

CONTRASTVRIJE OPLOSSINGEN VAN HET CANDECOMP/PARAFAC-MODEL

Wim P. Krijnen¹ en Jos M.F. Ten Berge

Rijksuniversiteit Groningen

Samenvatting

De CANDECOMP/PARAFAC-ontbinding van drieweg-arrays levert componenten en ladingen die uniek zijn in die zin dat rotatie een slechtere fit oplevert en dus niet is toegestaan. Onderzocht wordt of dit ontbreken van rotatievrijheid tot problemen leidt bij analyse van gegevens die bij gewone PCA juist dankzij de rotatievrijheid extra fraai geanalyseerd kunnen worden. Bedoeld worden variabelen die positief met elkaar correleren zodat er via rotatie van hoofdcomponenten contrastvrije ladingen gevonden kunnen worden. Inderdaad blijkt CANDECOMP/PARAFAC contrast-componenten op te kunnen leveren, zodat het ontbreken van rotatievrijheid een probleem vormt. Het probleem kan worden omzeild door randvoorwaarden van niet-negativiteit te introduceren. Het blijkt aldus mogelijk met een verwaarloosbaar klein verlies aan fit contrastvrije componenten te vinden. De uniciteit van de CANDECOMP/PARAFAC-oplossing verliest daarmee aan betekenis.

¹Met dank aan Henk Kiers voor stimulerend commentaar en Theo Nijse en Fred Wolters voor het beschikbaar stellen van de gegevens.

Harshman (1970) en Carroll en Chang (1970) hebben onafhankelijk van elkaar dezelfde methode ontworpen voor het analyseren van drieweg-arrays en hebben deze methode respectievelijk PARAFAC en CANDECOMP genoemd. Een drieweg-array ontstaat bijv. als van dezelfde n personen scores op dezelfde m variabelen worden verkregen op p verschillende tijdstippen. Als we de $n \times m$ scoresmatrix op tijdstip k met X_k aanduiden ($k=1, \dots, p$), dan berust CANDECOMP/PARAFAC (CP) op het minimaliseren van de functie

$$CP(A, B, C) = \sum_{k=1}^p \|X_k - AD_k B'\|^2 \quad (1)$$

waarbij A een $n \times q$ matrix is, B een $m \times q$ matrix is, C een $p \times q$ matrix is en $D_k = (\text{diag}(\mathbf{1}_q \mathbf{c}'_k))$, $k=1, \dots, p$, waarbij $\mathbf{c}'_k$ de k -de rij van C is en $\mathbf{1}_q$ een q -vector met éenen. Hoe lager de waarde van $CP(A, B, C)$ des te beter is de fit tussen het model $AD_k B'$ en de data X_k , $k=1, \dots, p$. De rang-parameter q dient van te voren door de gebruiker te worden vastgelegd. Het is bekend dat het CP-algoritme na een arbitraire start berust op het monotoon verminderen van de functiewaarde $CP(A, B, C)$ door alternerend A , B , en D_k , $k=1, \dots, p$ te berekenen via

$$A = \sum_{k=1}^p X_k B D_k \left(\sum_{l=1}^p D_l B' B D_l \right)^{-1}, \quad (2)$$

$$B = \sum_{k=1}^p X_k' A D_k \left(\sum_{l=1}^p D_l A' A D_l \right)^{-1} \quad (3)$$

en

$$D_k \mathbf{1}_q = (A A' * B B') [\text{diag}(A X_k B)] \mathbf{1}_q, \quad (4)$$

waarbij $*$ het elementsgewijze (Hadamard) product is.

Het is niet moeilijk in (1) een soort generalisatie van componentenanalyse te ontdekken, met A als componentenscoresmatrix voor alle tijdstippen, en met $D_k B' = P'_k$ als patroonmatrix, die van tijdstip tot tijdstip proportioneel varieert via D_k . De CP-methode lijkt met name dan op zijn plaats wanneer steeds dezelfde factoren (kolommen van A) een rol spelen, maar wanneer hun belang proportioneel varieert over tijdstippen via de diagonale matrixen D_k , $k=1, \dots, p$ (Harshman & Lundy, 1984).

