

HOMALS

Jan de Leeuw
Jan van Rijkevorsel
Afdeling Datatheorie RUL

Samenvatting:

We beschrijven de tegenwoordige stand van zaken van het HOMALS projekt, dat diverse vormen van gegeneraliseerde principale komponentenanalyse samenbrengt in een aantal efficiënte computer-programmaas.

1: Inleiding

Principale componentenanalyse is een van de meest gebruikte data-reduktie technieken in de sociale wetenschappen. De bestaande programmaas voor PKA leggen echter nogal strenge beperkingen op aan het meetnivo van de variabelen en zijn gedwongen allerlei ad hoc maatregelen te nemen in het geval van inkomplete datamatrices. In de sociale wetenschappen is de aanname van meting op een intervalschaal dikwijls volkomen uit de lucht gegrepen en is de datamatriks dikwijls verre van compleet. Het lijkt dus wenselijk om PKA te generaliseren zodat meer algemene typen variabelen geanalyseerd kunnen worden, met willekeurige patronen 'missing data'. Dit is op verschillende manieren in de psychometrische literatuur gebeurd. In de eerste plaats zijn er technieken voor PKA van een aantal puur nominale variabelen. Guttman heeft reeds in 1941 een tamelijk complete beschrijving van zo'n methode gegeven, deze methode is later door Mosteller, Lord, Guttman, Hayashi, Lingoes, Nishisato en anderen uitgewerkt. Voor meer uitvoerige referenties verwijzen we naar De Leeuw (1973). In Frankrijk werd rond 1960 door Benzécri een ekwivalente vorm van PKA voor nominale variabelen ontwikkeld, onder de naam 'analyse des correspondances'. Ook deze methode is in vele vormen uitgewerkt, een recent overzicht van de vele toepassingen, specialisaties, en interpretaties is Benzécri e.a. (1973). De meest complete en algemene beschrijving van de resultaten van de Franse School is Dauxois en Pousse (1976). Naast de Guttman-Hayashi-Benzécri vormen van PKA voor nominale gegevens bestaan er verschillende 'niet-metrische' generalisaties van PKA. De meest bekende is misschien wel die van Shepard en Kruskal (1974), maar er bestaan ook niet-metrische PKA technieken in de GL-SSA serie, in POLYCON, en in de serie programmaas beschreven door Roskam (1968).

Het doel van het HOMALS project is tweeledig. In de eerste plaats is een algemeen theoretisch kader ontwikkeld waarin beide vormen van gegeneraliseerde PKA beschreven en verder uitgewerkt kunnen worden. En in de tweede plaats wordt veel aandacht besteed aan de ontwikkeling van programmatuur. Hierbij staat voorop dat zeer grote problemen moeten kunnen worden geanalyseerd in een redelijk korte tijd. Dit stelt hoge eisen aan zowel de algorithmes als de uiteindelijke computer programma's. De HOMALS algorithmes zijn gekonstrueerd volgens de methode van alternerende kleinste kwadraten (De Leeuw, Young, en Takane, 1976), de HOMALS programma's zijn geschreven volgens de principes van gestructureerd programmeren. Alle deelprogramma's worden geanalyseerd met timers en op basis van de verkregen profielen geoptimaliseerd.

2: Numerieke gegevens

PKA kan op verschillende, ekwivalente manieren gedefinieerd worden. In het HOMALS project geven we de voorkeur aan meetkundige uitgangspunten en de daaruit volgende verliesfuncties, dat wil zeggen aan de interpretatie van PKA als een meerdimensionale schaalmethode

Stel we hebben n observaties y_{ij} op m numerieke variabelen ($i=1, \dots, n$; $j=1, \dots, m$). We willen de observaties afbeelden als punten in een ruimte met p dimensies, en we willen de variabelen afbeelden als richtingen in die ruimte, dat wil zeggen als lijnen door de oorsprong. Observatie i wordt gerepresenteerd als het punt $(x_{i1}, x_{i2}, \dots, x_{ip})$, variabele j wordt gerepresenteerd als de richtingskosinussen $(\alpha_{j1}, \alpha_{j2}, \dots, \alpha_{jp})$. De eis die we stellen is dat de orthogonale projecties van de observatie punten $x_1, x_2, \dots, x_n$ op de richting gedefinieerd door variabele j proportioneel zijn met de metingen y_{ij} . En perfecte representatie wordt dus gedefinieerd door de eisen

