$$

$$C_k \in R^{s \times t}$$

This minimization problem will be referred to as : (I)
A triplet $(\hat{G}, \hat{H}, \hat{C})$ is a solution of (I), if $\hat{G}$, $\hat{H}$, $\hat{C}$ satisfy the above constraints and if σ attains a minimum for these $\hat{G}$, $\hat{H}$ and $\hat{C}$.
The solution will be called exact if $\sigma = 0$, in other words if an exact factorisation exists.

3.1. There always exists a best approximate factorisation for the Tucker 2 model.

In this paragraph it will be shown, that (I) always has a solution. In order to prove this (in Theorem 2.6) we need a few preliminary lemmas.

Lemma 3.1

Be f a real function defined on T with

$$T = \{(G, H, C) \mid G \in K^{l \times s},\ H \in K^{m \times t},\ C \in R^{s \times t \times n}\},$$

then it follows that

$$\inf_{G,H,C} f(G,H,C) = \inf_{G,H} \left[\inf_{C} f(G,H,C)\right] \qquad (3.2)$$

The infima are to be taken over all G and/or H and/or C as defined in T.

Proof:

It will be shown that the left hand side of (3.2) is greater or equal to the right hand side and vice versa, which implies that (3.2.) is true.

$$f(G,H,C) \geq \inf_{G,H} \left[\inf_{C} f(G,H,C)\right] \quad \text{for all } G,H,C;$$

thus also:

$$\inf_{G,H,C} f(G,H,C) \geq \inf_{G,H} \left[\inf_{C} f(G,H,C)\right] \qquad (3.3.)$$

As the infimum is a lower bound, it follows that

$$\inf_{G,H,C} f(G,H,C) \leq f(G,H,C) \text{ for all } G,H,C;$$

this implies that

$$\inf_{G,H,C} f(G,H,C) \leq \inf_{C} f(G,H,C) \quad \text{for all } G,H;$$

and finally it follows that

$$\inf_{G,H,C} f(G,H,C) \leq \inf_{G,H} \left[\inf_{C} f(G,H,C)\right] \qquad (3.4)$$

combining (3.3) and (3.4) completes the proof.

▼▼

The following lemma is a special case of a more general theorem by Penrose (1955).

Lemma 3.2.

Be $G \in K^{l\times s}$, $H \in K^{m\times t}$ and $Z \in R^{l\times m}$ with arbitrary but fixed entries

a. $||GPH' + Z - GG'ZHH'||^2 = ||GPH'||^2 + ||Z - GG'ZHH'||^2$ (3.5.

is valid identity for all $P \in R^{s\times t}$

b. The solution of the equation

$$GCH' = Z$$

has a unique best approximation, viz.:

$$\hat{C} = G'ZH$$

c. The solution $\hat{C} = G'ZH$ is exact if and only if $Z = GG'ZHH'$

Proof:

a. $Tr\ (GPH' + Z - GG'ZHH')'(GPH' + Z - GG'ZHH') =$

$Tr\ (GPH')'(GPH') + Tr(Z - GG'ZHH')'(Z - GG'ZHH')$

$+ Tr(HP'G'Z) - Tr(HP'G'GG'ZHH') + Tr(Z'GPH').$

$- Tr(HH'Z'GG'GPH') =$

$Tr(GPH')'(GPH') + Tr(Z - GG'ZHH')'(Z - GG'ZHH'),$

because $Tr(HP'G'GG'ZHH') = Tr(HH'HP'G'Z) = Tr(HP'G'Z)$ and $Tr(HH'Z'GG'GPH') = Tr(Z'GPH')$ (see chapter 1, section 3)

The identity (3.5) follows from the fact that $||X||^2 = Tr(X'X)$ for all real X.

b. Now consider:

$$||GCH' - Z||^2 = ||G(C - G'ZH)H' - Z + GG'ZHH'||^2$$

$$= ||G(C - G'ZH)H'||^2 + ||GG'ZHH' - Z||^2 \text{ (see a.)}$$

$$= ||GCH' - GG'ZHH'||^2 + ||GG'ZHH' - Z||^2 \geq$$

$$\geq ||GG'ZHH' - Z||^2,$$

with equality if and only if $GCH' = GG'ZHH'$ or

$$C = G'ZH$$

Because there is only one $\hat{C} \in R^{s \times t}$ such that $\hat{C} = G'ZH$ this $\hat{C}$ is the unique best approximation of the solution of $GCH' = Z$

c. The solution, $\hat{C} = G'ZH$, of $Z = GCH'$ is exact if and only if $||Z - GCH'|| = 0$

Let $\hat{C} = G'ZH$ be an exact solution, then

$$||Z - G\hat{C}H'|| = ||Z - GG'ZHH'|| = 0 \Longrightarrow Z = GG'ZHH'$$

Conversely let $Z = GG'ZHH'$ and $\hat{C} = G'ZH$ be the best approximate solution of $Z = GCH'$, then $||Z - G\hat{C}H'|| = || GG'ZHH' - GG'ZHH'|| = 0$ thus $\hat{C} = G'ZH$ is the exact solution of $Z = GCH'$.

▼▼

We can now use the above lemma to show that the solution of (I) depends only on G and H, because the unique best C is completely determined by G, H and Z.

Lemma 3.3

Be $Z = (Z_1, Z_2, \ldots, Z_n)$ with $Z_k \in R^{l \times m}$ $(k = 1, \ldots, n)$, and the minimization problem (I) given.
There exists a unique $C = (C_1, C_2, \ldots, C_n)$ with $C_k \in R^{s \times t}$ $(k = 1, \ldots, n)$ such that $\inf_C \sigma(G,H,C)$ exists and is attained.

This unique C is:

$$\hat{C} = (\hat{C}_1, \hat{C}_2, \ldots, \hat{C}_n) \text{ with } \hat{C}_k = G'Z_kH \quad (k = 1, \ldots, n)$$

Proof:

$\inf_C \sigma(G,H,C)$ can be rewritten as

$$\inf_C \sigma(G,H,C) = \inf_C \sum_{k=1}^{n} \mathrm{Tr}\,(Z_k - GC_kH')'(Z_k - GC_kH')$$

$$= \inf_C \sum_{k=1}^{n} ||Z_k - GC_kH'||^2$$

$$= \sum_{k=1}^{k} \inf_{C_k} ||Z_k - GC_kH'||^2$$

The last equality follows because all terms of the summation are nonnegative.
According to lemma 3.2. it follows that, for fixed G and H, $\hat{C}_k = G'Z_kH$ is the unique best approximation of the solution of the equation $GC_kH' = Z_k$ $(k = 1,\ldots,n)$.

Thus:

$$\inf_C \sigma(G,H,C) = \sum_{k=1}^{n} \inf_{C_k} ||Z_k - GC_kH'||^2 =$$

$$= \sum_{k=1}^{n} ||Z_k - G\hat{C}_kH'||^2 =$$

$$= \sigma(G,H,\hat{C})$$

with $\hat{C} = (\hat{C}_1, \hat{C}_2, \ldots, \hat{C}_n)$ ▼▼

Because the minimization problem (I) is solely determined by G and H, we can redefine the problem only in terms of G, and H, which is done in the following lemma.

Lemma 3.4.

Be $Z = (Z_1, Z_2, \ldots, Z_n)$ with $Z_k \in R^{l\times m}$ $(k = 1,\ldots,n)$ and the minimization problem (I) given.

Be S defined as $S = \{(G,H) \mid G \in K^{l\times s}, H \in K^{m\times t}\}$

a. Be p defined as:

$$p(G,H) = \sum_{k=1}^{n} \operatorname{Tr} Z_k' GG' Z_k HH' \text{ for all } (G,H) \in S \qquad (3.6)$$

If p and σ are bounded functions, then (I) is equivalent to a maximization of p over S.

b. $p(G,H) = \sum_{k=1}^{n} \operatorname{Tr} H'Z_k GG' Z_k H = \sum_{k=1}^{n} \operatorname{Tr} G'Z_k HH'Z_k G$

c. $\tilde{\sigma} = 0 \Leftrightarrow p(G,H) = \sum_{k=1}^{n} \operatorname{Tr} Z_k'Z_k$

Proof:

Be p defined as above.

a. According to lemma 3.1:

$$\inf_{G,H,C} \sigma(G,H,C) = \inf_{G,H} \left[\inf_{C} \sigma(G,H,C)\right] .$$

According to lemma 3.3

$$\inf_{C} \sigma(G,H,C) = \sigma(G,H,\hat{C}) \text{ with } \hat{C} = (\hat{C}_1, \hat{C}_2, \ldots, \hat{C}_n)$$
$$\text{and } \hat{C}_k = G'Z_k H \ (k=1,\ldots,n)$$

Thus:

$$\inf_{G,H,C} \sigma(G,H,C) = \inf_{G,H} \sigma(G,H,\hat{C}) = \inf_{G,H} \tilde{\sigma}(G,H) \text{ with}$$

$$\tilde{\sigma}(G,H) = \sum_{k=1}^{n} \operatorname{Tr}(Z_k - G\hat{C}_k H')'(Z_k - G\hat{C}_k H') =$$

$$= \sum_{k=1}^{n} \operatorname{Tr}(Z_k - GG'Z_k HH')'(Z_k - GG'Z_k HH'). \qquad (3.7)$$

The latter expression can be rewritten as:

$$\tilde{\sigma}(G,H) = \sum_{k=1}^{n} \operatorname{Tr} Z_k'Z_k - \sum_{k=1}^{n} \operatorname{Tr} Z_k'GG'Z_k HH'$$

$$= \sum_{k=1}^{n} \operatorname{Tr} Z_k'Z_k - p(G,H).$$

Suppose p and σ (and thus $\tilde{\sigma}$) are bounded functions, then:

$$\inf_{G,H} \tilde{\sigma}(G,H) = \inf_{G,H} \left(\sum_{k=1}^{n} \operatorname{Tr} Z_k' Z_k - p(G,H) \right)$$

$$= \sum_{k=1}^{n} \operatorname{Tr} Z_k' Z_k - \sup_{G,H} p(G,H)$$

Thus if p and σ are bounded (I) is equivalent to the maximization of p over S.

b. In chapter 1, section 3 it is shown, that:

$$\operatorname{Tr} AB = \operatorname{Tr} BA \quad \text{for all } A, B' \in R^{n\times m}.$$

Thus by choosing for each k first $A = Z_k' G G' Z_k H$ and $B = H'$ it follows that:

$$p(G,H) = \sum_{k=1}^{n} \operatorname{Tr} H' Z_k' G G' Z_k H.$$

Choose subsequently for each k $A = H' Z_k G$ and $B = A'$, then it follows that:

$$p(G,H) = \sum_{k=1}^{n} \operatorname{Tr} G' Z_k H H' Z_k' G$$

c. The statement follows directly from the implications below:

$$\tilde{\sigma} = 0 \text{ iff } \sum_{k=1}^{n} \operatorname{Tr} Z_k' Z_k - p(G,H) = 0 \text{ iff } p(G,H) = \sum_{k=1}^{n} \operatorname{Tr} Z_k' Z_k$$

▼▼

As a final step before we can prove the main theorem of this section we will prove that p is a bounded, continuous function on a compact set.

Lemma 3.5

Be $Z = (Z_1, Z_2, \ldots, Z_n)$ with $Z_k \in R^{l\times m}$ $(k=1,\ldots,n)$, the set $S = \{(G,H) \mid G \in K^{l\times s}, H \in K^{m\times t}\}$ and the maximization problem of p over S given.

Then follows that

a. S is compact

b. p is continuous on S

c. p is bounded on S

Proof:

a. Define $||A||_a = \sup_{|x|=1} |Ax|$ as the matrix norm to

$|x| = \sqrt{(x,x)} = \sqrt{\sum_{i=1}^{m} x_i^2}$ with $x \varepsilon R^m$ and $A \varepsilon R^{n \times m}$.

The matrix norm $||A||_a = \sqrt{\lambda_1(A'A)}$, with $\lambda_1(A'A)$ is the largest eigenvalue of A'A (cf. Schwarz et al., 1968, p.31)

For all $G \varepsilon K^{l \times s}$ the matrix norm is equal to 1, as $G'G = I_s$. Thus $K^{l \times s}$ is uniformly bounded.

Be $G_1, G_2, \ldots\ldots$, ($G_i \varepsilon K^{l \times s}$; i=1, 2, ...) an arbitrary convergent sequence with limit $\tilde{G}$. Thus

$$\lim_{i \to \infty} ||G_i - \tilde{G}||_a = 0$$

For all $x \varepsilon R^s$ this gives:

$$||G_i x - \tilde{G}x||_a \leq ||G_i - \tilde{G}||_a . |x| \to 0 \text{ if } i \to \infty ,$$

and thus that:

$$\lim_{i \to \infty} G_i x = \tilde{G} x.$$

From this it follows that:

$$(\tilde{G}x, \tilde{G}x) = \lim_{i \to \infty} (G_i x, G_i x) = \lim_{i \to \infty} (x,x) = (x,x),$$

because for all $G \varepsilon K^{l \times s}$ $(Gx, Gy) = (x,y)$ for all $x, y \varepsilon R^s$.

This gives that $\tilde{G} \varepsilon K^{l \times s}$ and consequently $K^{l \times s}$ is closed.

As $K^{l \times s}$ is a subset of R^k with $k = l \times s$, $K^{l \times s}$ is a closed and bounded subset of a finite dimensional real space, and thus $K^{l \times s}$ is compact.

Analoguously it follows that $K^{m\times t}$ is compact and consequently their direct product, i.e. S, is compact.

b. As $||G||_a = 1$ for all $G \in K^{l\times s}$, it follows that G is uniformly continuous on R^s, because

$$|Gx - Gy| \leq ||G||_a . |x - y| = |x - y| \text{ for each pair } x, y \in R^s.$$

Analoguously it follows that H is uniformly continuous on R^t
Because p is comprised of continuous transformations, p itself is continuous on S.

c. Combining a. and b. gives that p is bounded as are all continuous functions on compact sets. ▼▼

We are now ready for the theorem that the minimization problem (I) always has a solution.

Theorem 3.6

Be $Z = (Z_1, Z_2 \ldots\ldots, Z_n)$ with $Z_k \in R^{l\times m}$ $(k= 1, \ldots., n)$ and the minimization problem (I) given.
The minimization problem (I) has a solution, i.e. there exists a triplet $(\hat{G}, \hat{H}, \hat{C})$ s.t.

$$\sigma(\hat{G}, \hat{H}, \hat{C}) = \min \sigma(G, H, C) \text{ over all G, H and C.}$$

Proof:

From the definition of the trace, it follows that p are σ nonnegative.
As p is nonnegative it follows:

$$\sigma(G,H,C) = \sum_{k=1}^{n} \operatorname{Tr} Z_k' Z_k - p(G,H) \leq \sum_{k=1}^{n} \operatorname{Tr} Z_k' Z_k < \infty ,$$

thus σ is bounded.
With the boundedness of p and σ the conditions of lemma 3.4. are fulfilled and thus (I) and the maximization of p over S are equivalent.