Een van de opmerkelijke eigenschappen van CP is de eigenschap van unieke assen (Harshman, 1972). Hiermee wordt bedoeld dat rotatie van een CP-oplossing tot een slechtere fit leidt en in die zin niet is toegestaan. Kruskal (1989) en Lundy, Harshman en Kruskal (1989) spreken van 'het ontbreken van rotatieproblemen' bij CP, en vatten de unieke assen-eigenschap dus op als een pluspunt van CP. Uitgangspunt van het huidige artikel is de gedachte dat het ontbreken van rotatievrijheid in CP eerder een nadeel is dan een voordeel en wel een dermate groot nadeel dat aanpassing van het CP-algoritme voor bepaalde toepassingen noodzakelijk is. Om deze gedachte duidelijk te maken bekijken we de rol van rotatie in gewone hoofdcomponentenanalyse op variabelen die

overwegend positief correleren.

Wanneer een hoofdcomponentenanalyse wordt uitgevoerd op variabelen die onderling positief correleren dan correleert de eerste hoofdcomponent positief met (vrijwel) alle variabelen, terwijl de tweede en volgende componenten contrast-componenten zijn, dat wil zeggen ze correleren positief met sommige en negatief met andere variabelen. Het opmerkelijke is nu dat rotatie (bijv. volgens het Varimax-criterium) de contrasten pleegt op te heffen, en componenten oplevert die elk met één groepje variabelen hoog (positief) correleren, en laag (negatief) correleren met de overige variabelen (Nunnally, 1978, pag. 367). Dit wordt geïllustreerd in Tabel 1 waar voor een correlatiematrix R de ladingenmatrix staat weergegeven die met PCA wordt gevonden, evenals diezelfde matrix, geroteerd volgens het Varimax-criterium (PCAVA).

Tabel 1. Een correlatiematrix (R) en ongeroteerde en geroteerde ladingenmatrixen.

	R			PCA		PCAVA	
				I	II	I	II
1 ^a	1.00	.59	.30	.81	-.45	.10	.92
2	.59	1.00	.43	.87	-.15	.38	.80
3	.30	.43	1.00	.69	.71	.98	.16

^a1=Woordenschat, 2=Verbale Analogieën, 3=Niet-Verbale Abstractie.

De eerste PCA-component kan geïnterpreteerd worden als 'algemene intelligentie' en de tweede component als 'niet-verbaal/verbaal contrast'. De eerste geroteerde component kan geïnterpreteerd worden als 'verbale intelligentie' en de tweede geroteerde component als 'niet-verbale intelligentie'. Hieruit kan geconcludeerd worden dat componenten zonder contrast fraaier te interpreteren zijn dan contrast-componenten, terwijl uit oogpunt van verklaarde variantie beide oplossingen precies even goed zijn.

De vraag rijst nu hoe de CP-oplossing eruit ziet, wanneer herhaalde metingen van bijv. intelligentietests worden geanalyseerd: Vertoont de CP-oplossing contrasten, of lijkt de CP-oplossing meer op een simple-structure oplossing zoals bij PCA met Varimax-rotatie op intelligentietests pleegt te ontstaan? Met een voorbeeld zal worden aangetoond dat de CP-oplossing wel degelijk contrast-componenten kan leveren. Bijgevolg is het

ontbreken van rotatievrijheid in CP een probleem, wanneer CP op intelligentietests wordt toegepast. Het huidige artikel heeft tot doel een oplossing voor dit probleem te bieden. Die oplossing wordt gezocht in een wijziging van de verliesfunctie, in die zin dat contrastvrijheid der componenten als randvoorwaarde aan (1) wordt toegevoegd.

CP onder randvoorwaarden van niet-negativiteit

Bij componentenanalyse is de structuurmatrix gedefinieerd als de matrix met correlatie-coëfficiënten tussen variabelen en factoren. Aldus kan bij CP de k -de structuurmatrix S_k gedefinieerd worden als $X_k'A$, waarbij elke kolom van zowel X_k als A een gemiddelde 0 en lengte 1 heeft. Bij componentenanalyse is een contrast-component gedefinieerd als een component die positief met minstens één variabele en negatief met minstens één andere variabele correleert. Het ligt voor de hand een contrast-component bij CP ook zo te definiëren. Dus een CP-component is een contrast-component als die component met minstens één variabele negatief en met minstens één andere variabele positief correleert, al dan niet op hetzelfde tijdstip.

