

**HORIZONPUNTEN IN HET EXTERNE
ONTVOUWINGSMODEL**

Liesbeth Wachmann

Department of Data Theory

University of Leiden

Inhoud

1.	Inleiding	1
2.	Horizonpunten	3
3.	Verliesfunctie WORK	5
4.	WORK in het externe ontvouwingsmodel	6
5.	Beperkte 'smooth' monotone regressie	10
6.	Toepassen van ADDIDEAL, MONIDEAL en SMIDEAL	12
6.1	Eerste voorbeeld	13
6.1.1	ADDIDEAL en MONIDEAL	14
6.1.2	MONIDEAL en SMIDEAL	18
6.2	Tweede voorbeeld	20
7.	Bespreking van de resultaten	21
8.	Conclusie	22
	Appendices	24
	Referenties	38

HORIZONPUNTEN IN HET EXTERNE ONVOUWINGSMODEL

In deze studie wordt onderzoek gedaan naar het vermijden van horizonpunten in het externe ontvouwingsmodel. Een horizonpunt is een ideaalpunt waar de afstand tot de targetconfiguratie groter is dan de gemiddelde afstand van de targetpunten die het convexe omhulsel vormen. Om horizonpunten te vermijden worden hier twee methoden geïntroduceerd in het externe ontvouwingsmodel: de verliesfunctie WORK en beperkte 'smooth' monotone regressie. Bij toepassing is gebleken dat beperkte 'smooth' monotone regressie een methode is om horizonpunten te vermijden.

1. Inleiding

In het ontvouwingsmodel worden de relaties tussen individuen en stimuli beschreven door zowel de stimuli als de individuen als punten in een ruimtelijk model af te beelden. De grootte van de afstanden tussen het individu (het ideaalpunt) en de stimuluspunten komt overeen met de voorkeuren van het individu, met andere woorden het ideaalpunt heeft de kleinste afstand tot het stimuluspunt van meeste voorkeur en de grootste afstand tot het stimuluspunt van minste voorkeur (Meerling, 1981).

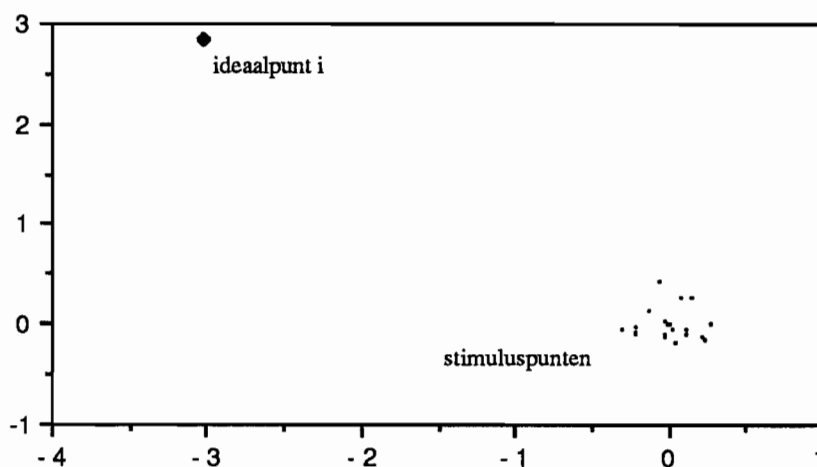
Met betrekking tot het afbeelden van de ideaalpunten en stimuluspunten zijn er twee vormen: *interne* analyse en *externe* analyse. Bij interne analyse worden zowel de ideaalpunten als de stimuluspunten gelijktijdig, rekening houdende met elkaar, vastgelegd ('reciprocal relocation'). De andere vorm is externe analyse. Hierbij hebben de stimuluspunten een vaste plaats in de ruimte en de ideaalpunten moeten hierin geplaatst worden (Heiser, 1987). Het onderwerp van deze studie heeft betrekking op het externe ontvouwingsmodel.

De soort data waarvoor ontvouwingsmodellen worden gebruikt, zijn onder meer preferentierangordeningen. Opgeslagen in een datamatrix blijkt dit een onvolledige datamatrix oftewel een deelmatrix van een volledige gelijkheidsmatrix te zijn. Er is alleen kennis over de gelijkheidsgegevens tussen een object en een set van stimuluspunten, en tussen de stimuluspunten, maar er is niets bekend over de relaties tussen de objecten onderling (Borg, 1981).

Een tot nu toe veel gebruikte techniek voor externe analyse is PREFMAP (Carroll & Chang, 1967). Er zijn echter twee nadelen verbonden aan deze techniek, namelijk het voorkomen van anti-ideaalpunten in plaats van ideaalpunten en het tweede nadeel is het voorkomen van punten die heel ver van de targetconfiguratie gaan afliggen. Door het SMACOF algoritme (Heiser en de Leeuw, 1979) toe te passen in het externe ontvouwingsmodel, dus gebruik te maken van een afstandsbenadering, is het eerste nadeel verdwenen. Voor wat betreft het tweede nadeel, punten die heel ver van de

targetconfiguratie gaan afliggen, heeft Heiser een extern ontvouwings algoritme ontwikkeld (ADDIDEAL) met het SMACOF algoritme waarin de verliesfunctie WORK (Heiser, 1987) wordt gebruikt. Bij toepassing van dit programma komen er echter nog steeds deze soort punten voor. Als zo'n ideaalpunt ver van de targetconfiguratie gaat afliggen dan wordt laatstgenoemde in elkaar gedrukt met als gevolg dat de punten van de targetconfiguratie nauwelijks van elkaar te onderscheiden zijn, evenals de volgorde van de stimuli ten opzichte van dit ideaalpunt (zie Figuur 1). Een gevolg is dat de relatie tussen het individu (weergegeven als ideaalpunt) en de stimuli (de punten van de targetconfiguratie) niet duidelijk kan worden weergegeven. Ideaalpunten die ver van de targetconfiguratie gaan afliggen worden hier voortaan horizonpunten genoemd.

Het onderwerp van deze studie is onderzoeken waar horizonpunten vandaan komen en een oplossing proberen te vinden om ze te vermijden. Horizonpunten worden gedefinieerd als ideaalpunten die 'aanzienlijk' buiten het convex omhulsel terechtkomen. Met aanzienlijk wordt hier bedoeld dat de afstand van het ideaalpunt ten opzichte van het convex omhulsel groter is dan de gemiddelde afstand van de punten die het convex omhulsel vormen.