Lemma 3.5. shows us that p is a continuous bounded function on a compact subset of a metric space and consequently has a supremum and attains it. (cf. e.g. Rudin (1953), p. 177) ▼▼

3.2. The best approximate factorisation of the Tucker 2 model.

In this section we will provide the best approximate factorisation of the Tucker 2 model.

From the proof of theorem 3.6. it followed that the best approximate factorisation of the Tucker 2 model is equivalent to the solution of the maxim isation problem of p (as defined in lemma 3.4) over $S = \{(G,H) \mid G \in K^{l\times s}$ and $H \in K^{m\times t}\}$. It will therefore be sufficient to concentrate on finding a solution for this maximization problem

Obviously a maximum of p is a stationary point of p, and in this case we will define a stationary point as follows:

Definition 3.7.

Let $T = \{(X,Y) \mid X \in K^{n\times m}$ and $Y \in K^{k\times l}\}$ and let f be a real continuous differentiable function on T, then $(X,Y) \in T$ is a stationary point of f if (X,Y) is a solution of the stationary equations:

$$\frac{\partial}{\partial X}\left[f(X,Y) - \mathrm{Tr}M(X'X - I_m)\right] = 0$$

$$\frac{\partial}{\partial Y}\left[f(X,Y) - \mathrm{Tr}\ N(Y'Y - I_l)\right] = 0$$

with M and N matrices of Lagrange multiplyers.

The following theorem provides the solution of the maximization problem of p.

Theorem 3.8.

Be $Z=(Z_1, Z_2, \ldots\ldots, Z_n)$ with $Z_k \in R^{l\times m}$ $(k= 1, \ldots\ldots, n)$ and $s \leq l$ and $t \leq m$ given and be p and S defined as in section 3.2. Let, in addition, P and Q be defined as:

$$P(G) = \sum_{k=1}^{n} Z_k' GG'Z_k \text{ and } Q(H) = \sum_{k=1}^{n} Z_k HH'Z_k' \text{ and finally be}$$

U and V defined as follows: U is any eigenvectormatrix of Q(V) and Y is any eigenvectormatrix of P(U) and $(U,Y) \in S$.

Then:

a. $(\hat{G},\hat{H})$ is a stationary point of p if and only if $\hat{G}= U$ and $\hat{H}= V$ or orthonormal rotations thereof.

b. $(\hat{G},\hat{H}) \in S$ maximizes p if and only if their columns are the eigenvectors associated with the largest s and t eigenvalues of $Q(\hat{H})$ and $P(\hat{G})$, respectively, or orthonormal rotations thereof.

Proof:

a. 1. Let us first determine the stationary equations for p.

$$p(G,H) = \text{Tr } H' \left(\sum_{k=1}^{n} Z_k' GG'Z_k \right) H = \text{Tr } H'P(G)H$$

$$\frac{\partial}{\partial H} \left[\text{Tr } H'P(G)H - M(H'H - I_t) \right] = 2P(G)H - 2HM$$

which gives: $P(G)\hat{H} = \hat{H}\hat{M}$ and $\hat{H}'\hat{H} = I_t$ (3.7)

$$p(G,H) = \text{Tr } G' \left(\sum_{k=1}^{n} Z_k HH'Z_k' \right) G = \text{Tr } G' Q(H)G$$

$$\frac{\partial}{\partial G} \left[\text{Tr } G'Q(H)G - N(G'G - I_s) \right] = 2 Q(H)G - 2GN$$

which gives: $Q(H)\hat{G} = \hat{G}\hat{N}$ and $\hat{G}'\hat{G} = I_s$ (3.8)

This leads to two sets equations from which $\hat{G}, \hat{H}, \hat{M}, \hat{N}$ must be solved:

$$P(\hat{G})\hat{H} = \hat{H}\hat{M} \quad ; \hat{M}'\hat{H} = I_t$$

$$Q(\hat{H})\hat{G} = \hat{G}\hat{N} \quad ; \hat{G}'\hat{G} = I_s$$

To simplify the notation we will drop the carets and just write P and Q for $P(\hat{G})$ and $Q(\hat{H})$.

Thus:

$$PH = HM \quad ; H'H = I_t \tag{3.10}$$

$$QG = GN \quad ; G'G = I_s \tag{3.11}$$

Let us note that M can be chosen to be symmetric because the condition $h_i'h_j = \delta_{ij}$ is identical to the condition $h_j'h_i = \delta_{ji}$, where h_i is the ith column of H. The analoguous result holds for N.

2. Let G, H, M, N be a solution of (3.10) and (3.11)

   First we will prove that: if H, M is a solution of (3.10) then there exists a $F \in K^{t\times t}$ s.t. $H = VF'$ and $M = F\Lambda F'$ with Λ an eigenvalue matrix of P and V the associated eigenvector matrix.

   As H, M is a solution of (3.10) it follows from PH = HM that $M = H'PH$. This goes to show that M is symmetric and positieve semi-definite and there exists a $F \in K^{t\times t}$ s.t. $M = F\Lambda F'$ with $\Lambda \in D^{t\times t}$. Substitute this into (3.10):

   $PH = HF\Lambda F'$. After postmultiplication with F we get: $P(HF) = (HF)\Lambda$. By defining $V = HF$ it follows that Λ is an eigenvalue matrix of P and Λ is the associated eigenvector matrix. Analoguously it follows that if G, N is a solution of (3.11), then there exists an $E \in K^{s\times s}$, s.t. $G = UE'$ and $N = E\tilde{\Lambda}E'$ with $\tilde{\Lambda}$ an eigenvalue matrix of Q and U the associated eigenvector matrix.

   Thus if (G, H) is a stationary point of p then G = U and H = V or orthonormal rotations thereof.

3. Conversely, if we let U, V, Λ, $\tilde{\Lambda}$ be the eigenvector and eigenvalue matrices of P and Q, then they (as well as orthonormal rotations of U and V) form a solution of (3.10) and (3.11).

   Let $G = UE'$, $N = E\tilde{\Lambda}E'$, $H = VF'$ and $M = F\Lambda F'$ with $E \in K^{s\times s}$ and $F \in K^{t\times t}$ arbitrary, then $H'H = I_t$ and $G'G = I_s$.

   And the following string of implications holds:

   $$QU = U\tilde{\Lambda} \Rightarrow QGE' = GE'\tilde{\Lambda} \Rightarrow QG = GE'\tilde{\Lambda}E \Rightarrow$$

   $$QG = GN' \Rightarrow QG = GN.$$

   Analoguously it follows from

   $PV = V\Lambda$ that $PH = HM$.

   Thus, if G = U and H = V or orthonormal rotations thereof, (G,H) is a stationary point of p.

Analoguously it follows that if G, N is a solution of (3.11), then there exists an $E \in K^{s\times s}$, s.t. $G = UE'$ and $N = E\tilde{\Lambda}E'$ with $\tilde{\Lambda}$ an eigenvalue matrix of Q and U the associated eigenvector matrix.

Thus if (G, H) is a stationary point of p then G = U and H = V or orthonormal rotations thereof.

Conversely if we let U, V, Λ, $\tilde{\Lambda}$ be the eigenvector and eigenvalue matrices of P and Q, then they (as well as orthonormal rotations of U and V) form a solution of (3.10) and (3.11).

Let $G = UE'$, $N = E\tilde{\Lambda}E'$, $H = VF'$ and $M = F\Lambda F'$ with $E \in K^{s\times s}$ and $F \in K^{t\times t}$ arbitrary, then $H'H = I_t$ and $G'G = I_s$.

And the following string of implications holds:

$QU = U\tilde{\Lambda} \rightarrow QGE' = GE'\tilde{\Lambda} \rightarrow QG = GE'\tilde{\Lambda}E \rightarrow$

$QG = GN' \rightarrow QG = GN$

Analoguously it follows from

$PV = V\Lambda$ that $PH = HM$.

Thus, if G = U and H = V or orthonormal rotations thereof, (G, H) is a stationary point of p

b. 1. In part a. it was shown that the collection of stationary points of p is the set $W = \{(G,H) \mid (G,H) \in S$; G is an eigenvector matrix of Q(H), H is an eigenvectormatrix of P(G) or orthonormal rotations there of.$\}$

In theorem 3.6 it was shown that there exist $(G,H) \in S$ s.t. p attains a maximum. From part a. it follows that this maximum will be attained for some $(G,H) \in W$.

2. We will investigate now the value of this maximum.

Let $(\hat{U}, \hat{V})$ be the point for which p attains its maximum

$$p(\hat{U}, \hat{V}) = \max_U p(U, \hat{V}) = \max_U \operatorname{Tr} U'Q(\hat{V})U =$$

$$= \max_\lambda \sum_{i=1}^{s} \lambda_i$$ (with the latter maximum taken over all possible ways to combine s of the total of l eigenvalues of $Q(\hat{V})$)

$= \sum_{i=1}^{s} \hat{\lambda}_i$ here $\hat{\lambda}_i$ (i=1,.....,s) are Λ the largest s eigenvalues of $Q(\hat{V})$).

Thus $\hat{U}$ must be the associated eigenvector matrix. Analoguously $\hat{V}$ must be the eigenvector matrix associated with the largest eigenvalues of $P(\hat{U})$. Let us note that $p(\hat{G}, \hat{H}) = \sum_{i=1}^{s} \hat{\lambda}_i = \sum_{j=1}^{t} \hat{\mu}_j$ ($\hat{\mu}_j$ is defined analoguously to $\hat{\lambda}_i$)

3. Conversely if $\hat{U}$, $\hat{V}$ are the eigenvector matrices associated with the largest s and t eigenvalues of $Q(\hat{V})$ and $P(\hat{U})$, then $p(\hat{U}, \hat{V}) = \sum_{i=1}^{s} \hat{\lambda}_i = \sum_{j=1}^{t} \hat{\mu}$

   $= \max_{(U,V)} p(U,V)$ and thus $(\hat{U}, \hat{V})$ maximizes p.

4. Be $E \in K^{s\times s}$ and $F \in K^{t\times t}$ arbitrary, and be U and V defined as above.

From

$$p(UE, UF) = \sum_{k=1}^{n} \mathrm{Tr}\ (UE)'\ Z_k (VF)\ (VF)'\ Z_k'\ (UE) =$$

$$= \sum_{k=1}^{n} \mathrm{Tr}\ E'U'Z_k\ VFF'V'\ Z_k UE =$$

$$= \sum_{k=1}^{n} \mathrm{Tr}\ U'Z_k VV'Z_k U = p(U,V)$$ it follows that if (U,V) maximizes p then orthonormal rotations of U and V do so as well, and vice versa.

This completes the proof.

▼▼

3.3. The exact factorisation of the Tucker 2 model.

In this section we will examine the conditions for the exact factorisation of the Tucker 2 model.

As in the previous section we first prove a few preliminary lemmas

Lemma 3.9.

Let the minimization problem (I) be given, then the following three statements are equivalent

a) $(\hat{G}, \hat{H}, \hat{C})$ with $\hat{C} = (\hat{C}_1, \hat{C}_2, \ldots, \hat{C}_n)$ and $C_k = \hat{G}'Z_kH$ $(k=1,\ldots,n)$ is an exact solution of (I), i.e. $\sigma(\hat{G}, \hat{H}, \hat{C}) = 0$

b) $\tilde{\sigma}(\hat{G}, \hat{H}) = 0$ ($\tilde{\sigma}$ defined as in (3.7))

c) $GG'Z_kHH = Z_k$ $(k = 1, \ldots, n)$

Proof: That (a) is true iff (b) is true follows from

$$\sigma(\hat{G}, \hat{H}, \hat{C}) = \sum_{k=1}^{n} \mathrm{Tr}\,(Z_k - \hat{G}\hat{C}_k\hat{H}')'(Z_k - \hat{G}\hat{C}_k\hat{H}') =$$

$$= \sum_{u=1}^{n} \mathrm{Tr}\,(Z_k - \hat{G}\hat{G}'Z_k\hat{H}\hat{H}')'(Z_k - \hat{G}\hat{G}'Z_k\hat{H}\hat{H}')$$

$$= \tilde{\sigma}(\hat{G}, \hat{H})$$

That (b) is true iff (c) is true follows from

$$0 = \tilde{\sigma}(\hat{G}, \hat{H}) = \sum_{k=1}^{n} \mathrm{Tr}\,(Z_k - \hat{G}\hat{G}'Z_k\hat{H}\hat{H}')'(Z_k - \hat{G}\hat{G}'Z_k\hat{H}\hat{H}')$$

and the non-negativeness of each element of the sum.

▼▼

Lemma 3.10

Let $Z = (Z_1, Z_2, \ldots, Z_n)$ with $Z_k \in R^{l\times m}$ $(k=1, \ldots, n)$ be given and $(G, H) \in S$ arbitrary.

Define $A = \sum_{k=1}^{n} Z_k Z_k'$ and $A \in R_p^{l\times l}$ and $B = \sum_{k=1}^{n} Z_k' Z_k$

and $B \in R_q^{m\times m}$

if $s < p$ and/or $t < q$ there exists at least one k $(1 \leq k \leq n)$ such that $GG' Z_k HH' \neq Z_k$.

Proof:

Let $s<p$ and suppose $GG'Z_kHH' = Z_k$ $(k=1, \ldots,n)$

Then

$$A = \sum_{k=1}^{n} Z_k Z_k' = \sum_{k=1}^{n} GG'Z_kHH'HH'Z_k' GG' =$$

$$= \sum_{k=1}^{n} GG'Z_kHH'Z_k GG' = G(\sum_{k=1}^{n} G'Z_kHH'Z_kG)G'$$

According to a well-known theorem about the rank of a product of matrices, we obtain

$$\text{rank } A \leq \text{rang } G \quad \text{or} \quad p \leq s.$$

This contradicts $p > s$

If $t < q$ and $GG'Z_kHH' = Z_k$ $(k=1, \ldots, n)$, we come to a similar contradiction.

Thus, if $s < p$ and/or $t < q$ at least one k $(1 \leq k \leq n)$ exists such that $GG'Z_kHH' \neq Z_k$.

▼▼

Lemma 3.10 shows that $s>p$ and $t>q$ is a necessary condition for $GG'Z_kHH' = Z_k$ $(k=1, \ldots,n)$ and, in view of lemma 3.9, also for the existence of an exact solution of (I).

In the rest of this section we therefore assume that $p \leq s \leq l$ and $q \leq t \leq m$.