$$y_{ij} = \beta_j \sum_{s=1}^p x_{is} \alpha_{js}, \quad (1a)$$

wat we ook kunnen formuleren als

$$y_{ij} = \sum_{s=1}^p x_{is} a_{js}. \quad (1b)$$

We zien dat een perfecte representatie bestaat als en alleen als we een $n \times p$ matrix X en een $m \times p$ matrix A kunnen vinden zodanig dat $Y = XA'$, dat wil zeggen als en alleen als $\text{rang}(Y) \leq p$. In het geval dat er geen perfecte oplossing bestaat voor de dimensionaliteit p die we gekozen hebben gebruiken we de verliesfunctie

$$\sigma = \text{tr} (Y - XA')'(Y - XA'), \quad (2)$$

waarbij $\text{tr}(U)$ de gebruikelijke aanduiding voor het spoor van de matrix U is. We willen deze verliesfunctie minimaliseren over X en over A . Dit probleem kan, zoals bekend, opgelost worden door de singuliere waarde ontbinding van Y of de eigenwaarde-eigenvector ontbinding van $Y'Y$ of YY' te berekenen. Deze rekenmethoden hebben echter twee nadelen. In de eerste plaats neemt de rekentijd toe als de derde macht van het aantal variabelen, en kunnen daardoor werkelijk grote voorbeelden niet geanalyseerd worden. In de tweede plaats kunnen de methoden niet gemakkelijk gegeneraliseerd worden tot meer algemene vormen van PKA.

3: Missing data

Een van deze meer algemene vormen is PKA waarbij observaties op verschillende variabelen ontbreken. We generaliseren (2) door gebruik te maken van diagonale matrices M_j , die aangeven welke observaties op variabele j ontbreken.

Met andere woorden diagonaal element i van M_j is gelijk aan één wanneer y_{ij} inderdaad geneten is, en diagonaal element i is gelijk aan nul wanneer y_{ij} ontbreekt. De voor de hand liggende generalisatie van (2) is nu

$$\sigma_1 = \sum_{j=1}^m (y_j - Xa_j)' M_j (y_j - Xa_j). \quad (3)$$

Deze verliesfunctie blijkt echter in de verdere ontwikkeling van HOMALS verschillende nadelen te hebben, die het moeilijk maken om efficiënte rekenmethoden te ontwikkelen. We bedenken daarom een alternatief. Dit is

$$\sigma_2 = \sum_{j=1}^m \text{tr}(X - y_j a'_j) M_j (X - y_j a'_j). \quad (4)$$

In de eerste plaats moeten we aantonen dat (4) een generalisatie van (2) is. Stel dat $M_j = I$ voor alle j (dwz geen missing data). Zonder verlies van algemeenheid kunnen we eisen dat $X'X = I$ en $y'_j y_j = 1$ voor alle j . Gebruik makend van deze normalisaties vinden we

$$\sigma_2 = \sigma_1 + m(p-1), \quad (5)$$

en dus komt het minimaliseren van σ_2 op hetzelfde neer op het minimaliseren van σ_1 in dit speciale geval. Als de M_j niet allemaal gelijk zijn, dan verschillen de twee problemen.

4: ALS algorithm

We herformuleren nu het rekenkundige probleem als minimalisatie van (4) over X en A onder de randvoorwaarden

$$X' M_{\star} X = I, \quad (6a)$$

$$e' M_{\star} X = 0, \quad (6b)$$

waarbij

$$M_{\star} = \sum_{j=1}^m M_j, \quad (7)$$

en waarbij e een vektor met n elementen gelijk aan één is. Alternierende kleinste kwadraten (ALS) methoden wisselen een reeks lineaire kleinste kwadraten problemen af. We beginnen met een beginschatting $X^{(0)}$, die voldoet aan (6). Dan berekenen we $A^{(0)}$ door (4) te minimaliseren over A , waarbij $X = X^{(0)}$ als een tijdelijke konstante opgevat wordt. Vervolgens berekenen we $X^{(1)}$ door (4) te minimaliseren over alle X die voldoen aan de randvoorwaarden (6), waarbij $A = A^{(0)}$ als tijdelijke konstante opgevat wordt. Deze twee soorten minimalisatieproblemen worden steeds afgewisseld. In De Leeuw, Young, en Takane (1976) en De Leeuw (1977) wordt aangetoond dat de reeks $(A^{(k)}, X^{(k)})$ naar een stationaire waarde toe konvergeert, die in het algemeen een lokaal minimum van de verliesfunctie (4) is.