Bij CP-analyse van een drieweg-array met scores op intelligentietests uit de praktijk worden contrast-componenten gevonden. We illustreren dit aan de hand van een drieweg-array met scores van 33 personen op 3 intelligentietests uit de Drenth-serie (Drenth & Hoolwerf, 1977) op 2 tijdstippen. Het oorspronkelijke drieweg-array met gegevens bevatte scores van 38 personen op 3 variabelen op 2 tijdstippen. Om vergelijking van contrastvrije PCA-componenten met CP-componenten mogelijk te maken zijn na het kolomsgewijs standaardiseren van de twee scores-matrixen de personen verwijderd met de grootste persoonsvariantie, totdat per tijdstip contrastvrije Varimax-geroteerde structuurmatrixen ontstonden. In Tabel 2 staan structuurmatrixen gevonden met PCA gevolgd door Varimax-rotatie, en met CP. Uit de linkerhelft van Tabel 2 kan worden afgelezen dat het CP-algoritme een oplossing met contrast-componenten geeft, terwijl elk van de Varimax-geroteerde structuurmatrixen contrastvrij is. Hieruit kan worden geconcludeerd dat het ontbreken van rotatievrijheid bij CP een probleem is. Een oplossing voor dit probleem wordt gezocht in het opleggen van niet-negativiteits-randvoorwaarden aan de elementen van bepaalde matrixen, want daarmee worden contrasten per definitie uitgesloten.

Tabel 2. Structuurmatrixen per tijdstip gevonden met Varimax-rotatie na PCA (PCAVA), CP, CPS⁺ en CPS⁺B⁺, de correlatie tussen componenten ($r_{I,II}$) en het percentage (%) verklaarde variantie.

	PCAVA		CP		CPS ⁺		CPS ⁺ B ⁺	
	I	II	I	II	I	II	I	II
tijdstip 1								
1 ^a	.92	.10	-.18	.53	.18	.33	.84	.19
2	.80	.38	.10	.28	.43	.56	.72	.44
3	.16	.98	.81	-.56	.86	.85	.00	.88
tijdstip 2								
1	.94	.06	-.42	.74	.00	.17	.92	.00
2	.87	.31	-.06	.45	.34	.49	.81	.35
3	.15	.98	.64	-.37	.82	.83	.09	.80
$r_{I,II}$	.00		-.90		.98		-.01	
%	88.97		75.42		75.22		75.09	

^a1=Woordenschat, 2=Verbale Analogieën, 3=Niet-Verbale Abstractie.

Er zijn globaal gesproken twee manieren om randvoorwaarden van niet-negativiteit bij CP te introduceren. We kunnen allereerst eisen dat A en/of B en/of C geen negatieve elementen hebben. We krijgen dan regressieproblemen waarbij gewichtenvectoren (bijv. rijen van B) gezocht worden onder de randvoorwaarde van niet-negatieve elementen. Een algoritme dat een oplossing voor dergelijke regressieproblemen geeft staat beschreven in Lawson en Hanson (1974, pag. 161). Zij noemen dit algoritme het 'NonNegative Least Squares algorithm' (NNLS). Het NNLS-algoritme convergeert monotoon naar een globaal minimum (Tenenhaus, 1988). In het geval van CP is het dus mogelijk de eis van niet-negativiteit rechtstreeks aan de elementen van A , B en C op te leggen.