Theorem 3.11

Let $Z = (Z_1, Z_2, \ldots Z_n)$ and $Z_k \in R^{l \times m}$ $(k = 1, \ldots,n)$ and the minimization problem (I) be given.

a) if $(\hat{G}, \hat{H}, \hat{C})$ is an exact solution of (I) then $\hat{G}$ and $\hat{H}$ are the eigenvectormatrices (or orthonormal rotations thereof) associated with the p and q largest eigenvalues of $A = \sum_{k=1}^{n} Z_k Z_k'$ and $B = \sum_{k=1}^{n} Z_k' Z_k$ respectively, where p=rank A and q= eank B, and $\hat{C} = (\hat{C}_1, \hat{C}_2, \ldots \hat{C}_n)$ with $\hat{C}_k = \hat{G} Z_k \hat{H}'$ $(k=1, \ldots,n)$

b) On the other hand if $\hat{G}$, $\hat{H}$ are the above mentioned eigenvectormatrices (or their orthonormal rotations) and if $\hat{C}$ is defined as above, $(\hat{G},\hat{H},\hat{C})$ is only an exact solution if also $\hat{G}\hat{G}'Z_k\hat{H}\hat{H}' = Z_k$ $(k=1, \ldots,n)$

Proof: b is trivial as it follows directly from lemma 3.9c.

Now suppose $(\hat{G}, \hat{H}, \hat{C})$ is an exact solution of (I), then according to lemma 3.9 $\tilde{\sigma}(\hat{G}, \hat{H}) = 0$ and $\hat{G}\hat{G}'Z_k\hat{H}\hat{H}' = Z_k$ $(k=1, \ldots,n)$. Lemmas 3.2 and 3.3 show that $\hat{C}_k = \hat{G} Z_k \hat{H}'$ $(k=1, \ldots,n)$.

Furthermore from lemma 3.4 it follows that $p(\hat{G}, \hat{H}) = \sum_{k=1}^{n} \mathrm{Tr}\, Z_k Z_k' = \sum_{k=1}^{n} \mathrm{Tr}\, Z_k' Z_k$ and the latter expressions are clearly the global maximum of p.

Theorem 3.8 states that for any solution (G,H) of the maximization of p, and thus a fortiori for $(\hat{G},\hat{H})$, $\hat{G}$ and $\hat{H}$ are the eigenvector matrices with the largest s and t eigenvalues of $Q(\hat{H})$ and $P(\hat{G})$ respectively. (For notation and definitions see theorem 3.8).

Thus, dropping the carets for the moment,

$$G\Lambda G' = \sum_{k=1}^{n} Z_k H H' Z_k'$$

$$H\tilde{\Lambda}H' = \sum_{k=1}^{n} Z_k' G G' Z_k$$

By pre- and postmultiplying with GG' and HH' respectively and noting that HH' = HH'HH' and GG' = GG'GG' and subsequently substituting $Z_k = GG'Z_kHH'$ we get $G\Lambda G' = \sum_{k=1}^{n} Z_k Z_k' = A$, and $H\tilde{\Lambda}H' = \sum_{k=1}^{n} Z_k' Z_k = B$

This shows that G and H are the eigenvectormatrices belonging to the largest s and t eigenvalues of A and B respectively, and thus

$$p(\hat{G}, \hat{H}) = \sum_{i=1}^{s} \lambda_i = \sum_{i=1}^{t} \tilde{\lambda}_i$$

As we have assumed that s > p and t > q and as the number of non-zero eigenvalues of a matrix is equal to its rank, we may restrict ourselves to s = p and t = q

As in theorem 3.8 G and H are determined up to an orthonormal transformation.

▼▼

Corollary 3.12

Let Z, (I), A, B be defined as above and let $\hat{G}$, $\hat{H}$ be the above mentioned eigenvector matrices.

If $\hat{G}\,\hat{G}'\,Z_k\,\hat{H}\,\hat{H}' = Z_k$ (k=1, ...,n), s=p and t=q and if there are no multiple non-zero eigenvalues of A and B, then $(\hat{G}, \hat{H}, \hat{C})$ with $\hat{C}$ defined as above is the unique exact solution of (I).

Proof

Let $G\,G'\,Z_k\,H\,H' = Z_k$ (k=1, ...,n), s=p, t=q and $\hat{C}$ be properly defined then $(\hat{G}, \hat{H}, \hat{C})$is an exact solution of (I) - lemma 3.11[b].
Let there be no multiple non-zero eigenvalues of A and B, then the eigenvectors in $\hat{G}$ and $\hat{H}$ are uniquely determined.
Because in addition the non-zero eigenvalues of A and B have a prescribed order on the main diagonals of the eigenvalue matrices (c.f. chapter 1, section 2), $\hat{G}$ and $\hat{H}$ are uniquely determined.
As finally $\hat{C}$ is the unique solution of $\inf_{C} \sigma(G, H, C)$ (lemma 3.3),
$(\hat{G}, \hat{H}, \hat{C})$is the unique exact solution of (I)

▼▼

The results of this chapter can be summarized as follows:

1. there always exists a best approximate factorisation for the Tucker 2 model, and under restricted conditions an exact factorisation exists;
2. the exact factoriation is either unique or there are infinitely many factorisations;
3. if infinitely many solutions exist they can all be generated by postmultiplying the eigenvector matrices of $\sum_{k=1}^{n} Z_k Z_k'$ and/or $\sum_{k=1}^{n} Z_k' Z_k$ with orthonormal matrices.

CHAPTER FOUR

TUCKALS2: An alternating least squares method for the solution of the Tucker 2 model

4.1. Introduction

As in section 3.2 we will solely be concerned with solving the maximization of p (as defined in lemma 3.4) over the set $S = \{(G,H) \mid G \in K^{l \times s}$ and $H \in K^{m \times t}\}$. In that section it was shown that p has at least one stationary point.

Obviously we would like to construct such an algorithm for the maximization of p that it converges to a global maximum of p. Unfortunately p is the cross-product term of a multivariate polynomial of the fourth degree, and in general it is not possible to prove that methods to solve such non-linear problems attain a global maximum. Also in the present case this happens to be true, and in order to speak of convergence at all we will have to be satisfied with a rather wider definition of convergence. The algorithm willbe said to converge if it converges to one of a finite number of stationary points of p, which is not a minimum.

The method to be used here utilizes the so-called alternating least squares (ALS) technique. The essential feature of the ALS approach is that in solving optimization problems with more than one set of parameters, each set of parameters is estimated in turn by applying least squares procedures holding the other sets fixed. After all sets have been estimated once the procedure is repeated again and again until convergence. Further details and references to applications of the ALS approach can, for instance, be found in Young, de Leeuw and Takane (1977).

In order to see how the ALS approach can be applied in the present problem, let us return briefly to the revised version of minimization problem (I):

$$\min_{G,H} \tilde{\sigma}\ (G,H) = \min_{G,H} \sum_{k=1}^{n} \mathrm{Tr}(Z_k - GG'Z_kHH')'(Z_k - GG'Z_kHH')$$

Clearly the sets of parameters are here G and H. Minimizing $\tilde{\sigma}$ over G holding H fixed is identical to solving one least squares problem, and minimizing $\tilde{\sigma}$ over H holding G fixed is another. In terms of the maximization problem, this means maximizing p over G holding H fixed and vice versa.

From the above discussion a rough outline for an algorithm is readily deduced. First choose an arbitrary H_0, maximize p over G with fixed H_0, maximize subsequently over H with fixed G_1, and iterate this procedure until - one hopes - convergence. According to theorem 3.8 the maximizations are essentially identical to searches for eigenvectors and eigenvalues. More in particular we seek the largest s (or t) eigenvalues of a matrix of order I (or J). As I and/or J can be quite large, while s and t are typically very small, say 2,3,or 4, we have to look for an eigentechnique which is particularly efficient in finding just a few of a possibly large number of eigenvectors. A very appropriate technique in this situation is the so-called simultaneous iteration method of Bauer-Rutishauser (cf. Rutishauser,1969).

The maximization of p consists thus of an in principle infinite iteration process, in which at each step two eigenproblems have to be solved. Clearly, solving these eigenproblems by yet another in principle infinite iteration process has its drawbacks. The whole procedure is likely to become computationally burdensome.

In order to avoid this we perform only one single step towards the solution of the eigenproblems, in stead of the complete iteration. A similar approach has been applied by de Leeuw c.s. in a number of cases when using an ALS technique. Their experience has been that carrying out the complete iteration for solving the eigenproblem only serves to decrease the overall efficiency of the procedure, and that it has no effect on the eventual convergence point (cf.Young, de Leeuw and Takane,1977). They suggest that the reason for this behaviour might be found in the same reasons that often cause

relaxation procedures to be more efficient than non-relaxation procedures.

In section 4.2 we will sketch the simultaneous iteration method of Bauer-Rutishauser, in section 4.3 we will present the method to solve the maximization of p, christened TUCKALS2,as well as discuss the initialization of the algorithm, in section 4.4 some terminology will be introduced as well as two convergence theorems due to Zangwill(1969) and Meyer(1976), in 4.5 we will show that TUCKALS2 satisfies the conditions of Zangwill's convergence theorem, in 4.6 we will discuss a practical procedure to ensure that Meyer's convergence theorem holds, and finally in section 4.7 a separate proof will be supplied that the fixed points of the TUCKALS2 algorithm are indeed the stationary points of p.

4.2. The simultaneous iteration method of Bauer-Rutishauser

Let $A \varepsilon R^{n\times n}$ be a symmetric positive definite matrix and let $X \varepsilon R^{n\times p}$ be defined as the matrix which has as its columns the orthonormal iteration vectors $x_1,x_2,\ldots,x_p$ with $1\leq p\leq n$. If we write X after k iterations as X_k then the method of Bauer-Rutishauser is described as:

a. Choose an arbitrary $X_0 \varepsilon K^{n\times p}$

b. $Z_k=AX_k$ (thus $Z_k \varepsilon R^{n\times p}$)

c. $G_k=Z_k'Z_k$ (thus $G_k \varepsilon R_p^{p\times p}$ and G symmetric)

d. Solve the eigenproblem for G_k i.e.

determine $T_k \varepsilon K^{p\times p}$ and $E_k \varepsilon D^{p\times p}$ with

$e_1^k \geq e_2^k \geq \ldots \geq e_p^k$ such that $T_k'G_kT_k = E_k$

with T_k the eigenvector matrix if G_k and E_k the eigenvalue matrix of G_k

e. $X_{k+1}=Z_kT_kE_k^{-\frac{1}{2}}T_k'$

Schwarz a.o. (1968) show that for $k\to\infty$ $E_k^{\frac{1}{2}}$ converges to the matrix with the largest p eigenvalues of A on its diagonal,and the columns of X_k converge to the associated eigenvectors, provided A is positive definite and the vectors of X are not orthogonal to one or more of the eigenvectors concerned, and in addition the pth and p+1st eigenvalues are different.

We may write a. through e. somewhat more concisely as:

$$X_{k+1} = Z_k T_k E_k^{-\frac{1}{2}} T_k' = A X_k G_k^{-\frac{1}{2}} = A X_k (X_k' A^2 X_k)^{-\frac{1}{2}} \qquad (4.1)$$

or

$$E(X_k) = A X_k (X_k' A^2 X_k)^{-\frac{1}{2}} \qquad (4.2)$$

When we use in the sequel a recursive formula like (4.2) we mean to say that X_k can be computed by carrying out one step of the Bauer-Rutishauser method.

It should be noted that the inverse square root of $X_k' A^2 X_k$ does exist for k=0,1,2,... if the expression is positive definite. A sufficient condition is that A is symmetric and positive definite. For if A is positive definite A^{-1} exists and

$$\text{rank } X = \text{rank } A^{-1}AX \leq \text{rank } A^{-1}X \leq \text{rank } X \text{ for all } X \in K^{n\times m}$$

Thus rank AX = rank X = m, consequently rank $(AX)'AX$ = rank $X'A^2X$ = m, and thus $X'A^2X$ is of full rank.
Inspecting the method of arriving at $(X'A^2X)^{-\frac{1}{2}}$ it is seen that in fact only the inverse square root of the eigenvalues of G are taken. If therefore A is singular we could take the Moore-Penrose generalized inverse of E (cf. e.g. Ben Israel and Greville ,1974) rather than the regular inverse. The generalized inverse D^{-1} of a diagonal matrix D is again diagonal, has the reciprocals of the non-zero elements of D on its diagonal and has zeroes everywhere else. We will assume from now on that if we write expressions like $(X'A^2X)^{-\frac{1}{2}}$ we mean the Moore-Penrose generalized inverse wherever necessary. It seems a bit superfluous to introduce some special notation in this case.

4.3. The TUCKALS2 algorithm

Suppose $Z = (Z_1, Z_2, \ldots, Z_n)$, $Z_k \in R^{l\times m}$ (k=1,...,n). Furthermore let $g_1, g_2, \ldots, g_s$ and $h_1, h_2, \ldots, h_t$ with $1 \leq s \leq l$ and $1 \leq t \leq m$ be systems of orthonormal iteration vectors which are combined into matrices $G \in K^{l\times s}$ and $H \in K^{m\times t}$ respectively. If we write the matrices G and H as they are after i iteration steps as G_i and H_i, then one main iteration step of the TUCKALS2 algorithm can be described as:

G substep

$$Q_i = \sum_{k=1}^{n} Z_k H_i H_i' Z_k' \qquad \text{(thus } Q_i \in R^{l\times l} \text{ and } Q_i \text{ symmetric)} \quad (4.3)$$

$$G_{i+1} = E_1(G_i) = Q_i G_i (G_i' Q_i^2 G_i)^{-\frac{1}{2}} \quad (4.4)$$

H substep

$$P_i = \sum_{k=1}^{n} Z_k' G_{i+1} G_{i+1}' Z_k \qquad \text{(thus } P_i \in R^{m\times m} \text{ and } P_i \text{ symmetric)} \quad (4.5)$$

$$H_{i+1} = E_2(H_i) = P_i H_i (H_i' P_i^2 H_i)^{-\frac{1}{2}} \quad (4.6)$$

Thus each G substep and each H substep are in fact one single step of an 'inner' iteration to find the eigenvectors of Q and P respectively, and together they define one step of the main iteration.

It should be noted that both Q and P at each substep can easily be singular. If we assume that Z_k is of full rank, as is true in most practical cases, rank $Z_k HH' Z_k' = t$ which means that rank Q can take on any value between 0 and min(l,nt). Analoguously $0 \leq$ rank $P \leq$ min(m,ns). The inverse in (4.4) and (4.6) will in general be the generalized inverse.

Initialization

Obviously we need a G_0 and an H_0 to initialize the procedure. Two options could be considered:

a. start with a completely arbitrary G_0 and H_0 generating them by means of a pseudo-random number generator
b. choose G_0 and H_0 in such a way that they optimize something relevant to the algorithm.

The second approach would be especially appropriate if an initialization could be found which would increase the chances of convergence to a global maximum rather than a local one. Ultimately the data and the number of components drawn determine what would have been the best initialization, but that is not a basis to work with. It seems sensible to choose a G_0 and an H_0 which would solve the problem if it had an exact solution. From theorem 3.11 it follows that if an exact factorisation exists, then the eigenvectors associated with the largest eigenvalues of $A = \sum_{k=1}^{n} Z_k Z_k'$ and $B = \sum_{k=1}^{n} Z_k' Z_k$ solve the maximization problem exactly. If we therefore choose these eigenvectors as starting vectors and an exact solution exists we have found the solution to our problem without even using the main iteration procedure.