De twee minimalisatie deelproblemen kunnen verduidelijkt worden door (4) op twee verschillende manieren te herschrijven. We gebruiken hierbij de bekende partitioneringen van kwadraten sommen in additieve componenten. In de eerste plaats

$$\sigma = \text{tr}(X - \hat{X})' M_{\star} (X - \hat{X}) + \text{tr} A'(V - C)A, \quad (8)$$

waarbij $\hat{X}$ gedefinieerd wordt door

$$\hat{X} = M_{\star}^{-1} \sum_{j=1}^m M_j y_j a_j'. \quad (9)$$

De matrix C heeft elementen

$$c_{jl} = y_j' M_j M_{\star}^{-1} M_l y_l, \quad (10)$$

en V is de diagonaal van C. In de tweede plaats kunnen we (4) ook herschrijven als

$$\sigma = \text{tr} X'(M_{\star} - G)X + \text{tr}(A - \hat{A})' D(A - \hat{A}), \quad (11)$$

waarbij $\hat{A}$ gedefinieerd wordt door

$$\hat{a}_j = \frac{X' M_j y_j}{y_j' M_j y_j}. \quad (12)$$

De matrix G is gelijk aan

$$G = \sum_{j=1}^m \frac{M_j y_j y_j' M_j}{y_j' M_j y_j}, \quad (13)$$

en D is diagonaal met elementen $y_j' M_j y_j$. Deze herschrijvingen zijn zowel

theoretisch als praktisch interessant. Gebruik makend van (8) en (11)

kunnen we eenvoudige benedengrenzen voor σ afleiden, en bovendien laten

(8) en (11) direct zien dat minimaliseren van σ op hetzelfde neer komt

als het vinden van eigenwaarden en eigenvectoren van $V^{-\frac{1}{2}} C V^{-\frac{1}{2}}$ ofwel

$M_{\star}^{-\frac{1}{2}} G M_{\star}^{-\frac{1}{2}}$. Uit (11) volgt direct dat minimalisatie van σ over A voor

gegeven X bereikt wordt door $A = \hat{A}$ te kiezen. Minimalisatie over X

voor gegeven A is minder eenvoudig door de randvoorwaarden. Het

is evenwel een variant van het orthogonale Procrustus probleem, en

de oplossing kan beschreven worden als

$$X = M_{\star}^{-1} J \hat{X} (\hat{X}' J M_{\star}^{-1} J \hat{X})^{-\frac{1}{2}}, \quad (14)$$

met

$$J = M_{\star} - \frac{M_{ee} e' M_{\star}}{e' M_{\star} e} \quad (15)$$

Het zal duidelijk zijn dat het berekenen van (14) in iedere HOMALS iteratie relatief het meeste tijd kost. Door voor deze deelbewerking goede algorithmes te kiezen en door vooral de desbetreffende subroutines te optimaliseren kunnen we de totale rekentijd binnen redelijke grenzen houden.

5: Optimale schaling

Tot nu toe hebben we steeds aangenomen dat alle variabelen numeriek zijn, dwz de y_{ij} zijn konstante getallen die tijdens het iteratieve proces niet veranderen. We generaliseren nu de theorie, zodat deze toepasbaar blijft onder de veel zwakkere aanname dat de vektoren $M_j y_j$ in gegeven konvekse kegels K_j moeten liggen (wat voor soort kegels dit zijn wordt in latere hoofdstukken behandeld). De verliesfunctie (4) moet nu geminimaliseerd worden over X , over A , en ook over Y . Het ALS algoritme krijgt in iedere iteratie een derde deelprobleem op te lossen, te weten minimalisatie over Y voor tijdelijk konstante X en A . Deze fase wordt soms wel de optimale scaling of optimale kwantifikatie van de variabelen genoemd, waarbij we natuurlijk moeten bedenken dat deze optimaliteit alleen maar gedefinieerd is in termen van een bepaalde verliesfunctie. Met dit derde type deelprobleem correspondeert ook een derde type partitionering van de residu kwadratensom.

$$\sigma = \sum_{j=1}^m (a_j' a_j) (y_j - \hat{y}_j)' M_j (y_j - \hat{y}_j) + \sum_{j=1}^m \text{tr } M_j X (I - \frac{a_j a_j'}{a_j' a_j}) X' M_j, \quad (16)$$

waarbij

$$\hat{y}_j = \frac{1}{a_j' a_j} M_j X a_j. \quad (17)$$

We minimaliseren nu de eerste term van (16) over y_j onder de voorwaarden dat y_j in K_j ligt, en dat $y_j' M_j y_j = 1$. Dit definieert een genormaliseerd kegel-regressie probleem, waarvoor de algemene theorie in De Leeuw (1977) besproken wordt. Numerieke PKA is nu het

speciale geval waarin de kegels K_j een-dimensionale deelruimtes (lijnen door de oorsprong) zijn. Bij ordinale variabelen zijn de kegels verzamelingen van isotone of monotone vektoren, dwz die vektoren waarop de proefpersonen op dezelfde manier gerangordend zijn als op de oorspronkelijke gegeven data vektor.