Niet-negativiteits-randvoorwaarden kunnen ook indirect aan A en/of B en/of C worden opgelegd, bijv. als we een A zoeken onder de randvoorwaarde dat de structuurmatrixen S_k , $k=1, \dots, p$, geen negatieve elementen hebben. Uit de definitie $S_k = X'_k A$ volgt dan dat elke kolom a_r ($r=1, \dots, q$) van A optimaal moet zijn voor CP, maar onder de randvoorwaarde dat $X'_k a_r$, $k=1, \dots, p$, geen negatieve elementen bevat. Een algoritme dat een oplossing voor dergelijke regressieproblemen geeft staat eveneens beschreven in Lawson en Hanson (1974,

pag. 167-168). Zij noemen dit een 'linear Least Squares with linear Inequality constraints' (LSI)-algoritme. Het LSI-algoritme stelt ons in staat een verbetering van A te vinden die voldoet aan de randvoorwaarde dat S_k , $k=1, \dots, p$, geen negatieve elementen bevat. Als we alternerend A verbeteren met het LSI-algoritme en B en $D_1, \dots, D_p$ met respectievelijk (3) en (4), dan hebben we een alternerend kleinste kwadraten (ALS)-algoritme. We zullen dit algoritme verder CPS⁺-algoritme noemen. Hierbij geeft S^+ aan dat de eis van niet-negativiteit is opgelegd aan de structuurmatrixen.

We weten dat een CPS⁺-oplossing per definitie geen contrasten in de structuurmatrixen heeft. Maar de vraag is of zo'n oplossing acceptabel is. In Tabel 2 staan de structuurmatrixen van de CPS⁺-oplossing van rang 2 voor de Drenth-gegevens. Wat opvalt is het geringe verlies aan verklaarde variantie ten opzichte van de CP-oplossing. Wat verder opvalt is de hoge correlatie tussen de twee componenten. Als we de randvoorwaarde van niet-negatieve structuurmatrixen aan (1) opleggen dan lopen we kennelijk het risico een oplossing te vinden met twee haast identieke componenten. Desondanks verklaart de oplossing met twee hoog correlerende componenten ongeveer 20% variantie meer dan de oplossing van rang 1. Dit is mogelijk doordat het contrast zich heeft verplaatst naar B . In Tabel 3 staat de gevonden B -matrix, waarbij de lengte van de kolommen van B op 1 is geschaald. Uit Tabel 3 en het feit dat de twee componenten extreem hoog correleren concluderen we dat het niet voldoende voor een acceptabele oplossing is om alleen een positieve structuur te eisen.

Tabel 3. Regressiegewichten-matrix B bij CPS⁺ op de Drenth-gegevens.

	I	II
1 ^a	-.77	.75
2	-.63	.66
3	.08	.05

^a1=Woordenschat, 2=Verbale Analogieën, 3=Niet-Verbale Abstractie.

Het feit dat het contrast zich kon verplaatsen werd mogelijk gemaakt doordat de elementen van B zowel positief als negatief mochten zijn. Door nu de extra randvoorwaarde dat B uitsluitend niet-negatieve elementen heeft op te leggen aan het CPS⁺-model, kunnen we voorkomen dat het contrast zich verplaatst naar B . Als we alternerend A verbeteren met het LSI-algoritme, B met het NNLS-algoritme en C volgens (4) dan hebben we een CPS⁺ B^+ -algoritme.

Geheel rechts in Tabel 3 staan de resultaten van het CPS^+B^+ -algoritme voor de Drenth-gegevens. De poging om laag correlerende componenten te krijgen blijkt volledig geslaagd, en gaat bovendien gepaard met een verwaarloosbaar verlies aan verklaarde variantie. Verder zien we dat de structuurmatrixen van de CPS^+B^+ -oplossing sterk lijken op die van PCAVA. Deze gelijkenis gaat samen met de gevonden hoge correlaties van de PCAVA-componenten met de componenten van de drie geschetste CP-algoritmen, zie Tabel 4.

Tabel 4. Correlaties per tijdstip van Varimax-geroteerde PCA-componenten met componenten gevonden met CP, CPS^+ en CPS^+B^+ .