Figuur 1 Ideaalpunt i is een horizonpunt

De indeling van deze studie is als volgt. In hoofdstuk 2 wordt dieper ingegaan op het verschijnsel horizonpunten. Vervolgens wordt in hoofdstuk 3 de verliesfunctie WORK geïntroduceerd. In hoofdstuk 4 wordt deze verliesfunctie in een externe ontvouwingsmodel ingepast en op basis hiervan zijn twee programma's ontwikkeld: een voor data op metrisch (ratio) en een voor data op niet-metrisch (ordinaal) meetniveau. In de volgende hoofdstuk, 5, wordt een beschrijving gegeven van de 'smooth' monotone regressie benadering en hoe dit wordt ingepast in het externe ontvouwingsmodel. In hoofdstuk 6 zijn de resultaten beschreven na toepassing van de programma's op twee voorbeelden met preferentiedata. Hierbij wordt ingegaan op het belang van de

beginconfiguratie en of er horizonpunten zijn voorkomen. In hoofdstuk 7 zijn de resultaten uit hoofdstuk 6 besproken. De studie wordt afgesloten met de conclusie.

2. Horizonpunten

In de vorige sectie zijn horizonpunten gedefinieerd als ideaalpunten die aanzienlijk buiten het convex omhulsel terecht komen. De eerste vraag die zich hier opwerpt is waar horizonpunten vandaan komen. Over het algemeen wordt aangenomen dat een horizonpunt wordt veroorzaakt doordat er binnen de targetconfiguratie geen punt bestaat die voldoet aan de eisen die door een ideaalpunt worden gesteld. Met andere woorden: de voorkeuren van beoordelaar i zijn zodanig dat er geen ideaalpunt i bestaat in de targetconfiguratie waarvan de grootte van de afstanden tot de stimuluspunten overeenkomen met de voorkeuren van beoordelaar i . Aan de hand van een voorbeeld wordt dit geïllustreerd.

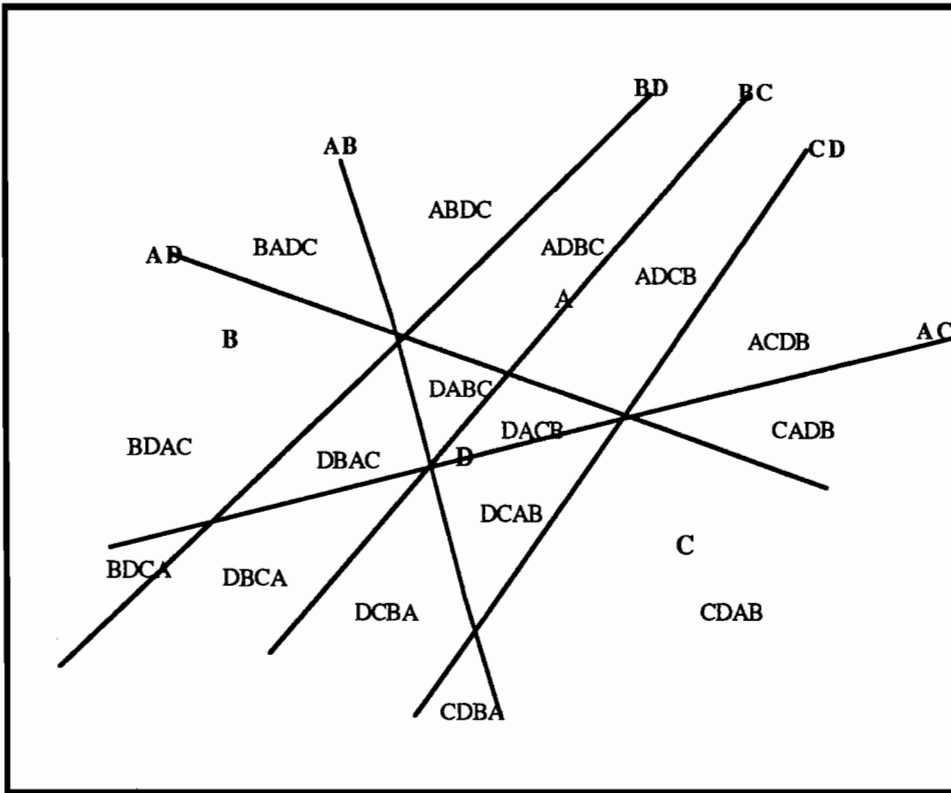
Stel een tweedimensionele targetconfiguratie bestaande uit een viertal punten: A(1,3) B(-5,2) C(3,-3) D(-1,-1). Het vlak kan in segmenten worden verdeeld door middel van het tekenen van middelloodlijnen tussen paren van stimuluspunten. Dit is in Figuur 2 afgebeeld.

Deze segmenten worden isotone gebieden genoemd: een isotoon gebied is een verzameling punten met dezelfde rangorde van afstanden tot de stimuluspunten. Er zijn vier stimuli dus $4! = 24$ voorkeursmogelijkheden. In de tekst wordt gerefereerd naar voorkeursrangordeningen of -mogelijkheden. Hiermee wordt het volgende bedoeld. Stel dat een willekeurige volgorde van 4 stimuli ABCD is, dan betekent dit hier dat A boven B wordt verkozen, B boven C etc. D is in dit geval de stimulus van de minste voorkeur.

Door de ligging van de stimuluspunten in de targetconfiguratie zijn er 18 voorkeursmogelijkheden, 18 mogelijkheden om 4 stimuli te rangordenen, die passen in het model. Dus het model (de targetconfiguratie) bevat 18 isotone gebieden, waarbij in elk gebied een verzameling punten ligt waarvan de grootte van de afstanden tot de stimuluspunten overeenkomen met een van de 18 voorkeuren. Een beoordelaar met een van deze 18 voorkeursmogelijkheden kan als een ideaalpunt in de targetconfiguratie perfect worden weergegeven, namelijk in het bijbehorende isotone gebied.

In de targetconfiguratie zijn zes gesloten -en 12 open gebieden te onderscheiden. In het algemeen geldt dat als n stimuli in $n-1$ dimensies zijn weergegeven alle rangordeningen voorkomen in de afbeelding. Dus alle rangordeningen van 4 stimuli kunnen in drie dimensies worden weergegeven. In dat geval is elk isotoon gebied open. Als n stimuli in minder dan $n-1$ dimensies worden afgebeeld betekent het dat niet alle rangordeningen van n stimuli kunnen worden weergegeven. In het laatste geval komen er gesloten gebieden voor. Dit onderscheid is van wezenlijk belang omdat elk open gebied een 'mirror image', een spiegelbeeld, heeft (Coombs, 1964). Bijvoorbeeld het open gebied BDAC heeft als spiegelbeeld CADB. De gesloten gebieden hebben geen 'mirror image' in de data. Het aantal gesloten gebieden geeft een aantal voorkeursrangordeningen aan die niet

voorkomen in de targetconfiguratie. Dit voorbeeld heeft zes gesloten gebieden; er zijn in ieder geval zes voorkeursrangordeningen die niet voorkomen. Er kunnen meer gebieden



Figuur 2 Een configuratie van de vier punten A, B, C en D in twee dimensies.