In all but a few cases we will not seek an exact solution of the maximization problem, therefore there is no real need to compute the eigenvectors completely, and only a few steps of a method to find them will suffice. In the programme we have also used the Bauer-Rutishauser method for the initialization.

4.4. Terminology, definitions and convergence theorems

First of all we need to introduce the concept of a point-to-set map, which is a direct generalization of the concept of a function. Whereas a function f (f: S → T) requires f(s) to be a single point in T, a point-to-set map F (F: S → T) allows F(s) to be a set in T. A good many algorithms use point-to-set maps in their recursive processes rather than functions; in other words given a point $s_i \varepsilon S$, the successor point $s_{i+1} \varepsilon S$ can be chosen from a set of possible successors. As we will only deal with recursive processes here T will always be the same as S.

Definition 4.1 point-to-set map

F is a point-to-set map from S to S if for any point $s \varepsilon S$ F(s) is a subset of S

The TUCKALS2 algorithm is a method that belongs to the class of algorithms addressing the problem:

$$\text{Maximize } p(s) \text{ subject to } s \varepsilon S, \tag{4.7}$$

where p is a continuous function defined on a closed subset T of R^n, and T is a subset of S. It has the characteristics common to the algorithms of this class:

a. it starts at an arbitrary $s \varepsilon T$
b. an iteration starting at s_i (i=0,1,2,....) yields a point $s_{i+1} \varepsilon F(s_i)$, where F is a point-to-set map from T into the nonempty subsets of T
c. there exists a continuous function $\pi : T \to R^1$ such that $\pi(s') \to \pi(s)$, where $s' \varepsilon F(s)$.

In fact TUCKALS2 is a so-called "primal" algorithm for which $p = \pi$ and $S = T$.

The procedures a. and b. are referred to as an algorithm correspon-

ding to F, and when c. holds the algorithm will be called monotonic.

In terms of the algorithm we will speak of a proper fixed point $s^* \in S$, if $F(s^*) = \{s^*\}$ and we will speak of a generalized fixed point $s^* \in S$, if $F(s^*) \supset \{s^*\}$.

Before we can present the convergence theorems we need to define a few more terms.

Definition 4.2 uniformly compact

A point-to-set map F is said to be uniformly compact on S if there exists a compact set T, independent of s, such that $F(s) \subset T$ for all $s \in S$.

In our case S itself is compact (lemma 3.5), thus T can be identified with S.

Definition 4.3 upper semicontinuous (u.s.c.)

A point-to-set map F is said to be u.s.c. at s if

$z_i \in F(s_i)$ $i=0,1,2,\ldots\ldots$,

$s_i \to s$,

$z_i \to z$ imply

$z \in F(s)$.

If F is u.s.c. for all points in S, then F is said to be u.s.c. on S.

Upper semicontinuity is an extension of the function continuity concept to maps (cf. Zangwill,1969,p.88).
The major problem in the sequel will be to show that the point-to-set map under consideration is u.s.c. on S.

Definition 4.4 strictly monotonic

A point-to-set map F is said to be strictly monotonic (with respect to a function p) at $s \in S$ if $s' \in F(s)$ implies $p(s') > p(s)$ whenever s is not a proper fixed point of F and $p(F(s)) = p(s)$ if s is a proper fixed point.

If F is strictly monotonic for all points in S, then F is said to be strictly monotonic on S.

Definition 4.5 essentially monotonic

A point-to-set map F is said to be essentially monotonic (with respect to a function p) at s, if $s' \in F(s)$ implies $p(s') > p(s)$ whenever s is not a generalized fixed point of F, and $p(F(s)) = p(s)$ if s is a generalized fixed point. If F is essentially monotonic for all points in S, then F is said to be essentially monotonic on S.

Convergence theorems

The notation of the following two convergence theorems has been changed slightly to stay in line with previous notation in this paper; moreover, they are stated in a less general form than in the original publications for greater clarity.

Theorem 4.6 (Zangwill's convergence theorem A, p.91)

Let $F: S \to S$ be a point-to-set map such that

a. F is uniformly compact on S,

b. F is u.s.c. on S,

c. F is essentially monotonic on S.

If $\{s_i\}$ is any sequence generated by the algorithm corresponding to F then

i. all accumulation points of $\{s_i\}$ will be generalized fixed points points,

ii. $p(s_i) \to p(s^*)$ where s^* is a generalized fixed point

iii. $\{s_i\}$ has convergent subsequences.

Theorem 4.7 (Meyer's convergence theorem 3.1, p.110)

Let $F: S \to S$ be a point-to-set such that

a. F is uniformly compact on S

b. F is u.s.c. on S

c. F is strictly monotonic on S

If $\{s_i\}$ is any sequence generated by the algorithm corresponding to F, then

i. all accumulation points of $\{s_i\}$ will be proper fixed points

ii. $p(s_i) \to p(s^*)$, where s^* is a proper fixed point

iii. $\{s_i\}$ has convergent subsequences

iv. $||s_{i+1} - s_i|| \to 0$

In both theorems iii. may also be stated as (cf. Meyer.1976):

iii'. either $\{s_i\}$ converges
or the accumulation points of $\{s_i\}$ form a continuum.

Note that neither of the theorems state that $\{s_i\}$ converges. This follows only if we make some arbitrary assumptions, for example, that there is only a finite number of fixed points or that one of the subsequential limits is an isolated fixed point. Note further that theorem 4.7 is a far more powerful theorem than theorem 4.6. In the former we know that $||s_{i+1} - s_i|| \to 0$, while in the latter we lack such information. The essential difference between the two theorems lies in the assumed nature of the fixed point. In theorem 4.6 we know that the algorithm has only generalized fixed points, while in theorem 4.7 we know that the algorithm has proper fixed points. This difference also plays a role in the definitions of the type of monotonicity as can be seen from the definitions 4.4 and 4.5.

4.5. TUCKALS2 satisfies Zangwill's convergence theorem

Let us first define or redefine the relevant parameters of the TUCKALS2 algorithm.

-- $S = \{(G,H) \mid G \in K^{l\times s},\ H \in K^{m\times t}\}$ (4.8)

-- $F: S \to S$ is a point-to-set map. F defines one complete step of the main iteration (4.9)

-- $F = F_2 \cdot F_1$ with $F_1: S \to S$ and $F_2: S \to S$ point-to-set maps such that

$$F_1(G_i,H_i) = \begin{pmatrix} E_1 \\ I_1 \end{pmatrix}(G_i,H_i) = (E_1(G_i),I_1(H_i)) = (G_{i+1},H_i) \quad (4.10)$$

$$F_2(G_{i+1},H_i) = \begin{pmatrix} I_2 \\ E_2 \end{pmatrix}(G_{i+1},H_i) = (I_2(G_{i+1}),E_2(H_i)) = (G_{i+1},H_{i+1}) \quad (4.11)$$

Thus $E_1: K^{l\times s} \to K^{l\times s}$ is a point-to-set map and so is $E_2: K^{m\times t} \to K^{m\times t}$, and they are defined by (4.4) and (4.6) respectively, while I_1 and I_2 are identity transformations.

-- $p: S \to R^1$ is the continuous bounded function to be maximized.

Because (4.4) and (4.6) are both of the form (4.2):

$$E(X) = AX(X'A^2X)^{-\frac{1}{2}} \quad (4.2)$$

we will often use (4.2) to prove properties of both (4.4) and (4.6). Moreover we will define the real continuous bounded function

$$r: K \to R^1 \text{ as } r(X) = \mathrm{Tr}\, X'AX, \quad (4.12)$$

where K is the space of all real orthonormal (a×b) matrices ($a \geq b$). We can in fact consider each G substep and each H substep one single step towards the maximization of just such a function as (4.12) (cf. also the proof of theorem 3.8). More in particular:

$$r_1: K^{l\times s} \to R^1 \quad \text{such that } r_1(G) = \text{Tr } G'QG$$

$$r_2: K^{m\times t} \to R^1 \quad \text{such that } r_2(H) = \text{Tr } H'PH$$

or formulated in terms of the algorithm:

$$r_1(G_i) = \text{Tr } G_i'Q_iG_i = p(G_i,H_i)$$

$$r_2(H_i) = \text{Tr } H_i'P_iH_i = p(G_{i+1},H_i) = p(E_1(G_i),H_i)$$

condition a.: uniform compactness

As S is compact (lemma 3.5), F: S → S defined in (4.9) is uniformly compact. Also F_1,E_1,F_2,E_2, and E are uniformly compact, because of the compactnes of $S,K^{l\times s},K^{m\times t}$ and K.

condition b.: upper semicontinuity

Note that if we can prove that E (and thus E_1 and E_2) are u.s.c., then ,because the identity transformations I_1 and I_2 are obviously continuous, F_1 and F_2 are u.s.c.; finally, if the conditions are fulfilled under which the composition $F_2 \cdot F_1$ is u.s.c.,F has the desired u.s.c property.

Clearly our first step therefore will be to show the upper semi-continuity of E. In order to do this we first need to show that, if

$$\hat{Y} \in E(X) \text{ then } \text{Tr } \hat{Y}'AX = \max \{\text{Tr } Y'AX \mid Y \in K\} .$$

Lemma 4.8

Let E: K → K be a point-to-set map defined as in (4.2), and let $\rho: K \times K \to R^1$ be defined as the continuous function

$$\rho(X,Y) = \text{Tr } Y'AX \text{ for all } X,Y \in K \tag{4.13}$$

then $\rho(X,\hat{Y}) = \max \{\rho(X,Y) \mid Y \in K\}$ for all $X \in K$, and all $\hat{Y} \in E(X)$

Proof:

Define $\tilde{\rho}(X,Y) = \rho(X,Y) - \frac{1}{2} \operatorname{Tr} M(Y'Y - I_b)$,
where M is a symmetric matrix with Lagrange multiplyers.
By equating the partial derivatives with respect to Y and
M to zero we obtain the following set of equations:

$$AX = YM \tag{4.14}$$

$$Y'Y = I_b \tag{4.15}$$

Suppose $(\hat{Y},\hat{M})$ is a solution of (4.14) and (4.15). Then by premultiplying (4.14) by its transposed equation, noting that both A and M are symmetric and by substituting (4.15) into (4.14) we obtain

$$X'A^2X = \hat{M}^2 \quad \text{thus} \quad \hat{M} = (X'A^2X)^{\frac{1}{2}}$$

Substituting this into (4.14) we get

$$AX = \hat{Y}(X'A^2X)^{\frac{1}{2}} \quad \text{and thus} \quad \hat{Y} = AX(X'A^2X)^{-\frac{1}{2}} \text{ and } \hat{Y} \in E(X). \tag{4.16}$$

This leads to

$$\rho(X,\hat{Y}) = \max \{\rho(X,Y) \mid Y \in K\} ,$$

as the Hessian of $\tilde{\rho}$ is negative.
Because the generalized inverse is not uniquely determined, $\hat{Y}$ is not uniquely determined either, and thus any $\hat{Y} \in E(X)$ maximizes ρ .

▼▼

The next step in proving that E is upper semicontinuous involves the following lemma due to Zangwill (1969), which has again been rephrased to suit our notation.

Lemma 4.9 (Zangwill's lemma 7.3, p.156)

Suppose $\rho : K \times K \to R^1$ is a continuous function on $K \times K$. For K, a compact set, let

$$f(X) = \max \{\rho(X,Y) \mid Y \in K \}$$

and define a point-to-set map $E : K \to K$ by

$$E(X) = \{Y \mid f(X) = \rho(X,Y),\ Y \in K \},$$

then the map E is u.s.c. on K.

In lemma 3.5 it was shown that K is compact and that p is continuous on S. From the proof of the lemma it can be readily seen that ρ must be continuous on K X K. Thus we can conclude using lemma 4.8:

Lemma 4.10

E is upper semicontinuous on K

From lemma 4.10 and the remarks above lemma 4.8 we conclude that F_1 and F_2 are u.s.c.

To show that F is u.s.c. we need another result from Zangwill (1969) - again slightly rephrased:

Lemma 4.11 (Zangwill's Corollary 4.2.1)

Let F_1: S → S and F_2: S → S be u.s.c. point-to-set maps on S.
If S is compact, then F: S → S with $F = F_2 \cdot F_1$ is u.s.c.

Clearly lemma 4.11 is immediately applicable, thus our F is u.s.c. on S.

condition c.: monotonicity

As with the proof of condition b. we will proceed stepwise via E, F_1 and F_2 towards the proof that F is monotonic.

Lemma 4.12

Let E be defined as in (4.2) and r as in (4.12), then

$$r(\hat{Y}) \geq r(X) \text{ for any } \hat{Y} \in E(X) \qquad (4.17)$$

with equality if and only if there exists a $Y \in E(X)$ s.t. $Y = X$.

Proof:

Let ρ be defined as in (4.13). In lemma 4.6 it was shown that

$$\rho(X,\hat{Y}) = \max \{\rho(X,Y) \mid Y \in K\} \text{ for any } \hat{Y} \in E(X)$$

From this it obviously follows that

$$\rho(X,\hat{Y}) \geq \rho(X,Y) \text{ for all } Y \in K \text{ and thus also}$$
$$\rho(X,\hat{Y}) \geq \rho(X,X) = r(X).$$

In order to prove the lemma we need to show that

$$\rho(\hat{Y},\hat{Y}) \geq \rho(X,X).$$

We will drop the carets in the following and just write Y for $\hat{Y}$.

Our next step is showing that

$$\rho(X,Y) \leq \rho(Y,Y)^{\frac{1}{2}}\ \rho(X,X)^{\frac{1}{2}} \text{ or}$$

$$\text{Tr } Y'AX \leq (\text{Tr } Y'AY)^{\frac{1}{2}}\ (\text{Tr } X'AX)^{\frac{1}{2}} \tag{4.18}$$

A is a symmetric matrix, therefore there exists an upper-triangular matrix R such that $A = R'R$ (4.19)

Substituting (4.19) in Tr Y'AX we get

$$\text{Tr } Y'AX = \text{Tr } Y'R'RX = \text{Tr } (RY)'(RX) \tag{4.20}$$

From the definition of a trace and the Cauchy-Schwarz inequality it follows that

$$\text{Tr } A'B = \sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij}b_{ij} \leq \Big(\sum_i\sum_j a_{ij}\Big)^{\frac{1}{2}}\ \Big(\sum_i\sum_j b_{ij}\Big)^{\frac{1}{2}} =$$

$$= (\text{Tr } A'A)^{\frac{1}{2}}\ (\text{Tr } B'B)^{\frac{1}{2}} \tag{4.21}$$

Combining (4.20) and (4.21):

$$\text{Tr } Y'AX = \text{Tr } (RY)'(RX) \leq (\text{Tr } Y'R'RY)^{\frac{1}{2}}\ (\text{Tr } X'R'RX)^{\frac{1}{2}} =$$

$$= (\text{Tr } Y'AY)^{\frac{1}{2}}\ (\text{Tr } X'AX)^{\frac{1}{2}}$$

From the following implications we derive the first part of the theorem:

$$\text{Tr } X'AX \leq \text{Tr}Y'AX \leq (\text{Tr } Y'AY)^{\frac{1}{2}}\ (\text{Tr } X'AX)^{\frac{1}{2}} \text{ ,thus}$$

$$(\text{Tr } X'AX)^{\frac{1}{2}} \leq (\text{Tr } Y'AY)^{\frac{1}{2}} \text{ , and}$$

$$\text{Tr } X'AX \leq \text{Tr } Y'AY \quad \text{for any } Y \in E(X) \text{ , and finally} \tag{4.22}$$

$$r(X) \leq r(Y) \text{ for any } Y \in E(X).$$

To investigate when the equality sign holds we first observe that if $X \in E(X)$ the equality sign holds trivially. To prove the converse we observe that the Cauchy-Schwarz inequality also becomes an equality if $r(X) = r(Y)$ for some $Y \in E(X)$:

$$\text{Tr } Y'AX = (\text{Tr } Y'AY)^{\frac{1}{2}}\ (\text{Tr } X'AX)^{\frac{1}{2}} \quad \text{for some } Y \in E(X).$$

This can only be the case if that Y is proportional to X, say $Y = \alpha X$. Substituting this in (4.22):

$$\text{Tr } X'AX = \text{Tr } Y'AY = \alpha^2\ \text{Tr } X'AX \tag{4.24}$$

and thus $\alpha^2 = 1$.