	CP		CPS^+		CPS^+B^+	
	I	II	I	II	I	II
tijdstip 1						
PCAVA I	-.29	.66	.12	.29	.95	.12
PCAVA II	.88	.68	.86	.83	-.15	.88
tijdstip 2						
PCAVA I	-.42	-.77	.01	.18	.97	.01
PCAVA II	.73	.51	.85	.83	-.06	.83

Uit Tabel 4 wordt duidelijk dat de CPS^+B^+ -componenten verreweg het sterkst overeenkomen met de PCAVA-componenten op tijdstip 1 en 2. Hieruit volgt dat de CPS^+B^+ -componenten het meest lijken op een simple structure oplossing zoals ontstaat na Varimax-rotatie van de PCA-oplossing. Meer in het algemeen kan worden geconcludeerd dat de toevoeging van niet-negativiteits randvoorwaarden een potentieel aantrekkelijke manier is om het ontbreken van rotatievrijheid in CP te compenseren, ten koste van een verwaarloosbare hoeveelheid verklaarde variantie.

Men kan zich nu afvragen of de tekencontrasten, die uit de structuurmatrixen en uit de matrix B verdreven zijn, zich misschien in de nog vrijgelaten matrixen D_1 t/m D_p genesteld hebben. Dat blijkt hier niet het geval te zijn: Bij de CPS^+B^+ -oplossing voor de Drenth-gegevens zijn D_1 en D_2 niet-negatief. Deze bevinding is nauwelijks verrassend te noemen. Uit de lage correlatie tussen de beide componenten (zie Tabel 2) volgt dat S_1 en S_2 zowel structuur als patroon zijn voor de scoresmatrixen apart. Uit de vrij goede fit van de CPS^+B^+ -oplossing volgt dat S_1 en S_2 weinig afwijken van BD_1

en BD_2 . Dit laat weinig ruimte over voor D_1 en D_2 om nog negatieve elementen te hebben.

Als C geen negatieve elementen heeft dan kunnen extreem hoog correlerende CPS^+B^+ -componenten niet veel meer variantie verklaren dan één component. Dit volgt uit het feit dat de kolommen van AD_kB' , $k=1, \dots, p$, in de kegel liggen die door de kolommen van A wordt opgespannen. Dit betekent dat de representaties van de kolommen van X_k , $k=1, \dots, p$, in de kegel liggen die door de kolommen van A wordt opgespannen. Naarmate de correlatie tussen de kolommen van A hoger wordt, is de ruimte om kolommen van X_k , $k=1, \dots, p$, te representeren kleiner en dus wordt ook de fit slechter. Dit gezamenlijk optreden van een slechtere fit met een hogere component-correlatie is een wenselijke eigenschap aangezien een tweede component die zowel qua interpretatie als qua hoeveelheid verklaarde variantie weinig aan de rang 1 oplossing toevoegt buiten beschouwing kan blijven.

Naast de Drenth-gegevens werden nog twee andere drieweg-arrays onderzocht. Bij beide arrays zijn de matrixen met scores van personen op variabelen per tijdstip kolomsgewijs gestandaardiseerd. Vervolgens zijn de personen verwijderd met de grootste persoonsvariantie, totdat contrastvrije Varimax-geroteerde structuurmatrixen ontstonden. Aldus werden scores van 20 personen verkregen op 4 subtests van de GIT (Luteijn & van der Ploeg, 1983), gemeten op twee tijdstippen, en werden scores van 87 personen verkregen op 9 subtests van de DAT (Evers & Lucassen, 1984), eveneens op twee tijdstippen.

De CP-oplossing leverde bij de GIT twee laag ($r=.15$) correlerende contrast-componenten, sterk lijkend op ongeroteerde hoofdcomponenten. De CPS^+ -oplossing leverde twee hoog correlerende componenten ($r=.85$), terwijl de CPS^+B^+ -oplossing mooie laag correlerende componenten opleverde, die evenals bij de Drenth-gegevens sterk overeenkwamen met de per tijdstip gevonden PCAVA-componenten. Het percentage verklaarde variantie van de beide CP-, de CPS^+ - en de CPS^+B^+ -componenten bedroeg respectievelijk 68.88, 68.76 en 68.64.