zijn die niet voorkomen in de targetconfiguratie. Als er 4 stimuli of minder zijn dan is het aantal gesloten gebieden gelijk aan het aantal gebieden dat niet voorkomt in de targetconfiguratie. Heeft men echter 5 of meer stimuli dan is het aantal gebieden dat niet voorkomt groter dan het aantal gesloten gebieden. Bekijken we de open gebieden dan zijn er drie gebieden (ABCD;ACBD;BDCA) die ver aflaggen van de targetconfiguratie. De eerstgenoemde definitie van 'aanzienlijk' buiten het convex omhulsel, is voor deze gebieden ook toepasbaar. Punten in deze drie gebieden liggen aanzienlijk buiten het convex omhulsel. Een beoordelaar i met een van deze drie voorkeuren zal een horizonpunt i hebben in plaats van een ideaalpunt i . Dat deze drie gebieden zo ver aflaggen komt door de ligging van de punten in de targetconfiguratie. De drie punten B C en D liggen bijna op een rechte lijn. Als de punten op een rechte lijn liggen dan lopen alle drie de middelloodlijnen parallel en ontmoeten elkaar in oneindig. Dus de vorm van de targetconfiguratie (Coombs, 1964) kan horizonpunten zelf veroorzaken. Aan de hand van dit voorbeeld is duidelijk dat er onderscheid kan worden gemaakt in de aard van horizonpunten. Enerzijds kunnen horizonpunten optreden als er geen punt i

bestaat in de targetconfiguratie waarvan de grootte van de afstanden tot de stimuluspunten overeenkomen met de voorkeuren van beoordelaar i . Met andere woorden: beoordelaar i heeft een niet in het model passende voorkeurrangorde, zoals we later zullen zien. Anderzijds zijn er voorkeuren die wel in het model passen, maar de punten die deze voorkeuren weergeven liggen ver van het convex omhulsel. Deze punten moeten wel in het model terug worden gevonden. Van welk belang dit onderscheid is, wordt in een later hoofdstuk behandeld.

3. Verliesfunctie WORK

Een verliesfunctie in MDS geeft aan in hoeverre de data op voldoende wijze wordt afgebeeld in afstanden. Hoe hoger de waarde van de verliesfunctie, hoe slechter de oplossing. Er zijn verschillende soorten MDS verliesfuncties. Een overzicht hiervan wordt gegeven in Heiser (1988), die een onderscheid maakt tussen enerzijds *absoluut versus relatief* en anderzijds *absoluut versus gekwadrateerd*. De verschillende verliesfuncties kunnen in termen van de eerst genoemde tegenstelling worden beschreven. In de tweede tegenstelling, absoluut versus gekwadrateerd, wordt met absoluut bedoeld dat in plaats van de residuen te kwadrateren, het absolute residu wordt genomen. In deze studie wordt dit onderscheid niet verder uitgewerkt. Voor verdere literatuur zie Heiser (1987). In de context van dit onderzoek wordt met absoluut bedoeld het verschil tussen de dissimilarities en de afstanden ($\delta_{ij}-d_{ij}$) tegenover de verhouding tussen deze twee waarden (δ_{ij}/d_{ij}). In dit onderzoek wordt een relatieve verliesfunctie gebruikt. De verliesfunctie WORK is gebaseerd op een idee van McGee (1966) en ziet er als volgt uit:

$$W = \sum_i \sum_j [1 - d(x_i, y_j) / \delta_{ij}(\alpha)]^2 \quad (1)$$

Deze verliesfunctie kan worden herleid tot de standaard kleinste kwadraten verliesfunctie door de functie te herschrijven door factor $1/\delta_{ij}^2(\alpha)$ buiten haakjes te halen. Daaruit volgt:

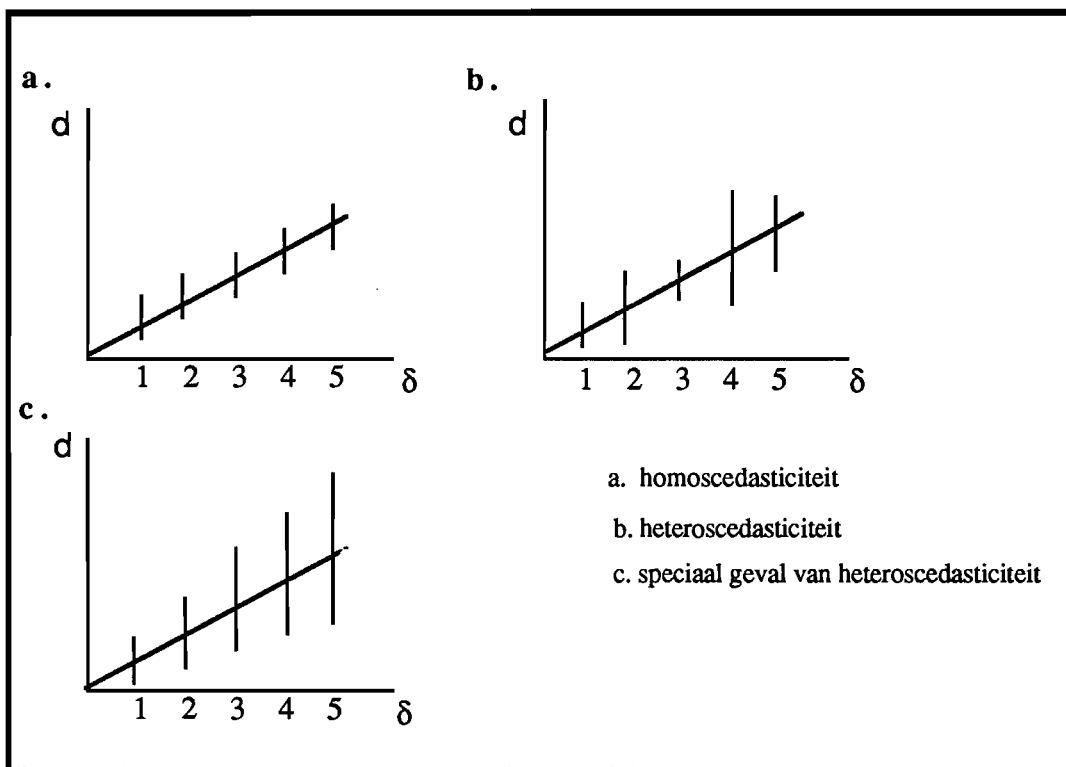
$$W = \sum_i \sum_j 1/\delta_{ij}^2(\alpha) \{ \delta_{ij}(\alpha) - d(x_i, y_j) \}^2 \quad (2)$$

Het idee achter deze verliesfunctie is dat kleinere afstanden beter gefit worden dan grotere afstanden; de grootheden $1/\delta_{ij}^2(\alpha)$, de gewichten genoemd, zorgen voor een meer nauwkeurige schatting van de kleine afstanden. Dus kleine dissimilarities hebben een zwaarder gewicht in de verliesfunctie dan grote dissimilarities.

Bij het absolute verschil tussen δ_{ij} en d_{ij} wordt *homoscedasticiteit* verondersteld. Bij de verschillende waarden van de dissimilarities is het bereik van de afstanden binnen een specifieke waarde van de dissimilarities even groot als voor een andere specifieke waarde van de dissimilarities. Zie figuur 3a.