As $\alpha = -1$ leads to a contradiction - both sides of (4.24) must be nonnegative - $\alpha = 1$ and $Y = X$ and thus $X \in E(X)$.

▼▼

Using the relationship between E_i and F_i (i=1,2) and lemma 4.12 we may conclude that E_1 and E_2 are monotonic, and therefore F_1 and F_2 are monotonic , and finally F is monotonic as well.

To extend the second part of lemma 4.12 to F is slightly less straightforward. We want to prove that

$$\text{if } (\tilde{G},\tilde{H}) \varepsilon F(G,H) \text{ then } p(\tilde{G},\tilde{H}) = p(G,H) \text{ iff } (\tilde{G},\tilde{H}) = (G,H) \quad (4.25)$$

or rewriting (4.25) using the following equalities

$$F(G,H) = F_2 \cdot F_1(G,H) = F_2(E_1(G),H) = (E_1(G),E_2(H))$$

we have to prove that

$$\text{if } \tilde{G} \varepsilon E_1(G) \text{ and } \tilde{H} \varepsilon E_2(H) \text{ then } p(\tilde{G},\tilde{H}) = p(G,H) \text{ iff } \tilde{G} = G \text{ and } \tilde{H} = H. \quad (4.26)$$

From lemma 4.12 we have already:

$$\text{if } \tilde{G} \varepsilon E_1(G) \text{ then } \tilde{G} = G \text{ iff } r_1(\tilde{G}) = r_1(G) \text{ iff } p(\tilde{G},H) = p(G,H) \quad (4.27)$$

$$\text{if } \tilde{H} \varepsilon E_2(H) \text{ then } \tilde{H} = H \text{ iff } r_2(\tilde{H}) = r_2(H) \text{ iff } p(G,\tilde{H}) = p(G,H) \quad (4.28)$$

Using (4.27) and (4.28) we can now prove (4.26).
Let $\tilde{G} \varepsilon E_1(G)$ and $\tilde{H} \varepsilon E_2(H)$.
If $p(\tilde{G},\tilde{H}) = p(G,H)$ then according to (4.27) $\tilde{G} = G$ and according to (4.28) $\tilde{H} = H$.
Conversely if $\tilde{G} = G$ and $\tilde{H} = H$ then $p(\tilde{G},\tilde{H}) = p(G,\tilde{H})$ according to (4.27), and $p(G,\tilde{H}) = p(G,H)$ according to (4.28).

Thus we have established that

$$\begin{aligned} &\text{a. F is monotonic and} \\ &\text{b. if } (\tilde{G},\tilde{H}) \varepsilon F(G,H) \text{ then } p(\tilde{G},\tilde{H}) = p(G,H) \text{ iff } \tilde{G} = G \text{ and } \tilde{H} = H. \end{aligned} \quad (4.29)$$

According to definition 4.5 F is thus essentially monotonic.

We have thus shown that the algorithm corresponding to F, i.e., TUCKALS2, satisfies all the conditions of theorem 4.6.

4.6. Practical measures to ensure fulfillment of the conditions of Meyer's convergence theorem

It would be attractive if ways could be found to have the TUCKALS2 algorithm satisfy Meyer's convergence theorem (theorem 4.7). If that were the case we would also know that

$$||(G_{i+1},H_{i+1}) - (G_i,H_i)|| \to 0.$$

It will be shown that it is sufficient to require that in equations like (4.16) $X'A^2X$ is positive definite for the application of Meyer's theorem to be come possible.

If $X'A^2X$ is positive definite its inverse is uniquely defined and therefore $\hat{Y} = AX(X'A^2X)^{-\frac{1}{2}}$ is uniquely determined. We have, by the way, already seen in section 4.2 that the nonsingularity of A is a sufficient condition for $X'A^2X$ to be nonsingular, but that it is not a necessary one. Because $\hat{Y}$ is uniquely determined in the nonsingular case E becomes a function rather than merely a point-to-set map,and a continuous one at that. The latter fact follows because E is then composed of continuous functions. We could therefore dispense with the machinery around point-to-set maps and use simply theorems about continuous functions. We will not pursue such a course, because we want to maintain a certain parallel with the previous developments and, of course, make use of them.

For each substep in the main iteration process the uniqueness of $\hat{Y}$ means that $E_1(G_i) = G_{i+1}$ and $E_2(H_i) = H_{i+1}$ for i=0,1,2,...,and thus $(G_{i+1},H_{i+1}) = F(G_i,H_i)$. Clearly this implies that there can only be one (G,H) such that (G,H) = F(G,H), in other words the algorithm corresponding to F can only have proper fixed points. Moreover, we can rewrite (4.29) in the following way:

a. F is monotonic, and

b. (G,H) = F(G,H) iff p(F(G,H)) = p(G,H),

which means as much as F is strictly monotonic.

This leads to the conclusion that if $X'A^2X$ is positive definite at each substep theorem 4.7 can be applied and in particular it is true that

$$||(G_{i+1},H_{i+1}) - (G_i,H_i)|| \to 0.$$

It is easy to include a check on positive definiteness of $X'A^2X$ at each substep of the algorithm. After computing the eigenvalues of $X'A^2X$ theoretically one has to check if all eigenvalues are larger than zero, and in practice one checks if all eigenvalues are larger than some small number. If one or more eigenvalues are too small one can restart the procedure with a smaller number of components - i.e. with a smaller s or t. There is, however, no guarantee that this will solve the singularity problem, nor is it certain that such a procedure is necessary, because the singularity might only be a problem in that particular substep. On the other hand, if no singularities have occurred the above development shows that the conclusions of convergence theorem 4.7 apply.

So far no problems with the convergence of the algorithm have been encountered other than those in the first example of the following chapter, where they were expected.

4.7. Fixed points of the algorithm corresponding to F are stationary points of p

Theorem 4.13

Let Z,G,H,F,p be defined as in the previous sections, then
if (G,H) is a fixed point of F, then (G,H) is a stationary point of p

Proof:

Let (G,H) be a fixed point (generalized or proper) of F,

then $G = QG(G'Q^2G)^{-\frac{1}{2}}$ with $Q = \sum_{k=1}^{n} Z_k HH'Z_k'$

and $H = PH(H'P^2H)^{-\frac{1}{2}}$ with $P = \sum_{k=1}^{n} Z_k'GG'Z_k$

First consider G.

Define $M = (G'Q^2G)^{\frac{1}{2}}$, then (G,M) is a solution of

$$\tilde{Q}\tilde{G} = \tilde{G}\tilde{M} \qquad (4.30)$$
$$\tilde{G}'\tilde{G} = I_s ,$$

because $G'G = I_s$, and $GM = QG(G'Q^2G)^{-\frac{1}{2}} (G'Q^2G)^{\frac{1}{2}} = QG$

As M is symmetric there exists a $P \varepsilon K^{s\times s}$ such that

$M = P\Lambda P'$ with $\Lambda \varepsilon D^{s\times s}$. (4.31)

Substituting (4.31) into (4.30) we get

$$QG = GP\Lambda P'$$

which leads to

$$Q(GP) = (GP)\Lambda \quad \text{or} \quad Q\hat{G} = \hat{G}\Lambda \text{ with } \hat{G} = GP.$$

Thus $\hat{G}$ is a matrix with eigenvectors of Q and G is an orthonormal rotation of $\hat{G}$.

The analoguous result holds for H.

Theorem 3.8 then tells us that (G,H) is a stationary point of p.

▼▼

CHAPTER FIVE

Illustrations of the working of the TUCKALS2 algorithm

5.1. Introduction

It is our intention to illustrate the working of a programme written in Fortran based on the algorithm developed in the previous chapter. As it is not our intention to give a full account of the data used, we have not tried to give more than a sketchy interpretation of the results.

The opening examples are a first attempt to investigate the accuracy of the algorithm, the last two analyze some real data of the Dutch political scene in 1968.

5.2. Hilbert matrices

The Hilbert matrix of the order n, H_n, is defined as:

$$H_n = \{h^n_{ij}\} \text{ with } h^n_{ij} = \frac{1}{i+j-1} \quad i,j = 1,\ldots,n$$

A n×n×2 matrix Z was constructed such that $Z = (Z_1,Z_2)$ with $Z_1 = Z_2 = H_n$.

An exact factorisation of Z is possible if s=t=n (cf. corollary 3.12). The solution will be the triplet (K_n,K_n,C) where K_n is the eigenvector matrix of H_n, and $C = (C_1,C_2)$ with $C_1 = C_2 = \Lambda_n$ the eigenvalue matrix of H_n.

This can be shown as follows:

H_n is symmetric and positive definite, thus there exists an eigendecomposition of H_n: $H_n = K_n\Lambda_nK_n'$. Using the terminology of lemma 3.9,

$A_n = \sum_{k=1}^{2} H_nH_n' = \sum_{k=1}^{2} H_n^2 = 2H_n^2 = 2K_n\Lambda_nK_n' = K_n(\Lambda_n\sqrt{2})^2K_n'$. Furthermore

$B_n = \sum_{k=1}^{2} H_n'H_n = K_n(\Lambda_n\sqrt{2})^2K_n'$. Thus K_n is the eigenvector matrix of both A_n and B_n and finally $C_k = K_n'H_nK_n = \Lambda_n$.

Several tests have been carried out with Hilbert matrices of various orders, both with the initialization procedure described in section 4.3 and with random start matrices. In both cases the same results were obtained, be it that a solution was found almost without using the main iteration procedure, when the initialization routine was employed. The solution found was a rotated version of the original eigenvectors, as was ascertained through the use of a diagonalization procedure which is described in Appendix 1. In section 3.2 it was already pointed out that the solution was determined up to orthonormal transformations of the G and H matrices, which are here identical. One might be worried that inserting this extra procedure before checking the accuracy of the algorithm might impair just this accuracy, but the table below shows that this has not been the case.

TABLE 5.1 Accuracy of eigenvectors computed for Hilbert matrices of order 4

literature eigenvectors*		computed eigenvectors	
e_1	e_2	e'_1	e'_2
.79260 8291	.58207 5699	.79260 8296	.58207 5693
.45192 3121	-.37050 2185	.45192 3118	-.37050 2193
.32241 6398	-.50957 8634	.32241 6394	-.50957 8637
.25216 1170	-.51404 8272	.25216 1165	-.51404 8271
e_3		e'_3	
-.17918 629		-.17918 627	
.74191 779		.74191 771	
-.10022 814		-.10022 813	
-.63828 253		-.63828 247	
literature eigenvalues*		computed eigenvalues	
1.50021 42800 59		1.50021 42800 53	
.16914 12202 2		.16914 12202 6	
.00673 82736		.00673 82721	

* Source: Gregory, R.T. and D.L. Karney, A collection of matrices for testing computational algorithms. New York: Wiley-Interscience, 1969

As is clear from table 5.1 the computed eigenvectors and eigenvalues are a good approximation to the ones from the literature. Differences in the eigenvectors only show up in the 8th of 9th decimal. For the Hilbert 4 problem it was not possible to find the eigenvector associated with the smallest eigenvalue. The latter was apparently so small ($1.87 * 10^{-8}$), that the Jacobi routine used in the simultaneous iteration failed to find it. In fact it was so small that the matrix Q_i of which the eigenvectors were to be computed was considered singular (cf. the discussion of this problem in section 4.6). Investigations of higher order Hilbert matrices showed that their very small eigenvalues caused a similar breakdown of the procedure. A restart with fewer components solved the problem in all cases.

5.3. Data with known exact factorisation - with and without error added

Error free

This data set was constructed as follows:

- construct a random $\hat{Z} \in R^{l \times m}$;
- compute eigenvector matrices G of $\hat{Z}\hat{Z}'$ and H of $\hat{Z}'\hat{Z}$;
- construct m diagonal matrices C_k (k=1,...,n) with the random diagonal elements arranged in descending order;
- construct Z_k (k=1,...,n) such that $Z_k = G'C_kH$.

Theorem 3.11 guarantees an exact solution if s=1 and t=m, and this solution is the triplet(G,H,C) with $C = (C_1, C_2, \ldots, C_n)$ or an orthonormally rotated version thereof.

Both with the initialization procedure and with a random start an exact solution was found. The initialization proved to be very efficient, just as with the Hilbert matrices.

The 10×10×3 set of data is considerably more demanding than the previous set as for a full solution an eigenproblem of 10×10 had to be solved. Due to the rapid convergence this was not a very serious problem in the main iteration, but in the diagonalization routine a large amount of central processing time was required. In fact the latter routine was so

slow, that it took some 200 iterations and 3 minutes computing time to arrive at an 1-5 digit accuracy of the eigenvectors and a 3-5 digit accuracy of the eigenvalues. The iteration record of the orthonormalization routine showed that after 200 iterations the improvement in one step was still considerable. A less time consuming approach could be followed to show that the G and H matrices were in fact the rotated eigenvectors by rotating them towards the $\hat{G}$ and $\hat{H}$ matrix respectively. This has, however, not yet been done.

With the data set under consideration it is hardly feasible to compare The structure of the complete set of components to a restricted set of, say, three components. The randomness in the data does not provide us with a structure in the first place, as was confirmed by the very gradual decline of the component weights. The only reasonable reduction one could obtain would be by deleting the last two or three components, but there seems hardly any point in doing so.

Error added

After the construction of the Z_k's (see above) an error with a uniform distribution on the interval (0,.01) was added to each element of the Z_k matrices. The perturbed data were analyzed in the same manner as the unperturbed data.

The results can be summarized as follows:

- differences between the perturbed and unperturbed data showed up only in the tenth decimal;
- differences in the frontal planes of the core matrix showed up in the second decimal, i.e. all error was transferred to the core matix;
- after orthonormal rotation the error appeared also in the component scores (second or third decimal), and the diagonalization of the core matrix proved to be more difficult than with the unperturbed data.