Bij de DAT-gegevens vertoonde de CP-componenten al vrijwel géén contrast, en waren de beide CPS^+B^+ -componenten slechts in geringe mate fraaiër te interpreteren dan de CP- en de CPS^+ - componenten. Voor deze gegevens lijkt de CPS^+B^+ -oplossing nauwelijks rendabel, omdat de contrast-problemen zich hier ook nauwelijks manifesteren. Het percentage verklaarde variantie van de twee CP-, de CPS^+ - en de CPS^+B^+ -componenten bedroeg respectievelijk, 51.36, 51.35 en 51.34.

Discussie

Bij de drie onderzochte arrays vonden we in verschillende mate contrast-componenten bij gebruik van CP, die weggewerkt konden worden via

CP^+B^+ . Daarmee lijkt een uitstekend alternatief voor de bij CP ontbrekende rotatievrijheid gevonden te zijn: Tegen inlevering van een kleine hoeveelheid verklaarde variantie verkrijgen we aanzienlijk beter interpreteerbare componenten. De uniciteit van de CP-oplossing blijkt daarmee haar betekenis te verliezen.

De ontwikkelde algoritmen zijn alle van het Alternating-Least-Squares-type, en convergeren dientengevolge monotoon. Het gevaar van lokale minima is echter aanwezig: Dit gevaar kan worden verkleind door herhaalde malen te starten met verschillende start-matrixen.

Indien n , het aantal personen, zeer groot is kunnen er problemen ontstaan bij het opslaan van arrays in het werkgeheugen van een computer. Indien het aantal variabelen m en het aantal tijdstippen p niet al te groot is kan dit probleem opgelost worden. Het is namelijk mogelijk een CP^+B^+ -algoritme te construeren waarbij het grootste array dat nodig is van orde mp bij mp is. Bij de eerste auteur is een testprogramma op aanvraag verkrijgbaar.

Verwijzingen

- Carroll, J.D., & Chang, J.J. (1970). Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. *Psychometrika*, 35, 283-319.
- Drenth, P.J.D. & Hoolwerf, G. (1977). *Hernormering testserie hoger nivo. Een hernormeringsonderzoek met betrekking tot de tests: Woordenschat, Functies van Woorden, Verbale Analogieën (VAT'69), Test voor Niet Verbale Abstractie (TNVA), Cijferreeksen en Rekenvaardigheid (NAT'70)*. Amsterdam: Swets & Zeitlinger.
- Evers, A. & Lucassen, W. (1984). *Voorlopige Handleiding DAT'83, Differentiële Aanleg Testserie*. Lisse: Swets & Zeitlinger.
- Harshman, R.A. (1970). Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multi-mode factor analysis. *UCLA-Working Papers in Phonetics*, 16, 1-84.
- Harshman, R.A. (1972). Determination and proof of minimum uniqueness conditions for PARAFAC1. *UCLA-Working Papers in Phonetics*, 22, 111-117.
- Harshman, R.A., & Lundy, M.E. (1984). The PARAFAC model for Three-Way Factor Analysis and Multidimensional Scaling. In H.G. Law, C.W. Snyder, J.A. Hattie, & R.P. McDonald (Eds.), *Research methods for multimode data analysis* (pp. 122-215). New York: Praeger.
- Kruskal, J.B. (1989). Rank, decomposition, and uniqueness for 3-way and N-way arrays. In R. Coppi, and S. Bolasco (Eds.), *Multway Data Analysis* (pp. 7-18). Amsterdam: Elsevier Science Publishers B.V.
- Lawson, C.L., & Hanson, R.J. (1974). *Solving Least Squares Problems*

- New-Jersey: Prentice-Hall.
- Lundy, M.E., Harshman, R.A., & Kruskal, J.B. (1989). A two stage procedure incorporating good features of both trilinear and quadrilinear models In R. Coppi, and S. Bolasco (Eds.), *Multiway Data Analysis* (pp. 123-130). Amsterdam: Elsevier Science Publishers B.V.
- Luteijn, F. & Ploeg, F.A.E. van der (1983). *Handleiding GIT*. Lisse: Swets & Zeitlinger.
- Nunnally, J. C. (1978). *Psychometric Theory*. New York: McGraw-Hill.
- Tenenhaus, M. (1988). Canonical analysis of two convex polyhedral cones and applications. *Psychometrika*, 53, 503-524.

Geaccepteerd: 04-03-1991