Bij heteroscedasticiteit wordt verondersteld dat het bereik van de afstanden binnen een specifieke waarde van de dissimilarities niet gelijk is tussen de verschillende

dissimilarities; bij verschillende dissimilarities zal ook het bereik van de afstanden die horen bij de dissimilarity verschillend zijn. Met andere woorden het bereik van de afstanden binnen de dissimilarity met waarde 6 is niet gelijk aan het bereik van de afstanden binnen de dissimilarity met waarde 3. Zie Figuur 3b. Een specifiek geval van heteroscedasticiteit (Figuur 3c) is dat bij stijgende dissimilarities het bereik van de afstanden die horen bij de dissimilarities groter worden. Met de verliesfunctie WORK wordt met dit laatste geval van heteroscedasticiteit rekening gehouden doordat de kleine afstanden worden beter worden gefit dan de grote afstanden.



Figuur 3 Homoscedasticiteit en heteroscedasticiteit

4. WORK in het externe ontvouwingsmodel

Het algemene externe ontvouwingsprobleem kan als volgt worden beschreven. Stel er zijn m -punten in een p -dimensionele ruimte beschikbaar in de $m \times p$ matrix Y (de j^{de} rij wordt genoteerd met y_j .) Deze m -punten liggen vast en noemen we hier de targetconfiguratie. Vervolgens zijn er n onbekende punten, geïndexeerd door i met coördinaten x_i verzameld in de rijen van de $n \times p$ matrix X . Deze onbekende punten moeten in de targetconfiguratie

worden ingepast. Hierbij wordt de euclidische afstandsfunctie $d(x,y)$ gebruikt, gedefinieerd door

$$d^2(x_i,y_j)=(x_i-y_j)'(x_i-y_j)$$

De n onbekende punten van matrix \mathbf{X} moeten worden berekend vanuit de data. De data zijn voorkeurgegevens over m stimuli van n beoordelaars. De kleinste beoordeling geeft de grootste voorkeur weer: de beoordelingen worden dissimilarities genoemd. Deze dissimilarities staan in een matrix, \mathbf{Z} , waarbij voor elke rij een punt met coördinaten wordt gezocht die zodanig past in de targetconfiguratie dat de grootte van de afstanden overeenkomen met de dissimilarities van de bijbehorende rij.

Om een optimale oplossing te vinden voor matrix \mathbf{Z} , wordt de verliesfunctie $WORK$ geminimaliseerd. Dit wordt aangepakt door middel van het *Alternating Least Squares* algoritme (De Leeuw, Young & Takane, 1976; Young, 1981). Dat wil zeggen het iteratief oplossen van het kleinste kwadraten probleem door afwisselend twee subproblemen voor verschillende (sets van) parameters op te lossen. In het eerste subprobleem wordt een optimale schaling van de dissimilarities berekend. Omdat de laatste afhankelijk is van het meetniveau is dit probleem dan ook verschillend bij een verschillend meetniveau. Het hieronder beschreven algoritme is geschikt gemaakt om zowel de data op een metrisch (ratio) als een niet-metrische (ordinaal) meetniveau te behandelen. Allereerst wordt hier het eerste geval behandeld, dus de metrische versie, en vervolgens wordt de niet-metrische versie besproken. Vervolgens zijn de parameters die in het tweede subprobleem moeten worden opgelost, de coördinaten van n punten; dus er wordt een op dat moment optimale configuratie geschat.

De twee subproblemen noemen we de *normalized regression* en de *relocation* (Heiser, 1987). In de eerst genoemde wordt gezocht naar optimale waarden voor de α , dus een optimale schaling van de dissimilarities, gegeven de vaste waarden van d_{ij} . Dit wordt bereikt door de dissimilarities te vermenigvuldigen met de factor α . De optimale keuze van α is dan:

$$\alpha = \frac{\sum_i \sum_j \{d_{ij} / \delta_{ij}\}^2}{\sum_i \sum_j \{d_{ij} / \delta_{ij}\}} \quad (3)$$

De herschaalde dissimilarities zijn dan

$$d^+_{ij} = \alpha \delta_{ij} \quad (4)$$

Nadat de herschaalde dissimilarities berekend zijn worden deze vastgehouden en de parameters in het subprobleem relocation worden opgelost. Hierin worden nieuwe locaties berekend. De verliesfunctie heeft de vorm die beschreven staat in formule (2). In dit geval zijn de gewichten gelijk aan $1/\delta^2_{ij}(\alpha)$ en zorgen voor een meer nauwkeurige schatting van de kleine afstanden. Voor het *relocation* probleem worden de volgende stappen ondernomen. De gewichten worden verzameld in matrix \mathbf{W} met de elementen

$$w_{ij}=1 / (d^{+}_{ij})^2$$

Vervolgens wordt een matrix **A** aangemaakt die informatie bevat over de grootte van de huidige afstanden, d_{ij} , en over wat deze zouden moeten zijn, d^{+}_{ij} , en de gewichten, w_{ij} . Matrix **A** heeft de volgende elementen (Heiser, 1987)

$$\begin{aligned} a_{ij} &= w_{ij} d^{+}_{ij} / d_{ij} && \text{als } d_{ij} > 0 \\ a_{ij} &= 0 && \text{als } d_{ij} = 0 \end{aligned}$$

Vervolgens worden twee diagonale matrices, **P** en **R**, gedefinieerd als

$$\begin{aligned} \mathbf{P} &= \text{diag}(\mathbf{A}e_m) \\ \mathbf{R} &= \text{diag}(\mathbf{W}e_m) \end{aligned}$$

waarbij e_m een m -vector is met enen.

Het algoritme dat hier gebruikt wordt is het SMACOF algoritme (Heiser, 1987). Samengevat bevat het algoritme de volgende operatie

$$\mathbf{X}^+ = \mathbf{R}^{-1}(\mathbf{P}\mathbf{X} - \mathbf{A}\mathbf{Y} + \mathbf{W}\mathbf{Y}) \quad (5)$$

Waarbij \mathbf{X} de matrix is waarin de schattingen van n onbekende punten staan van de vorige iteratie. \mathbf{X}^+ is de matrix die berekend moet worden, dus hierin komen de nieuwe schattingen te staan van de n -onbekende punten. \mathbf{Y} is de matrix die de m -bekende punten bevat. Het programma dat bovenstaand algoritme bevat heet ADDIDEAL.