5.4. Similarity of political parties (V.d.Kamp data)

In 1968 dr.L.v.d.Kamp used the method of successive intervals for pairs of stimuli to obtain similarity judgements of the major political parties. A group of hundred psychology students were used as judges.

We can store these data in a 9×9×100 block matrix of which the frontal planes are symmetric. The data were centred per frontal plane over both rows and columns (i.e. $z^k_{ij} = z^k_{ij} - z^k_{.j} - z^k_{i.} + z^k_{..}$), and analyzed by the TUCKALS2 programme. The results of the analysis are tabulated in table 5.2 and depicted in figures 5.3 and 5.4 .

TABLE 5.2 Similarity of Dutch political parties (stimulus space)

(V.d.Kamp data)

party	components		
KVP	.15	.33	-.03
PvdA	.30	-.16	.27
VVD	.02	-.16	-.54
ARP	.19	.50	.02
CHU	.06	.46	-.12
CPN	-.20	-.21	.49
PSP	.07	-.22	.65
BP	-.86	.00	-.13
D'66	.27	-.54	-.40
component weight	.39	.15	.12

As was to be expected G and H are identical so that we can limit our discussion to one of the two modes.

The most striking result is that the psychology students considered the Boerenpartij (BP - farmers' party) different from all other parties. This effect is so pronounced that it overshadows any other characteristics that differentiate between the other parties. An earlier analysis of the same data by de Leeuw (1973) confirms the unique position of the BP in this set.

Figure 5.4 shows the relationships between the other parties. The vertical axis could tentatively be labelled the left-right axis, while the horizontal one corresponds to somthing like religious affiliation or inclination. Many other studies have shown that these two attributes are the

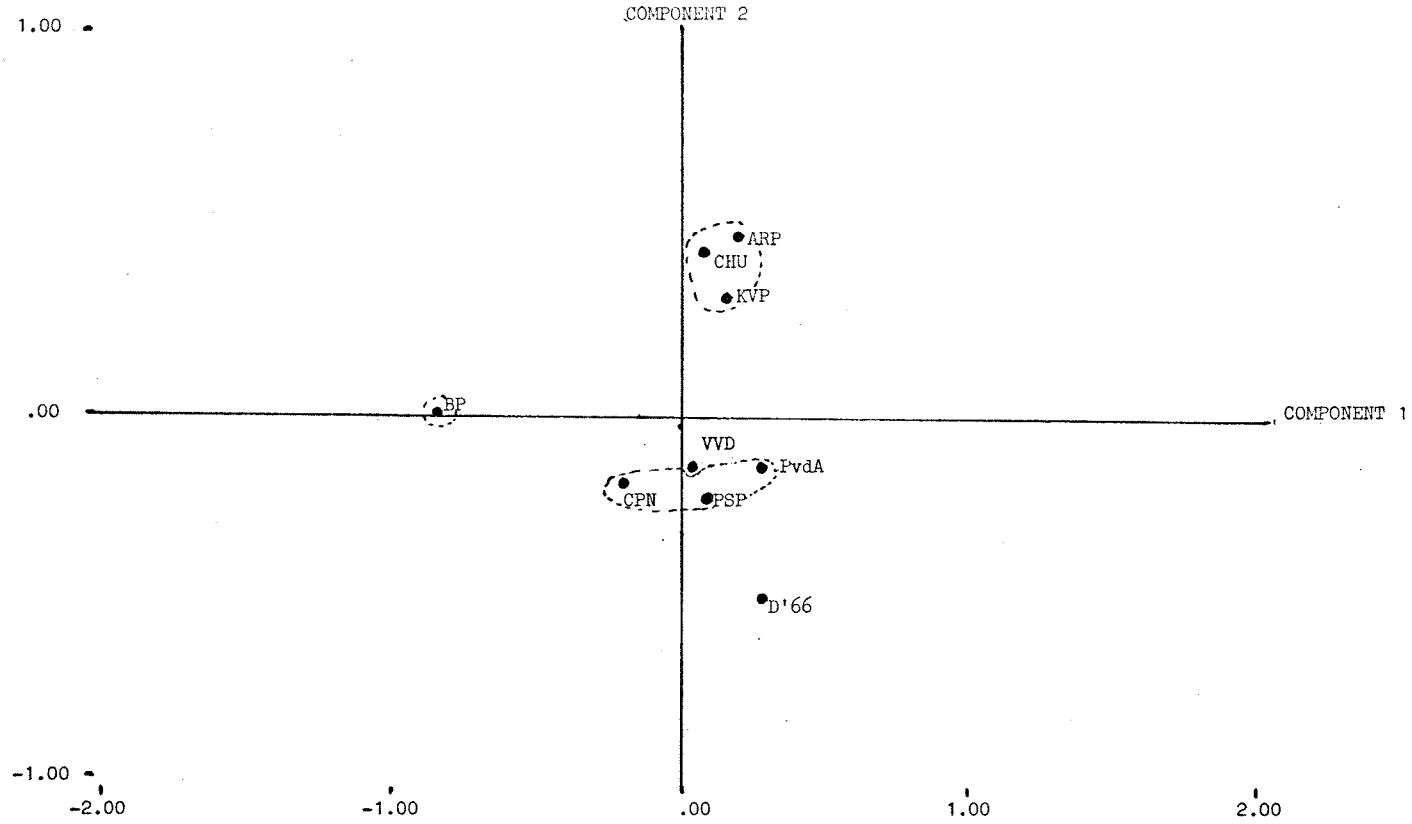


FIG. 5.3 SIMILARITIES OF DUTCH POLITICAL PARTIES

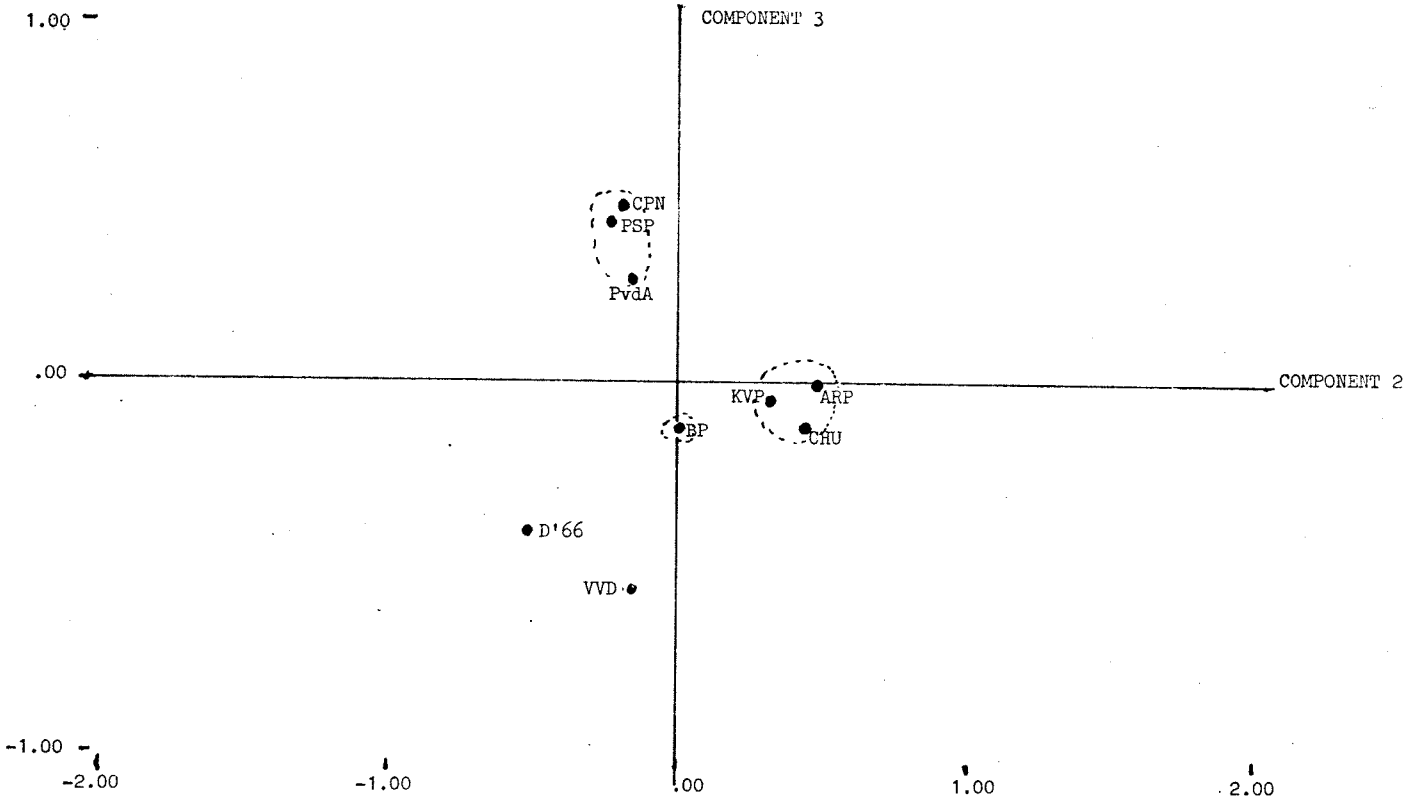


FIG. 5.4 SIMILARITIES OF DUTCH POLITICAL PARTIES

main determinants of the Dutch political system. It may be noted in passing that D'66 was considered a much more right-wing party in 1968 than at present, in view of its present membership of a left-christian democrat government.

Rather than labelling the axes it might be more useful to label the groups of parties in the manner displayed in the two figures.

Noting the more or less equality of the weights of the second and third components we might conjecture that both the left-right attribute and the religious inclination one played a comparable role in the similarity judgements.

Finally the frontal planes of the core matrix (not shown) could be used for further investigations into the differences between judges by submitting them to a cluster analysis or another principle components analysis. Ideally the latter procedure should be done at the same time as the computation of the principle components of the party space, but that would require solving the general Tucker model. At a later stage we will attempt to do just that.

5.5. Attributes, parties and psychologists (de Leeuw data)

Also in 1968 de Leeuw (1973) performed a short survey among eleven members of the staff and graduate student body of the Psychologisch Instituut of the University of Leiden to find out which attributes would be associated with which political party. All possible combinations of the twelve parties and the seventeen attributes were presented to each subject, who had to indicate whether an attribute did or did not belong to a particular party. The data were therefore in the form of eleven 12×17 matrices of zeroes and ones. These data were analyzed after double centring by the TUCKALS2 programme. The principal components of the party and of the attribute mode are given in table 5.5 and are depicted in figures 5.6 and 5.7.

TABLE 5.5 Principal components of the party and attribute space

(de Leeuw data)

party	components			attribute	components		
	1	2	3		1	2	3
D'66	.50	-.17	.19	homogeneous	.40	.09	-.01
PPR	.40	.10	.02	clear	.34	.02	-.18
PvdA	.35	-.05	-.33	consistent	.32	-.13	-.27
PSP	.15	.37	-.30	negative	.30	.04	.42
CPN	-.02	.54	.03	dogmatic	.23	-.46	-.17
ARP	-.04	-.33	-.36	left-wing	.09	.19	-.29
VVD	-.05	.04	.19	conservative	.08	-.41	.41
KVP	-.07	-.48	.26	up-to-date	-.02	.41	.01
BP	-.10	.08	.69	progressive	-.06	.23	-.18
CHU	-.28	-.38	-.21	opportunistic	-.08	.16	.59
GPV	-.41	.14	-.07	important	-.08	.27	.08
SGP	-.42	.14	-.11	sympathetic	-.16	.17	-.14
				constructive	-.22	.01	-.08
				intelligent	-.22	.11	-.01
				responsible	-.25	-.32	-.11
				tolerant	-.35	-.28	.04
				democratic	-.36	-.10	-.11
weight	.30	.15	.06		.32	.13	.06

Figure 5.6 shows the political party space, and, if we note that the weight of the third component is rather smaller than the other two, we can distinguish the following groups of parties:

-- christian democratic parties (KVP,ARP,CHU)

-- moderate left-wing parties (PvdA,PPR,D'66)

-- 'real' left-wing parties (CPN,PSP)

The GPV and the SGP are parties which are only distinguishable for insiders, as is confirmed by the analysis (in general psychologists do not tend to be affiliated to these parties).Although the third axis does not carry much weight, it is clear that it serves to set the BP apart from the other parties, and thus fulfills a similar role as the

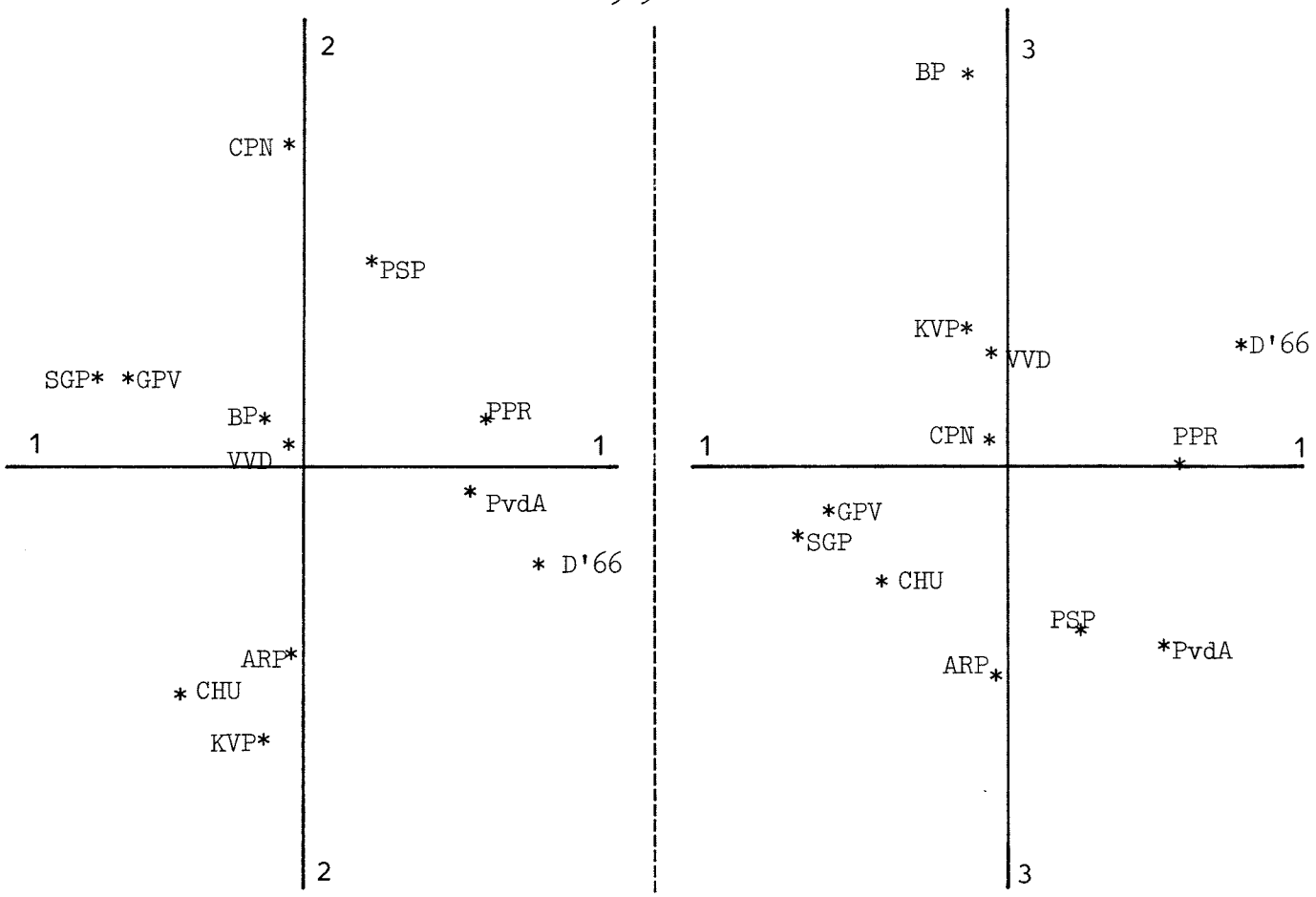


FIG. 5.6 POLITICAL PARTY SPACE (de Leeuw data)