6: Diskreet en kontinu

In De Leeuw, Young, en Takane (1976) worden de verschillende schaaltypen beschreven waarop ALS algorithmen toegepast kunnen worden. We stellen hier een wat andere klassifikatie voor, hoewel we dezelfde uitgangspunten gebruiken. Alle variabelen worden als kategories beschouwd. Kategoriese variabelen kunnen diskreet of kontinu geïnterpreteerd worden. Bij diskrete kategoriese variabelen worden kategorieën als punten op een schaal gerepresenteerd, bij kontinue kategoriese variabelen worden kategorieën als niet-overlappende intervallen gerepresenteerd. In beide gevallen moeten de schaalwaarden van de individuele observaties binnen de kwantifikaties van de bijbehorende kategorieën vallen, wat in het diskrete geval betekent dat de kwantifikaties van observaties en de kategorie waartoe ze behoren moeten samenvallen. Om de relatie tussen kwantifikaties van observaties en kategorieën weer te geven gebruiken we indikator matriksen (De Leeuw, 1973). Een indikator matriks G voor n observaties op een kategorische variabele met k kategorieën is een $n \times k$ binaire matriks met $g_{ic} = 1$ als observatie i in kategorie c valt, en $g_{ic} = 0$ als i in een andere kategorie als c valt. We onderscheiden nu diskrete en kontinue variabelen van interval, ordinaal, en nominaal nivo. Kontinu interval is een nieuw type, nog niet besproken in De Leeuw, Young, Takane (1976), maar recentelijk volledig besproken in De Leeuw (1977b). De kontinue en diskrete nominale variabelen in De Leeuw, Young, Takane (1976) zijn bij nader inzien van zeer ondergeschikt belang, en we bespreken hier daarom een alternatieve definitie van deze twee belangrijke typen.

Stel een variabele is ordinaal, met k categorieën. In het continue geval kwantificeren we categorie c als een interval (z_c^-, z_c^+) , en we eisen

$$z_1^- \leq z_1^+ \leq z_2^- \leq z_2^+ \leq \dots \leq z_k^- \leq z_k^+. \quad (18)$$

En bovendien, voor alle i ,

$$\sum_{c=1}^k g_{ic} z_c^- \leq y_i \leq \sum_{c=1}^k g_{ic} z_c^+. \quad (19)$$

Voor een diskrete ordinale variabele hebben we de additionele eis dat $z_c^- = z_c^+ (= z_c)$ voor alle c , en dus

$$y_i = \sum_{c=1}^k g_{ic} z_c. \quad (20)$$

Voor numerieke variabelen zijn z_c^- en z_c^+ bekende getallen, met $z_c^- \leq z_c^+$. We eisen dat er een $\alpha \geq 0$ en een β zijn, zodanig dat

$$\alpha \sum_{c=1}^k g_{ic} z_c^- + \beta \leq y_i \leq \alpha \sum_{c=1}^k g_{ic} z_c^+ + \beta. \quad (21)$$

In het diskrete geval geldt weer dat $z_c^- = z_c^+ (= z_c)$ voor alle c , en dus

$$y_i = \alpha \sum_{c=1}^k g_{ic} z_c + \beta. \quad (22)$$

Voor continue interval data zal het bovendien dikwijls het geval zijn dat de bovengrens van een categorie gelijk is aan de benedengrens van de erop volgende categorie.