Als de data op niet-metrisch niveau wordt behandeld, blijft het subprobleem relocation hetzelfde, maar het tweede subprobleem verandert en dit noemen we *monotone regression*. Hierbij gaat het niet om een optimale waarde van α te vinden, maar om een zodanige transformatie van de dissimilarities te vinden, dat er een monotone relatie tussen de pseudo-afstanden en de dissimilarities ontstaat (Kruskal, 1964). De pseudo-afstanden worden genoteerd als \hat{d} . Voor het berekenen hiervan wordt de verliesfunctie zodanig herleid, dat de gekwadrateerde afstanden als gewichten in de functie voorkomen.

$$W = \sum_i \sum_j d^2_{ij} (1/d_{ij} - 1/\delta_{ij}(\alpha))^2 \quad (6)$$

In het normalized regression probleem worden de volgende stappen ondernomen. De gewichten worden verzameld in matrix \mathbf{W}^* met de elementen

$$w^*_{ij} = d^2_{ij}$$

De verliesfunctie (2) is zodanig herleid dat er in plaats van d_{ij} en δ_{ij} , $1/d_{ij}$ en $1/\delta_{ij}$ staat. Hierdoor komt, volgens een suggestie van De Leeuw (1986), de invoer voor de monotone regressie er als volgt uit te zien. De volgende monotoniceits eis geldt:

$$\delta_{ij} \leq \delta_{ik} \leq \dots \leq \delta_{ir} \quad \Leftrightarrow \quad \hat{d}_{ij} \leq \hat{d}_{ik} \leq \dots \leq \hat{d}_{ir} \quad (7)$$

Er moet een zodanige transformatie van de dissimilarities plaats vinden dat de rangordening gelijk blijft. Dit toegepast op onze speciale verliesfunctie geeft het volgende

$$\begin{aligned} \delta_{ij} \leq \delta_{ik} & \Leftrightarrow \hat{d}_{ij} \leq \hat{d}_{ik} \\ 1/\delta_{ik} \leq 1/\delta_{ij} & \Leftrightarrow 1/\hat{d}_{ik} \leq 1/\hat{d}_{ij} \end{aligned}$$

Stel $\beta_{ij} = 1/\delta_{ij}$ en $\zeta_{ij} = 1/\hat{d}_{ij}$, daaruit volgt

$$\beta_{ik} \leq \beta_{ij} \quad \Leftrightarrow \quad \zeta_{ik} \leq \zeta_{ij} \quad (8)$$

Duidelijk is dat de grootste dissimilariteit door gebruik van deze verliesfunctie de kleinste wordt. Aan de hand van onderstaand voorbeeld, Tabel 1, wordt duidelijk gemaakt hoe de fase van monotone regressie werkt.

Tabel 1 Een voorbeeld van monotone regressie.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
δ	d	W	$1/\delta$	$1/d$	β	ζ	W^*	z	$1/z$	\hat{d}
1	3	9	1	1/3	1/5	1/4	16	0.195	5.138	2.519
2	1	1	1/2	1	1/4	1/6	36	0.195	5.138	2.519
3	3	9	1/3	1/3	1/3	1/3	9	0.330	3.030	3.030
4	6	36	1/4	1/6	1/2	1	1	0.397	2.519	5.138
5	4	16	1/5	1/4	1	1/3	9	0.397	2.519	5.138

In kolom (1) staan de voorkeuren (δ), in kolom (2) de afstanden (d). In kolom (3) worden de gewichten berekend. Dit zijn de gekwadrateerde afstanden, dus de getallen van kolom (2) zijn gekwadrateerd. In kolom (4) en (5) worden δ en d door één gedeeld. In kolom (6) worden de $1/\delta$ op volgorde van klein naar groot gezet, de grootste dissimilariteit is nu de kleinste geworden. In kolommen (7) en (8) worden respectievelijk de afstanden en de gewichten op volgorde van deze 'nieuwe' data, dus van kolom (6),

gezet. De data zijn nu klaar om in de monotone regressie te worden ingevoerd. Hierin wordt het volgende gedaan. In kolom (7) voldoen de eerste twee afstanden niet aan de ordeningseisen, $1/4 (=0.25)$ is groter dan $1/6 (=0.17)$. Om deze volgorde kloppend te maken, zodat er aan de eisen wordt voldaan, is de volgende operatie uitgevoerd: $(16 \times 0.25 + 36 \times 0.17) / 16 + 36 = 0.195$. De beide afstanden worden opgeteld en gedeeld door de som van de afstanden. Dit wordt gedaan voor de hele kolom (7). In kolom (9) staat het resultaat van de monotone regressie. Vervolgens wordt, in kolom (10), dit resultaat gedeeld door 1, zoals ook in de kolommen (4) en (5) gebeurd is. In kolom (11) worden de resultaten op de volgorde van kolom (1) gezet en dit zijn de berekende pseudo-afstanden.

In dit bovenstaande voorbeeld bevatten de data geen 'ties', dat wil zeggen er komen geen gelijke dissimilarities voor. In veel sociaal-wetenschappelijk onderzoek kunnen gelijke beoordelingen voor verschillende stimuli worden gegeven, waardoor er gelijke dissimilarities zijn. In de literatuur zijn twee benaderingen bekend voor dit geval, namelijk *primary approach* en *secondary approach*. In de eerstgenoemde benadering hoeven gelijke dissimilarities niet noodzakelijkerwijze in dezelfde pseudo-afstanden te worden afgebeeld. Het maakt niet uit welke van de twee afstanden korter is en of deze gelijk zijn of niet. In de verliesfunctie wordt dit niet als verlies meegerekend. De wijze om dit te bereiken is door geen beperkingen op te leggen aan de \hat{d} :

$$\begin{array}{ll} \delta_{ij} < \delta_{ik} & \hat{d}_{ij} \leq \hat{d}_{ik} \\ \delta_{ij} = \delta_{ik} & \hat{d}_{ij} \leq \hat{d}_{ik} \quad \text{of} \quad \hat{d}_{ij} \geq \hat{d}_{ik} \end{array}$$

Voor wat betreft de *secondary approach* geldt dat bij het voorkomen van gelijke dissimilarities er gelijke pseudo-afstanden *moeten* zijn. In dit geval is er wel sprake van een effect op de configuratie als $d_{ij} \neq d_{ik}$; Het wordt als verlies meegerekend in de verliesfunctie. Dit wordt bereikt door beperkingen op te leggen aan de \hat{d} (Kruskal, 1964):

$$\begin{array}{ll} \delta_{ij} < \delta_{ik} & \hat{d}_{ij} \leq \hat{d}_{ik} \\ \delta_{ij} = \delta_{ik} & \hat{d}_{ij} = \hat{d}_{ik} \end{array}$$

Het programma dat deze monotone regressie bevat is MONIDEAL. De beide benaderingen zijn als opties opgenomen in dit programma.

5. Beperkte smooth monotone regressie

In dit hoofdstuk wordt eerst de 'smooth' monotone regressie benadering beschreven en vervolgens de beperkte smooth monotone regressie. Het introduceren van deze benadering in het externe ontvouwingsmodel, waarin WORK is toegepast, heeft als doel horizonpunten te voorkomen.