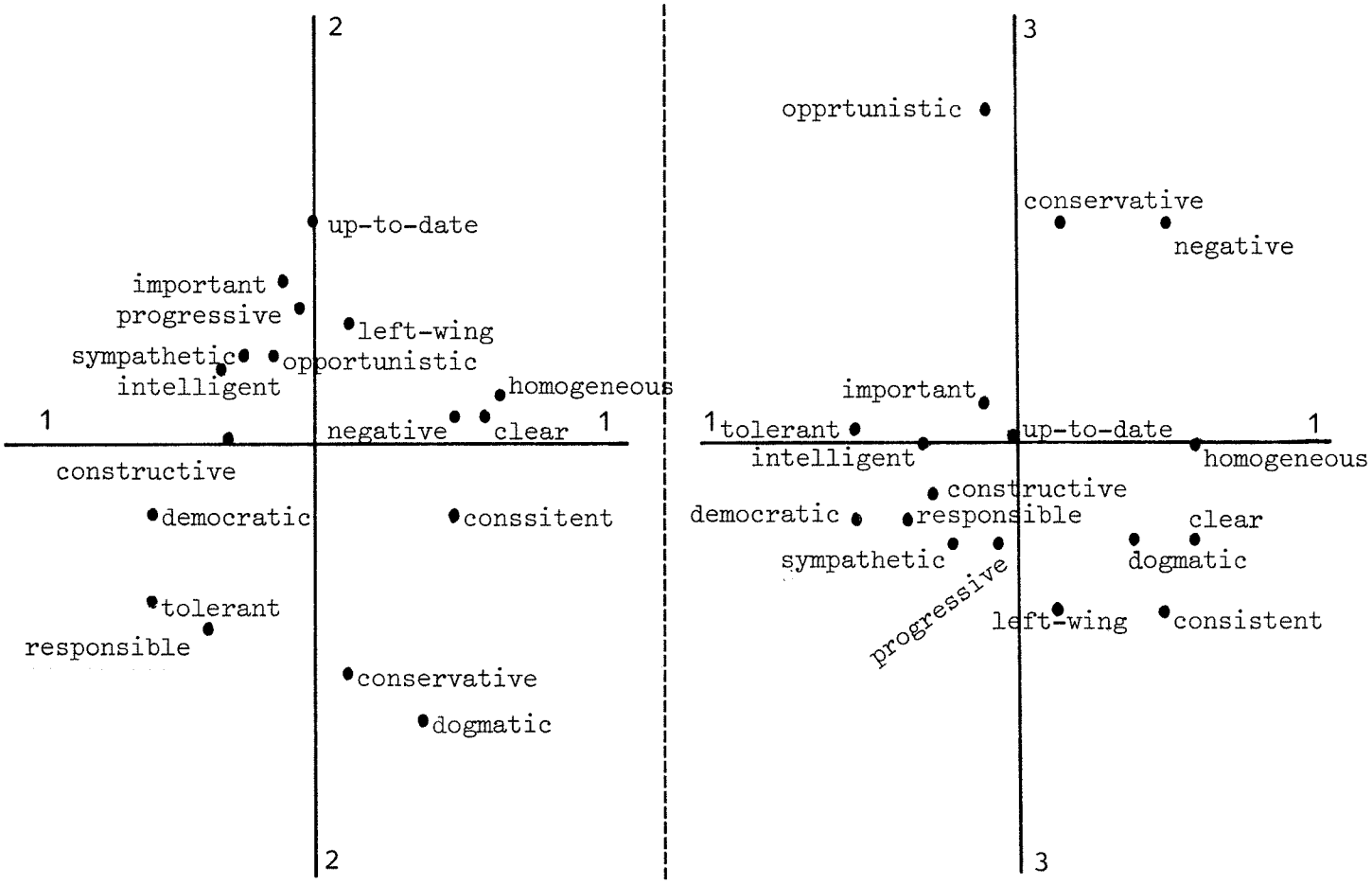


FIG. 5.7 ATTRIBUTE SPACE (de Leeuw data)

first axis in the previous example. Furthermore it should be noted that the relative position of D'66 in this example is entirely different from its position in the previous example. The reasons for this phenomenon are largely speculative. One speculation is that the religious side of the Dutch political scene is conspicuously absent from this study.
Figure 5.7 is much less telling. The attributes seem to be a rather amorphous set with a slight tendency of similar attributes to go together. This is hardly surprising as the task set was defined to give insight in the attributes of parties and not to scale the attributes themselves. The full meaning of the attribute only comes to light if we connect it in some way with the party space.

In appendix 2 a general procedure for transforming two spaces so that they can be plotted in one figure is presented, as well as an adaptation of this procedure to fit the occassion. Figure 5.8 shows the results of combining the political party and the attribute space. From this figure we acquire an indication of which attributes were seen to be associated with which parties. It is clear that the eleven psychologists favoured parties like the PvdA, PPR, and D'66. Interesting is that the Dutch communist party (CPN) and the extreme, dogmatic christian parties (SGP,GPV) were considered equally homogeneous, clear and consistent, and thus in those aspects very much alike, while on many political issues they are miles apart.

An orthonormal transformation of the core matrix was also performed, but this did not change the component spaces substancially. However, the average core matrix was transformed into a simpler structure, as can be seen from table 5.9 where the average frontal planes are given.

TABLE 5.9 Average frontal plane of core matrix before and after rotation

(de Leeuw data)

		ATTRIBUTES					
		before rotation			after rotation		
		1	2	3	1	2	3
PARTIES	1	-5.3	4.4	-2.6	-7.5	- .3	.2
	2	3.5	1.5	-2.6	.2	4.5	.1
	3	1.3	.7	2.8	- .2	.3	2.7

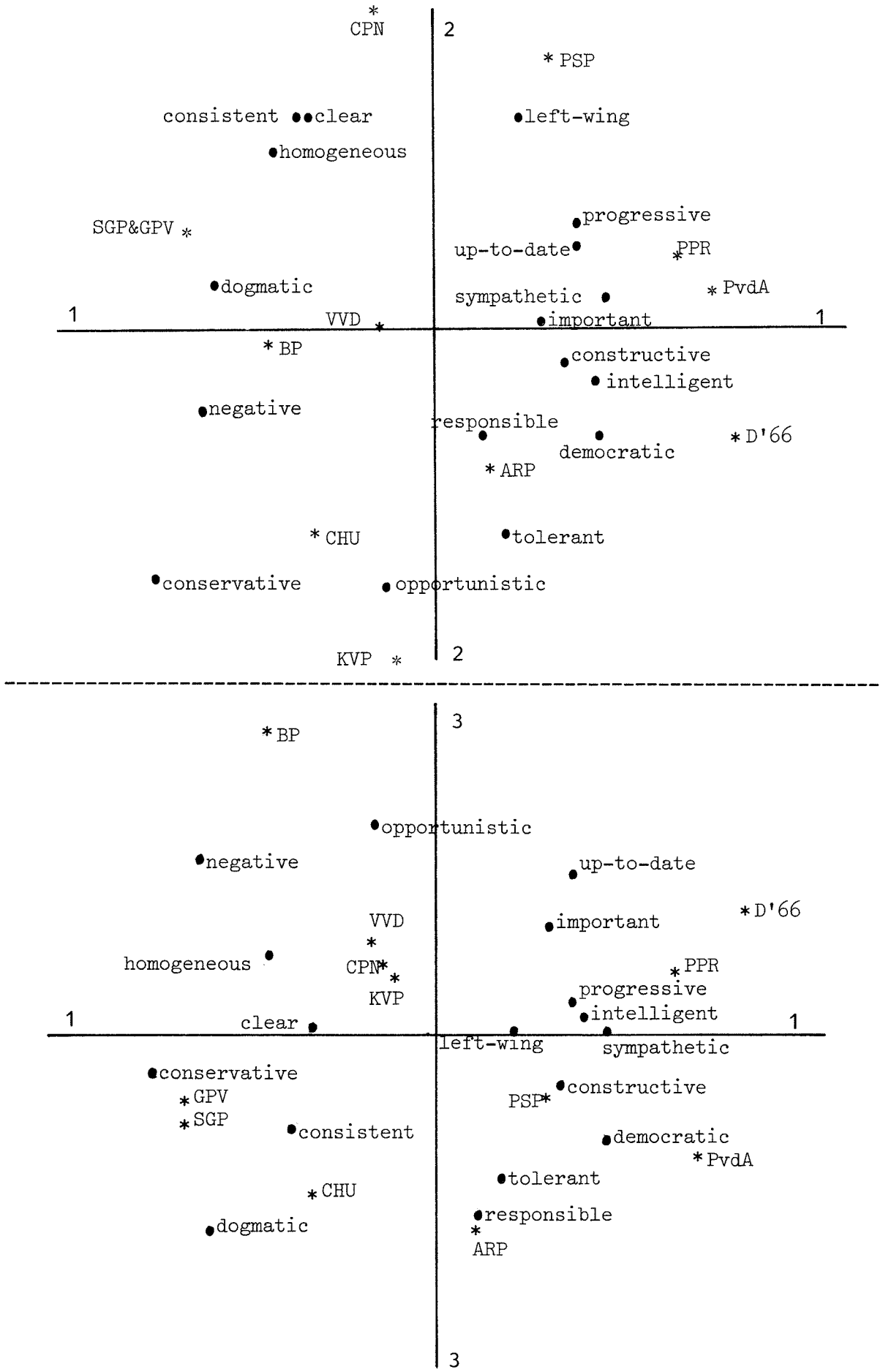


FIG. 5.8 JOINT PLOT OF ATTRIBUTE AND POLITICAL PARTY SPACE (de Leeuw data)

From the average frontal plane one might want to deduce that the core matrix has been diagonalized reasonably well inspection of the individual frontal planes does not support this conclusion. There are various ways to assess the differences between the average and the individual frontal planes , and to analyze them further, but as said earlier it is not our intention in this paper to fully explore such possibilities. We do, however, want to note that if a full diagonalization of the core matrix has been achieved, i.e. all individual frontal planes are diagonal, then the solution should be identical to an INDSCAL solution of the same data set. If the diagonalization is not perfect the applicability of the INDSCAL model might be in question. In this way the present programme could serve as a check on the appropriateness of the INDSCAL and similar models.

APPENDIX ONE

Orthonormal transformation of a core matrix to achieve optimal diagonality

In this appendix a procedure is outlined to transform a core matrix to an optimal diagonal form - in a least squares sense - by using two orthonormal transformations. It is not the intention to provide an extensive discussion of the procedure, nor to provide proofs about the convergence of the algorithm to compute the optimal transformations. These will be presented elsewhere.

Diagonality problem (D)

Be $C = (C_1, C_2, \ldots, C_n)$ with $C_k \in R^{s \times t}$ $(k=1,\ldots,n)$ given.

Find matrices $\hat{K} \in K^{s \times s}$ and $\hat{L} \in K^{t \times t}$, such that

$$\sigma(K,L) = \sum_{k=1}^{n} \mathrm{Tr}(D_k - KC_kL')'(D_k - KC_kL') \qquad (A1.1)$$

with $D_k \in D^{s \times t}$ $(k=1,\ldots,n)$, is a small as possible.

Theorem A1.1

Be $C = (C_1, C_2, \ldots, C_n)$ with $C_k \in R^{s \times t}$ $(k=1,\ldots,n)$, and the problem D given. Let D_k be defined for $k=1,\ldots,n$ as $D_k = \mathrm{diag}(KC_kL')$. Then

$$\hat{K} = \hat{U}(\hat{U}'\hat{U})^{-\frac{1}{2}} \text{ and } \hat{L} = \hat{V}(\hat{V}'\hat{V})^{-\frac{1}{2}}$$

with

$$\hat{U} = \sum_{k=1}^{n} D_k\hat{L}C_k' \text{ and } \hat{V} = \sum_{k=1}^{n} D_k'\hat{K}C_k$$

solve problem D.