Uit (17) volgt dat een noodzakelijke voorwaarde voor een perfecte oplossing is dat Xa_j in de kegel K_j ligt. Voor ordinale diskrete data wil dat zeggen dat de y_i in k parallelle hypervlakken (van dimensie $p-1$) moeten liggen, die loodrecht staan op de lijn gedefinieerd door a_j . De projecties op deze lijn moeten bovendien in de 'goede' volgorde liggen. Als de ordinale variabele continu is dan moeten er $k-1$ van deze hypervlakken zijn, die de puntenwolken behorend bij de k categorieën scheiden. De y_i in categorie c vallen

tussen de hypervlakken behorend bij categorie $c-1$ en categorie $c+1$. Voor diskrete interval data veronderstellen we weer dat de y_i in de bijbehorende hypervlakken vallen, maar bovendien dat de hypervlakken op gegeven afstanden tot elkaar liggen (waarbij afstanden in de interval schaal betekenis worden opgevat, dwz gegeven tot op een multiplikatieve konstante). Voor continue interval data tenslotte moeten de y_i tussen een boven- en een onder-hypervlak liggen met de voorgeschreven numerieke eigenschappen. In het geval dat het boven-hypervlak van een categorie samenvalt met het onder-hypervlak van de volgende categorie betekent dit weer dat er $k-1$ hypervlakken zijn die de puntenwolken behorend bij de categorieën scheiden.

6: Nominale variabelen

In de voorafgaande hoofdstukken hebben we een generalisatie van niet-metrische PKA beschreven, compleet met verliesfuncties, een systeem van meetniveaus, algorithmes, en geometrische interpretaties. Een verdere opgave is om aan te tonen hoe nominale data ingepast kunnen worden, en hoe HOMALS gerelateerd is aan de Guttman-Hayashi-Benzécri vormen van categorische PKA. We behandelen eerst diskrete nominale data. Stel onze variabele heeft k categorieën. We definiëren k binaire diagonale matrices M_c , met diagonaal element i van M_c gelijk aan één als observatie i in categorie c valt, en diagonaal element i gelijk aan nul als i niet in categorie c valt. Als er geen missing data zijn is de som van alle M_c gelijk aan de eenheidsmatriks. De HOMALS verliesfunctie voor deze variabele is

$$\sigma = \sum_{c=1}^k \text{tr}(X - y_c a'_c) M_c (X - y_c a'_c). \quad (23)$$

De eis die we aan de optimale schalingen van de y_c stellen is dat alle observaties in categorie c een zelfde score y_{ci} krijgen, en dat bovendien $y'_c M_c y_c = 1$ voor alle c . We vatten dus als het ware

onze nominale variabele met k categorieën op als k binaire variabelen met missing data. Dit is vanzelfsprekend iets heel anders als een k -aire variabele opdelen in k complete binaire variabelen. We kunnen nu σ herschrijven, gebruik makend van de indikator matrix G en de diagonale matrix van categorie frekwenties $D = G'G$,

$$\sigma = \text{tr } X'X - 2 \text{tr } X'GD^{-\frac{1}{2}}A + \text{tr } A'A. \quad (24)$$

Als we $E = D^{-\frac{1}{2}}A$ definiëren dan

$$\sigma = \text{tr}(X - GE)'(X - GE). \quad (25)$$

Formule (25) laat twee dingen zien. Als alle variabelen diskreet nominaal zijn, als er geen missing data zijn, dan is HOMALS ekwivalent aan een eigenvektor-eigenwaarde analyse van de supermatrks C met deelmatriksen

$$C_{j\ell} = D_j^{-\frac{1}{2}}G_j'G_\ell D_\ell^{-\frac{1}{2}}, \quad (26)$$

dwz aan Guttman-Hayashi-Benzécri PKA. In de tweede plaats kunnen we de geometrische interpretaties van deze vorm van PKA dus direkt overnemen. Formule (25) laat zien dat perfecte fit gedefinieerd wordt als het samenvallen van alle observatiepunten met hun desbetreffende categoriepunten, en minimalisatie van (25) over alle variabelen betekent minimalisatie van de binnen-kategorie variantie in de representatie. Voor meer details verwijzen we naar De Leeuw (1973, 1976).

Een nominaal continue variabele wordt geanalyseerd door een k -aire nominale variabele op te splitsen in $\binom{k}{2}$ binaire variabelen. Ieder paar categorieën (c, c') definieert een binaire diagonale matrix $M_{cc'}$, met element i gelijk aan nul wanneer i niet in c en ook niet in c' valt. De verliesfunctie is weer

$$\sigma = \sum_{c=1}^k \sum_{c'=1}^k \text{tr}(X - y_{cc'}a'_{cc'})'M_{cc'}(X - y_{cc'}a'_{cc'}). \quad (27)$$

Voor $y_{cc'}$ stellen we de gebruikelijke ordinaal continue eisen, dwz de y_i korresponderend met c mogen niet kleiner zijn dan de y_i

korresponderend met c' . Geometries gezien betekent dit dat we eisen dat alle paren categorieën door hypervlakken gescheiden kunnen worden, met andere woorden dat de konvekse uitbreidingen van de k puntenwolken elkaar niet overlappen. Dit is weer heel wat anders als een k -aire variabele opsplitsen in k binaire continue variabelen, want dit laatste impliceert alleen maar dat iedere categorie gescheiden kan worden van de overige $k-1$ tezamen.