Smooth monotone regressie is monotone regressie zoals geformuleerd in hoofdstuk 3, maar met meerdere restricties. Stel $r_1 \dots r_m$ zijn de indexen die de dissimilarities ordenen van klein naar groot, dan worden de gewone monotoniteitseisen gedefinieerd als

$$\hat{d}(r_1) \leq \hat{d}(r_2) \leq \dots \leq \hat{d}(r_m) \quad (9)$$

Alleen de restrictie van rangordering wordt geëist. Met andere woorden als de dissimilarities worden getransformeerd moet de rangorde gehandhaafd blijven. De waarden die de pseudo-afstanden (\hat{d}) kunnen aannemen is onbeperkt zolang aan de bovenstaande restrictie wordt voldaan. Het idee achter de 'smooth' monotone regressie is om juist beperkingen op te leggen aan de waarden die de pseudo-afstanden aannemen. Om een onderscheid te maken tussen de bovengenoemde pseudo-afstanden (\hat{d}) en de pseudo-afstanden die door middel van de 'smooth' monotone regressie worden berekend, zijn de laatste *toegestane* pseudo-afstanden genoemd (Heiser, 1985) en hun notatie is γ . De 'smoothness' benadering richt zich op de stappen tussen de opeenvolgende γ 's, de toegestane pseudo-afstanden. Voor $s=1, \dots, m$ worden de toegestane stappen als volgt gedefinieerd

$$\theta_s = \gamma_s - \gamma_{s-1} \quad (10)$$

De pseudo-afstanden worden aan nul gebonden. Behalve een ondergrens waar de pseudo-afstanden aan zijn gebonden, wordt tevens een bovengrens gedefinieerd. Een essentieel aspect in de 'smoothness' benadering is dus de keuze van de onder- en bovengrens. Deze moeten bepaald worden aan de hand van een intern criterium. Dit kan worden bereikt door het volgende te eisen

$$|\theta_s - \theta_{s-1}| \leq \theta \quad (11)$$

waarbij

$$\theta = 1/m \sum_{t=1}^i \theta_t \quad \text{en} \quad \theta_0 = 0$$

waarbij θ de gemiddelde stap is. Het verschil tussen twee opeenvolgende stappen moet kleiner of gelijk zijn aan de gemiddelde stap. Niet de toegestane pseudo-afstanden zelf worden beperkt, maar het verschil tussen de opeenvolgende pseudo-afstanden; met andere woorden niet de eerste, maar de tweede orde verschillen worden begrensd. Invariantie van de monotone transformatie van δ is gegarandeerd door de volgorde waarop de δ 's (11) ingaan. Telkens wordt het verschil genomen tussen twee opeenvolgende stappen. De berekende γ 's worden als het ware naar elkaar toe getrokken.

De 'smoothness'-eisen kunnen in termen van toegestane pseudo-afstanden worden geschreven door de γ in te vullen in (11) waaruit volgt

$$-1/m \gamma_m \leq \gamma_s - 2\gamma_{s-1} + \gamma_{s-2} \leq 1/m \gamma_m \quad (12)$$

In deze formule zijn de smoothness restricties aangegeven. Stel dat we m stimuli hebben dan kunnen we de de monotoniciteitseisen (9) en de 'smoothness'-eisen (12) in een $(3m-1) \times m$ matrix D , waarbij de eerste m rijen zijn gedefinieerd door de monotoniciteitseisen en de rest van de rijen worden gedefinieerd door de 'smoothness'-condities.

Het smooth monotone regressie probleem kan worden geformuleerd als het vinden van een toegestane pseudo-afstand, zodanig dat de volgende functie wordt geminimaliseerd:

$$\tau(\gamma) = \|\gamma - d\|_{\mathbf{W}}^2 = \text{tr}(\gamma - d)' \mathbf{W} (\gamma - d) \quad (13)$$

\mathbf{W} is een diagonale $m \times m$ matrix die de niet-negatieve gewichten bevat. Het 'smooth' monotone regressie probleem is beschreven door Heiser (1985). In appendix A staat beschreven hoe zijn methode wordt gevolgd om een algoritme af te leiden.

De beperkte 'smooth' monotone regressie is alleen verschillend van de 'smooth' monotone regressie voor wat betreft het aantal 'smoothness' condities. De matrix D is hier een $(m+1) \times m$ matrix, waarbij de eerste m rijen gedefinieerd zijn door de monotoniciteitseisen en de laatste rij door een 'smoothness'-conditie. Deze 'smoothness'-eis kan als volgt worden gedefinieerd:

$$\gamma_l \leq 1/m \gamma_m \quad (14)$$

Dus de kleinste pseudo-afstand mag niet groter worden dan $1/m \gamma_m$ en deze blijft klein, waardoor de andere pseudo-afstanden ook klein blijven. De reden voor een beperkte 'smooth' monotone regressie is van pragmatische aard: in de praktijk blijkt dat als er een groot aantal stimuli is, het algoritme langzaam werkt en veel rekentijd (en geld) verbruikt.

De beperkte 'smooth' monotone regressie wordt als volgt ingepast. In sectie 4 is beschreven dat de verliesfunctie WORK in het externe ontvouwingsmodel wordt geminimaliseerd door middel van het oplossen van twee subproblemen. Het eerste subprobleem is de relocation en dit is ook hier een van de twee subproblemen. Het tweede subprobleem is de beperkte 'smooth' monotone regressie, voortaan genoteerd als 'smoothness'. Het algoritme dat de beide subproblemen bevat heet SMIDEAL. Het algoritme is alleen nog geschikt voor data waarin geen gelijke beoordelingen in voorkomen.

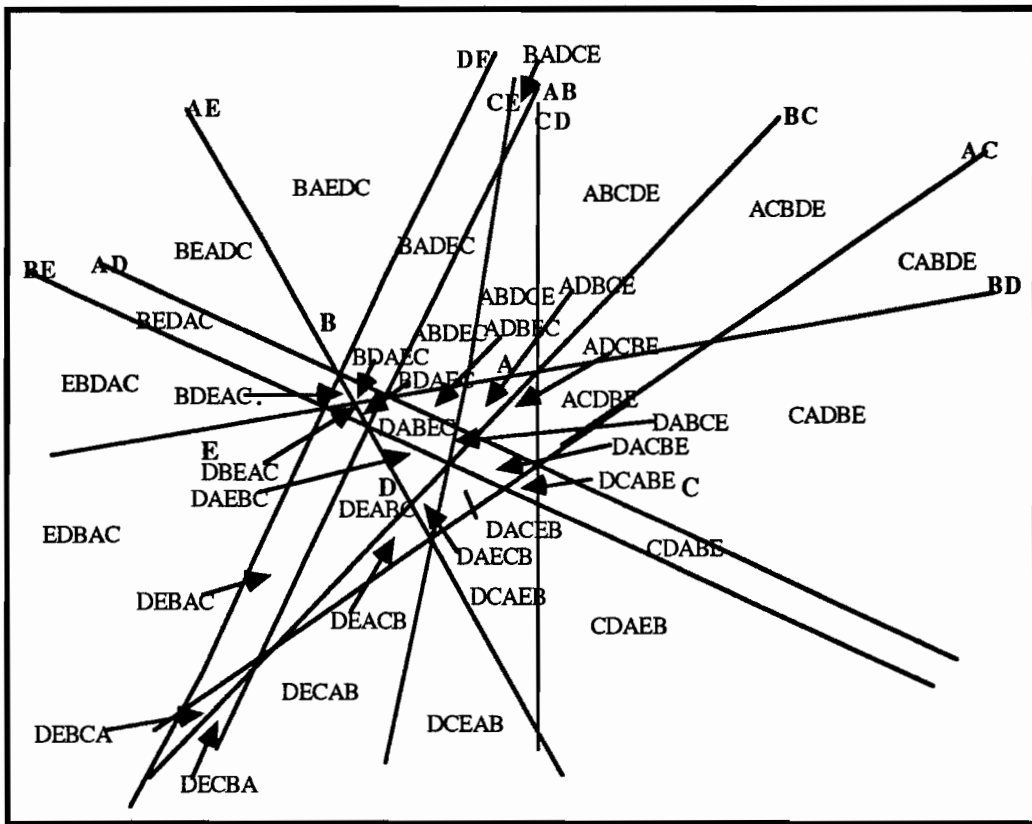
6. Toepassen van ADDIDEAL, MONIDEAL en SMIDEAL