Proof:

The solution of problem D is equivalent to the minimization of $\tilde{\sigma}$:

$$\tilde{\sigma}(K,L,M,N) = \sum_{k=1}^{n} \mathrm{Tr}(D_k - KC\ L')'(D\ - KC\ L') - \\ - \tfrac{1}{2}\mathrm{Tr}\ M(K'K - I_s) - \tfrac{1}{2}\mathrm{Tr}\ N(L'L - I_t) \qquad (A1.2)$$

where M and N are symmetric matrices of Lagrange multiplyers.
Let us rewrite σ in two ways:

$$1.\ \sigma(K,L) = \sum_{k=1}^{n} \mathrm{Tr}(D_k'D_k + LC_k'K'KC_kL' - D_k'KC_kL' - LC_k'K'D_k)$$

$$= \sum_{k=1}^{n} \mathrm{Tr}\, D_k'D_k - 2\sum_{k=1}^{n} \mathrm{Tr}\, K'D_kLC_k' + \sum_{k=1}^{n} \mathrm{Tr}\, C_k'C_k$$

$$= \sum_{k=1}^{n} \mathrm{Tr}\, D_k'D_k - 2\mathrm{Tr}K'(\sum_{k=1}^{n} D_kLC_k' + \sum_{k=1}^{n} \mathrm{Tr}\, C_k'C_k)$$

$$= \sum_{k=1}^{n} \mathrm{Tr}\, D_k'D_k - 2\mathrm{Tr}\, K'U + \sum_{k=1}^{n} \mathrm{Tr}\, C_k'C_k$$

$$2.\ \sigma(K,L) = \sum_{k=1}^{n} \mathrm{Tr}\, D_k'D_k - 2\sum_{k=1}^{n} \mathrm{Tr}\, L'D_k'KC_k + \sum_{k=1}^{n} \mathrm{Tr}\, C_k'C_k$$

$$= \sum_{k=1}^{n} \mathrm{Tr}\, D_k'D_k - 2\mathrm{Tr}\, L'\sum_{k=1}^{n} D_k'KC_k + \sum_{k=1}^{n} \mathrm{Tr}\, C_k'C_k$$

$$= \sum_{k=1}^{n} \mathrm{Tr}\, D_k'D_k - 2\mathrm{Tr}\, L'V + \sum_{k=1}^{n} \mathrm{Tr}\, C_k'C_k$$

Substituting these expressions succesively into (A1.2) and differentiating $\tilde{\sigma}$ with respect to K,M,L,N and equating these partial derivatives to zero, we obtain the following set of equations from the stationary equations:

$$\hat{U} = \hat{K}\hat{M} \tag{A1.3}$$
$$\hat{K}'\hat{K} = I_s \tag{A1.4}$$
$$\hat{V} = \hat{L}\hat{N} \tag{A1.5}$$
$$\hat{L}'\hat{L} = I_t \tag{A1.6}$$

Premultiplying (A1.3) with its transposed equations and using (A1.4) we get:

$$\hat{U}'\hat{U} = \hat{M}\hat{K}'\hat{K}\hat{M} = \hat{M}'\hat{M} = \hat{M}^2 \quad \Rightarrow \quad \hat{M} = (\hat{U}'\hat{U})^{\frac{1}{2}} \tag{A1.7}$$

Substituting (A1.7) into (A1.3) and isolating $\hat{K}$:

$$\hat{K} = \hat{U}(\hat{U}'\hat{U})^{-\frac{1}{2}}.$$

Analoguously:

$$\hat{L} = \hat{V}(\hat{V}'\hat{V})^{-\frac{1}{2}}$$

▼▼

From the above theorem a computational procedure can easily be derived, resulting in an ALS algorithm similar to the TUCKALS2 one.
A main iteration step of the diagonalisation routine can be described as:

K substep

$$U_i = \sum_{k=1}^{n} D_k L_i C_k'$$

$$K_{i+1} = U_i (U_i' U_i)^{-\frac{1}{2}}$$

L substep

$$V_i = \sum_{k=1}^{n} D_k' K_{i+1} C_k$$

D substep

$$D_{i+1} = \text{diag}\,[K_{i+1} C_k L_{i+1}']$$

Both $(U_i' U_i)^{-\frac{1}{2}}$ and $(V_i' V_i)^{-\frac{1}{2}}$ can be computed in the same manner as in the equivalent expression in the TUCKALS2 algorithm, i.e. by solving the eigenproblem of $U_i' U_i$ and $V_i' V_i$, and taking the inverse square root of the eigenvalues. Clearly also here problems of nonuniqueness occur in the case of singularities in $U_i' U_i$ and $V_i' V_i$, but these problems can be overcome in the same manner as in the TUCKALS2 algorithm.

As mentioned before, proof of the convergence of this algorithm will not be presented in this paper.

APPENDIX TWO

Joint plot of X and Y from the decomposition Z = XY'

A2.1. General solution

A very common problem in data analysis is to find the best rank p approximation to a given rectangular matrix. Thus, if $Z \in R_p^{n\times m}$ we want to find $\hat{Z} \in R_p^{n\times m}$ such that

$$\sigma(\hat{Z}) = \mathrm{Tr}\ (Z - \hat{Z})'(Z - \hat{Z})$$

is as small as possible. It is well known that a solution to this problem can be computed from the singular decomposition of Z (cf. section 1.4 and 2.2B). Under weak restrictions, which are almost always satisfied in practical applications, the solution for $\hat{Z}$ is unique.

A second related problem is to represent $\hat{Z} \in R_p^{n\times m}$, the solution of the first problem, in the form $\hat{Z} = XY'$, with $X \in R_p^{n\times p}$ and $Y \in R_p^{m\times p}$. This full rank decompositon of $\hat{Z}$ is in general not unique: if $\hat{Z} = XY'$ then for each $A \in R_p^{p\times p}$ we can also write $Z = \tilde{X}\tilde{Y}'$ with $\tilde{X} = XA$ and $\tilde{Y} = Y(A')^{-1}$. If $\hat{Z} = XY'$ and $\hat{Z} = \tilde{X}\tilde{Y}'$ with $X,\tilde{X} \in R_p^{n\times p}$ and $Y,\tilde{Y} \in R_p^{m\times p}$, then there is an $A \in R_p^{p\times p}$ such that $\tilde{X} = XA$ and $\tilde{Y} = Y(A')^{-1}$. A particular full rank decomposition of $\hat{Z}$ can always be found from the singular value decomposition of $\hat{Z}$. If $\hat{Z}$ is the best rank p approximation to a matrix Z, then the singular value decomposition of $\hat{Z}$ can be found from that of Z (cf. section 2.2B).

It is often desirable to represent the rows of the matrices X and Y as points in a p dimensional space. The geometrical interpretation of the ensuing picture follows from the approximation formula $Z \approx XY'$. That is to say the element z_{ij} of Z is approximately equal to the inner product of the vectors x_i and y_j $(i=1,\ldots,n;j=1,\ldots,m)$.
A basic difficulty in making a joint plot of X and Y is their indeterminacy. If we choose an arbitrary $A \in R_p^{p\times p}$ then the joint plot of $\tilde{X} = XA$ and $\tilde{Y} = Y(A')^{-1}$ can look very different. By using $A = K\Lambda L'$, the singular value decomposition of A, we can rewrite the transformations as $X = XK\Lambda L'$ and $Y = YK\Lambda^{-1}L'$. Thus first we rotate the joint plot of X

and Y, then apply stretching to X and shrinking to Y (or vice versa), and finally rotate again. It is difficult to see how the final results will be related to the original plot.

Consequently,it is desirable to choose A in a more or less rational way, thus eliminating some or all of the indeterminacy in the equation $\hat{Z} = XY'$. Because we want to make a joint plot it seems desirable to find A in such a way that the points x_i (i=1,..,n) and y_j(j=1,..,m) do not lie in separate regions of the space, but are close to each other. We measure closeness as the sum of all n×m squared distances $d^2(\tilde{x}_i,\tilde{y}_j)$, and we want to minimize this measure over A. Clearly the problem is equivalent to minimizing the function

$$\delta(A) = m\mathrm{Tr}\ \tilde{X}\tilde{X}' + n\mathrm{Tr}\ \tilde{Y}\tilde{Y}' = m\mathrm{Tr}\ AA'X'X + n\mathrm{Tr}(A')^{-1}A^{-1}Y'Y.$$

This is because the inner product components of the distances do not change by types of transformations as we are considering, only the length of the transformed vector changes. Thus minimizing the sum of the squared distances is equivalent to minimizing the length of the transformed configurations. If we now define:

$$E = mX'X \tag{A2.1}$$

$$D = nY'Y \tag{A2.2}$$

$$S = AA' \tag{A2.3}$$

the problem can be redefined as the minimization of

$$\lambda(A) = \mathrm{Tr}\ SE + \mathrm{Tr}\ S^{-1}D$$

over all symmetric and positive definite matrices S. Given the solution $\hat{S}$ to this problem we can recover A by factoring $\hat{S}$ as $\hat{S}$= LL', and by defining $\hat{A}$ = LM, with M an arbitrary rotation matrix. Therefore it is clear that our procedure does not determine $\tilde{X}$ and $\tilde{Y}$ completely, but only up to a joint rotation.

From the developement

$$\lambda(S + \varepsilon\ ) = \lambda(S) + \varepsilon\mathrm{Tr}\ \Delta(C - S^{-1}ES^{-1}) + \varepsilon^2\mathrm{Tr}\ S^{-1}\Delta S^{-1}\Delta S^{-1}D + o(\varepsilon^3)$$

it follows easily that λ is stationary at S if and only if $E = S^{-1}DS^{-1}$. Because C and D are both positive definite the unique solution ot this equation is

$$\hat{S} = D^{\frac{1}{2}}(D^{\frac{1}{2}}ED^{\frac{1}{2}})^{-\frac{1}{2}}D^{\frac{1}{2}}, \tag{A2.4}$$

which corresponds to the global minimum of λ. Thus $\tilde{X}$ = XLM and

$Y = Y(L')^{-1}M$ are the transformed solutions. Of course L can be computed from the eigendecomposition of $\hat{S}$ or from the Cholesky decomposition of $\hat{S}$ or as $\hat{S}^{\frac{1}{2}}$.

The development simplifies in an important special case. If X and Y are computed from the singular value decomposition of Z, they are represented in such a way that X'X and Y'Y are equal to the diagonal matrix with the largest p singular values (ordered in such a way that they decrease along the diagonal). This gives the solution

$$\hat{S} = \left(\frac{n}{m}\right)^{\frac{1}{2}} I_p$$

and the transformed solution $\hat{X} = \left(\frac{n}{m}\right)^{\frac{1}{4}}XM$ and $\hat{Y} = \left(\frac{m}{n}\right)^{\frac{1}{4}}YM$. Consequently, it simplifies the problem enormously if we use the singular value decomposition to start with.

A2.2. The TUCKALS2 case

The TUCKALS2 problem is considerably more complex than the general case, and a simplification is necessary if we want to plot G and H in one figure. The easiest way of doing so is by computing the average frontal plane of the core matrix $\bar{C}$, and using $\bar{C}$ to obtain an average best approximate matrix $\hat{Z} = G\bar{C}H'$. We will write C in stead of $\bar{C}$ in the following for simplicity.

Define $X = GC$ ($X \in R^{l\times t}$) and $Y = H$ ($Y \in K^{m\times t}$). We can now apply the procedure of the previous section.

Using this X and Y (A2.1) and (A2.2) become:

$$E = lX'X = lC'G'GC = lC'C \qquad \text{and } E^{-\frac{1}{2}} = l^{-\frac{1}{2}}(C'C)^{-\frac{1}{2}}$$

$$D = mY'Y = mH'H = mI_t \qquad \text{and } D^{\frac{1}{2}} = m^{\frac{1}{2}}I_t$$

Thus

$$\hat{S} = D^{\frac{1}{2}}(D^{\frac{1}{2}}CD^{\frac{1}{2}})^{-\frac{1}{2}}D^{\frac{1}{2}} = m^{\frac{1}{2}}(m^{\frac{1}{2}}Em^{\frac{1}{2}})^{-\frac{1}{2}}m^{\frac{1}{2}} = m(mE)^{-\frac{1}{2}} = m^{\frac{1}{2}}E^{-\frac{1}{2}} =$$

$$= \left(\frac{m}{l}\right)^{\frac{1}{2}}(C'C)^{-\frac{1}{2}}.$$

We recover A by factoring $\hat{S} = LL'$.

Let $U\Lambda U' = C'C$ be the eigendecomposition of C'C, then

$$(C'C)^{-\frac{1}{2}} = U\Lambda^{-\frac{1}{2}}U'$$

Thus

$$\hat{S}^{\frac{1}{2}} = L = \left(\frac{m}{l}\right)^{\frac{1}{4}}((C'C)^{-\frac{1}{2}})^{\frac{1}{2}} = \left(\frac{m}{l}\right)^{\frac{1}{4}}U\Lambda^{-\frac{1}{4}}U'$$

and

$$(L')^{-1} = \left(\frac{l}{m}\right)^{\frac{1}{4}}U\Lambda^{\frac{1}{4}}U'.$$

As M was arbitrary we may as well set it equal to U and thus we obtain:

$$\hat{X} = \left(\frac{m}{l}\right)^{\frac{1}{4}}XU\Lambda^{-\frac{1}{4}}\ U'U = \left(\frac{m}{l}\right)^{\frac{1}{4}}XU\Lambda^{-\frac{1}{4}} = \left(\frac{m}{l}\right)^{\frac{1}{4}}GCU\Lambda^{-\frac{1}{4}}$$

$$\hat{Y} = \left(\frac{l}{m}\right)^{\frac{1}{4}}YU\Lambda^{\frac{1}{4}}U'U \quad = \left(\frac{l}{m}\right)^{\frac{1}{4}}YU\Lambda^{\frac{1}{4}} \quad = \left(\frac{l}{m}\right)^{\frac{1}{4}}HU\Lambda^{\frac{1}{4}}\ .$$

A computational procedure follows directly from the above considerations.

REFERENCES

Ben Israel,A and T.N.E. Greville, Generalized inverses. Theory and applications. New York: Wiley,1974

Carroll,J.D. and J.J. Chang, Analysis of individual differences via n-way generalisation of "Eckart-Young" decomposition, Psychometrika, 35(1970),p.283-319

Carroll,J.D. and J.J. Chang, Individual differences in orientation scaling (IDIOSCAL). Paper presented at the spring meeting of the Psychometric Society, Princeton,1972

Harshman,R.A., Foundations of the PARAFAC procedure: Models and conditions for an 'explanatory' multimode factor analysis, Working papers in phonetics, No. 17,1970

Israelsson,A., Three-way (or second order component analysis. In Wold,H and E. Lyttkens (Eds.), Nonlinear iterative partial least squares (NIPALS) estimation procedures, Bulletin of International Statistical Institute,43 (1969),p.29-51

Jennrich,R., A generalisation of the multidimensional scaling model of Carroll and Chang, Working papers in phonetics, No. 22,1972

Keller,J.B., Factorization of matrices by least squares, Biometrika,49 (1962), p.239-242

Levin,J., Three mode factor analysis, Psychological Bulletin,64(1965), p.442-452

Leeuw,J. de, Some three way generalisations of principal component analysis, unpublished manuscript.

Leeuw,J. de, Canonical Analysis of Categorical Data, dissertation. Leiden, 1973

Meyer,R.R., Sufficient conditions for the convergence of monotonic mathematical programming algorithms, J.of Computer and System Sciences,12 (1976),p.108-121

Penrose,R., On the best approximate solutions of linear matrix equations, Proc.Camb.Phil.Soc.,51 (1955),p.406-413

Rudin,W., Principles of Mathematical Analysis.(2nd edition) New York: McGraw Hill Book Cy.,1964

Rutishauser,H., Computational aspects of F.L.Bauer's simultaneous iteration method, Numer. Math.,13 (1969), p.4-13

Schwarz,H.R., Rutishauser,H. and Stiefel,E. Numerik. Symmetrischer Matrizen. Stuttgart: Teubner, 1968

Tucker,L.R., Implications of factor analysis of three way matrices for measurement of change. In C.W.Harris (Ed.), Problems in measuring change. Madison: Univ. of Wisconsin Press, 1963, p.122-137

Tucker,L.R., The extension of factor analysis to three dimensional matrices. In Gullikson,H. and Frederiksen,N. (Eds.), Contributions to Mathematical Psychology. New York: Holt,Rinehart and Winston,1964

Tucker,L.R., Some mathematical notes on three mode factor analysis, Psychomertika,31 (1966),p.279-312

Young,F.W.,J. de Leeuw and Y. Takane, Quantifying qualitative data, Psychometrika,42 (1977), in press

Zangwill, W.I., Nonlinear programming. Englewood Cliffs: Prentice Hall,1969