7: Diversen

7:1 Rotatie: HOMALS oplossingen kunnen geroteerd worden zonder dat de verliesfunctie verandert. Algemene rotatieprogramma's zoals VARIMAX zouden in principe gebruikt kunnen worden, maar het idee van een 'simpele structuur' is in HOMALS nog aanzienlijk simpeler dan in de gebruikelijke vormen van PKA en factor analyse. Uitbreidingen van HOMALS die de oplossingen op andere wijzen identificeren zijn in voorbereiding.

7:2 Modellen: Het is nuttig om uit te vinden wat HOMALS oplevert als het toegepast wordt op data gegenereerd volgens een bepaald analytisch of stochastisch model. In sommige gevallen (multinormaal verdeling, Guttman schaal) zijn de uitkomsten van te voren te voorspellen, in andere gevallen (Rasch model, Radex of Circumplex, Konjunktief-Disjunktief model) moeten we Monte Carlo methoden gebruiken.

7:3 Signifikantie, etc: Het is in sommige gevallen mogelijk asymptoties optimale toetsen en betrouwbaarheidsintervallen op HOMALS resultaten los te laten. Over het algemeen is dit soort aanvullende informatie echter van zeer beperkte waarde, omdat de aanname van een onderliggende werkelijke structuur alleen interessant is als we een specifiek stochastisch model aannemen. En als we eenmaal bereid zijn een dergelijk model aan te nemen, dan zijn er over het algemeen betere technieken beschikbaar.

7:4 Hoeveel dimensies: Er bestaat in de meeste gevallen niet zoiets als een werkelijke dimensionaliteit, en die hoeven we

dus ook niet te vinden. Gewoonlijk kiezen we $p = 2$ of $p = 3$, omdat dit overzichtelijke plaatjes oplevert. HOMALS kan het beste opgevat worden als een techniek voor het maken van dit soort plaatjes.

Induktieve generalisatie van een 'steekproef' naar een 'populatie' is over het algemeen niet mogelijk, althans niet zonder ekstra empiriese informatie of zonder sterke aanvullende aannamen.

7:5 Diskreet of kontinu: De continue opties lijken aantrekkelijker.

Ze zijn echter veel duurder, en empiries hebben we gevonden dat de continue optie dikwijls oninteressante plaatjes oplevert (zg 'triviale oplossingen'). En als dit laatste niet het geval is, blijken de resultaten van de diskrete optie vrijwel niet te onderscheiden van de continue.

7:6 Schaalnivo: Hetzelfde geldt in meer of mindere mate voor schaalnivo. Zwakke schaalnivoos geven soms oninteressante oplossingen, als zwakke schaalnivoos interessante oplossingen geven, dan vinden we dikwijls dezelfde oplossingen met sterkere schaalnivoos.

Referenties:

- Benzécri, J.P. Analyse des Donnees, tome II.
Paris, Dunod, 1973
- Dauvois, J. & Pousse, A. Les analyses factorielles en calcul des
probabilités et en statistique: essai d'étude
synthetique.
Thèse, Univ. Paul-Sabatier, Toulouse, 1976
- De Leeuw, J. Canonical analysis of categorical data.
Leiden, Psychologisch Instituut, 1973.
- De Leeuw, J. HOMALS.
Paper presented at Psychometric Society Meeting,
april 1976, Murray Hill, N.J.
- De Leeuw, J. A normalized cone regression approach to alternating
least squares algorithms.
Submitted, 1977.
- De Leeuw, J. Optimal scaling of continuous numerical data.
Submitted, 1977 (b).
- De Leeuw, J., Young, F.W., & Takane, Y. Additive structure in
qualitative data: an alternating least squares
approach with optimal scaling features.
Psychometrika, 41, 1976, 471-503.
- Roskam. E.E.Ch.I. Metric analysis of ordinal data in psychology.
Voorschoten, VAM, 1968.
- Shepard, R.N. & Kruskal, J.B. A nonmetric variety of linear factor
analysis.
Psychometrika, 39, 1974, 123-157.