Dit hoofdstuk is verdeeld in twee delen: in het eerste deel worden de drie programma's ADDIDEAL, MONIDEAL en SMIDEAL toegepast op een kunstmatig voorbeeld; in dit geval is er een willekeurig targetconfiguratie gekozen. In het tweede deel worden twee

programma's, ADDIDEAL en MONIDEAL toegepast op data die in een onderzoek van Vonk (1988) zijn verzameld.

6.1 Eerste voorbeeld

Het eerste voorbeeld bevat een willekeurig gekozen targetconfiguratie van vijf targetpunten. De vijf punten zijn A (0,2) B (-3,3) C (3,-1) D (-2,-1) E (-5,0). Figuur 6.1 is een afbeelding van deze targetconfiguratie, waarin ook de isotone gebieden zijn aangegeven. Er zijn in totaal 41 isotone gebieden, 20 open en 21 gesloten gebieden. In ieder geval weten we een aantal rangordeningen (21) die niet voorkomen in het model.



Figuur 4 Een configuratie van vijf punten A, B, C, D en E in twee dimensies.

De datamatrix is als volgt samengesteld. Van de 41 isotone gebieden zijn een willekeurig aantal gebieden, 8, gekozen. Deze isotone gebieden fungeren als data voor de drie programma's: bijvoorbeeld het open isotone gebied BAEDC geldt als een voorkeur van een individu i en er wordt gekeken of het ideaalpunt i , na toepassing van de drie programma's in het bijbehorende isotone gebied terecht komen. Omdat het aantal stimuli hier laag is, vijf, wordt heel duidelijk in welk isotone gebied een ideaalpunt terecht moet

komen en of zo'n ideaalpunt daar ook terecht komt. Tevens zijn er een aantal willekeurige voorkeuren gekozen die niet passen in de targetconfiguratie. In dit geval zijn de spiegelbeelden van drie gesloten isotone gebieden (DABCE, ACDBE, CBAED) gekozen. Door middel van het toepassen van de drie programma's willen we vier zaken onderzoeken. Ten eerste willen we weten of de ideaalpunten in de bijbehorende isotone gebieden terecht komen. Ten tweede of er horizonpunten voorkomen en zo ja, wanneer deze optreden. Vervolgens willen we controleren of er sprake is van lokale minima. Dit wordt gedaan door de beginconfiguratie willekeurig te variëren. Dit laatste is gedaan voor de beide programma's ADDIDEAL en MONIDEAL. Ten vierde, in hoeverre de resultaten van de programma's MONIDEAL en SMIDEAL van elkaar verschillen, omdat het beide data op ordinaal meetniveau betreft.

Eerst worden de programma's ADDIDEAL en MONIDEAL vergeleken, waarbij de eerste drie genoemde zaken worden onderzocht. Vervolgens worden MONIDEAL en SMIDEAL vergeleken, waarbij het vierde punt wordt onderzocht. Voor de overzichtelijkheid is onderscheid gemaakt in het soort gebied waaruit de voorkeur is gekozen, een open of een gesloten gebied, en er is een aparte tabel voor de niet-passende voorkeur.

6.1.1 ADDIDEAL en MONIDEAL

Allereerst worden vier voorkeuren onderzocht waarvan de ideaalpunten in de bijbehorende *open* gebieden zouden moeten liggen, als er sprake zou zijn van een perfecte oplossing. De vier willekeurig gekozen voorkeuren zijn BAEDC, BEADC, BEDAC en EDCBA. De laatste voorkeur is met name interessant omdat het bijbehorende isotone gebied ver van het convexe omhulsel afligt. In onderstaande Tabel 2 zijn de resultaten vermeld na toepassing van ADDIDEAL. In de eerste kolom in Tabel 2 staat voor elke voorkeur de beginconfiguratie aangegeven en de coördinaten. Deze beginconfiguratie is voor zowel ADDIDEAL als MONIDEAL hetzelfde. De beginconfiguratie wordt zodanig gekozen dat het aanvangspunt van elke voorkeur ligt in het zwaartepunt van de eerste twee voorkeuren. Dus bij de voorkeur BAEDC wordt het aanvangspunt gelegd in het zwaartepunt BA, van de voorkeur BEADC in BE etc.. Deze berekening van de beginconfiguratie wordt in de beide programma's standaard gebruikt. Dit wordt voortaan de standaard beginconfiguratie genoemd. Het resultaat is weergegeven in kolom (2). Dit zijn de gebieden waarin de ideaalpunten uiteindelijk zijn terechtgekomen, gegeven de beginconfiguratie in kolom (1). Alleen het tweede ideaalpunt komt terecht in het juiste isotone gebied. Er komen geen horizonpunten voor. Als de beginconfiguratie wordt gevarieerd (waarbij het aanvangspunt voor alle voorkeuren gelijk is), kolommen (3) t/m (5), blijken de resultaten niet van elkaar te verschillen, op een uitzondering na. Voor het eerste ideaalpunt zijn de resultaten beter als een andere beginconfiguratie wordt gekozen, bij de andere drie ideaalpunten is er geen verschil.

Voor MONIDEAL staan de resultaten van deze vier voorkeuren in Tabel 3. De beginconfiguratie in kolom (1) is hetzelfde als bij ADDIDEAL; in dit geval wordt dus de standaard beginconfiguratie gebruikt. In kolom (2) is te zien dat de eerste drie

ideaalpunten in het bijbehorende gebied worden neergelegd; het verlies voor alle drie is dan ook nul. In hoofdstuk 7 wordt dit resultaat verder besproken. Het eerste ideaalpunt wordt op de lijn ED gelegd, met andere woorden het ideaalpunt heeft dezelfde afstand tot E als tot D. In het programma worden voor wat dit betreft geen eisen gesteld. Het vierde

Tabel 2 Resultaten ADDIDEAL voor vier voorkeuren. Hier wordt weergegeven het gebied waarin het ideaalpunt valt, de coördinaten van het ideaalpunt en het verlies.

voorkeur	beginconfiguratie				
	(1) standaard beginconf	(2) resultaat beginconf	(3) CADBE (2.0,0.0)	(4) EDCBA (-6.5,-9.0)	(5) DAEBC (-1.5,-0.5)
BAEDC	BADEC (-2.4,1.7)	BADEC (-2.1, 2.4) .29	BAEDC (-2.3,4.4) .18	BAEDC (-2.3,4.4) .18	BADEC (-2.1,2.4) .29
BEADC	BEDAC (-2.9,1.5)	BEADC (-4.6,3.5) .24	BEADC (-4.6,3.5) .24	BEADC (-4.6,3.5) .24	BEADC (-4.6,3.5) .24
BEDAC	BEDAC (-3.1,1.2)	BEADC (-4.6,3.0) .27	BEADC (-4.6,3.0) .27	BEADC (-4.6,3.0) .27	BEADC (-4.6,3.0) .27
EDCBA	DEABC (-2.3,-0.2)	EDBAC (-5.8,-2.4) .60	EDBAC (-5.8,-2.4) .60	EDBAC (-5.8,-2.4) .60	EDBAC (-5.8,-2.4) .60

Tabel 3 Resultaten MONIDEAL voor vier voorkeuren. Hier wordt weergegeven het gebied waarin het ideaalpunt valt, de coördinaten van het ideaalpunt en het verlies.

voorkeur	beginconfiguratie				
	(1) standaard beginconf	(2) resultaat beginconf	(3) CADBE (2.0,0.0)	(4) EDCBA (-6.5,-9.0)	(5) DAEBC (-1.5,-0.5)
BAEDC	BADEC (-2.4,1.7)	BAEDC (-2.7,1.8) .00	BAEDC (-2.7,1.8) .00	DECAB (-10.8,-46.8) .07	BAEDC (-2.7,1.8) .00
BEADC	BEDAC (-2.9,1.5)	BEADC (-2.8,1.7) .00	DACBE (-0.4,-0.3) .70	DECAB (-10.8,-46.9) .06	BDAEC (-2.8,1.7) .001
BEDAC	BEDAC (-3.1,1.2)	BEDAC (-3.1,1.2) .00	DACBE (-0.4,-0.3) .70	EDBCA (-28.9,-29.9) .06	BEDAC (-2.9,1.4) .00
EDCBA	DEABC (-2.3,-0.2)	DEBAC (-5.3,-6.6) .02	DACBE (-0.4,-0.3) .70	EDCBA (-6.5,-9.0) .00	DACBE (-0.4,-0.3) .70

ideaalpunt gaat verder van het convex omhulsel afliggen dan de andere drie ideaalpunten, maar is geen horizonpunt. Als de beginconfiguratie veranderd wordt, zoals in Tabel 2, komen er horizonpunten voor. De verschillende beginconfiguraties brengen ook verschillende resultaten met zich mee, in tegenstelling tot de resultaten in ADDIDEAL.

Vanuit de *gesloten* gebieden zijn vier willekeurige voorkeuren gekozen. Deze staan in Tabel 4 voor ADDIDEAL. Er komen geen horizonpunten voor. In de kolommen (3) t/m (5) staan de resultaten van de verschillende aanvangsconfiguraties. De ideaalpunten, behalve het eerste punt, komen telkens in hetzelfde gebied terecht bij variërende beginconfiguraties en het verlies blijft hetzelfde.

De resultaten van het programma MONIDEAL voor deze 4 voorkeuren staan in Tabel 5. Kolom (2) geeft de resultaten als de beginconfiguratie in kolom (1) wordt gebruikt. Alleen het derde ideaalpunt komt niet in het juiste gebied, en heeft ook een verlies van .03. Het eerste ideaalpunt valt op het snijpunt van de twee lijnstukken AD en CE. Dus het ideaalpunt heeft gelijke afstand tot A als tot D en het heeft een gelijke afstand tot C en tot E (maar niet tussen A en C of D en C). Als de beginconfiguratie van kolommen (3) en (5) gebruikt worden, blijken horizonpunten het resultaat te zijn. Zoals opgemerkt is over de resultaten uit tabel 5, geldt ook hier, dat bij verschillende beginconfiguraties, er verschillende resultaten zijn.

Tabel 4 Resultaten ADDIDEAL voor vier voorkeuren. Hier wordt weergegeven het gebied waarin het ideaalpunt ligt, de coördinaten van het ideaalpunt en het verlies.

voorkeur	beginconfiguratie				
	(1) standaard beginconf	(2) resultaat beginconf	(3) CDAEB (1.0,-3.0)	(4) ACDBE (1.0,1.0)	(5) DEACB (-2.0,-2.5)
DABCE	ADBEC (-1.1,0.7)	DABEC (-1.2,-0.1) .35	DABEC (-1.2,-0.1) .35	BAEDC (-1.2,-0.1) .35	BADEC (-1.2,-0.1) .35
ABDEC	ABDEC (-1.8,1.5)	ABDEC (-1.1,2.1) .06	ABDEC (-1.1,2.1) .06	ABDEC (-1.1,2.1) .06	ABDEC (-1.1,2.1) .06
DEBCA	DEBAC (-2.6,0.1)	DEBAC (-3.1,-2.0) .35	DEBAC (-3.1,-2.0) .35	DEBAC (-3.1,-2.0) .35	DEBAC (-3.1,-2.0) .35
ADBEC	ADBEC (-1.7,1.1)	ADBCE (-0.7,1.1) .27	ADBCE (-0.7,1.1) .27	ADBCE (-0.7,1.1) .27	ADBCE (-0.7,1.1) .27

Tabel 5 Resultaten MONIDEAL voor vier voorkeuren. Hier wordt weergegeven het gebied waarin het ideaalpunt ligt, de coördinaten van het ideaalpunt en het verlies.

voorkeur	beginconfiguratie				
	(1) standaard beginconf	(2) resultaat beginconf	(3) CDAEB (1.0,-3.0)	(4) ACDBE (1.0,1.0)	(5) DEACB (-2.0,-2.5)
DABCE	ADBEC (-1.1,0.7)	DABCE (-0.9,0.4) .00	DAECB (-1.0,-0.5) .001	DABCE (-0.5,-0.2) .00	DAECB (-1.0,-0.5) .001
ABDEC	ABDEC (-1.8,1.5)	ABDEC (-1.8,1.5) .00	DECAB (-9.7,-44.0) .07	ABDEC (-0.7,1.8) .00	DECAB (-9.7,-44.8) .07
DEBCA	DEBAC (-2.6,0.1)	DEACB (-3.0,-3.6) .03	DEACB (-3.3,-4.0) .02	DEACB (-2.7,-3.2) .001	DEACB (-3.2,-3.8) .02
ADBEC	ADBEC (-1.7,1.1)	ADBEC (-1.7,1.0) .00	DECAB (-9.1,-44.1) .07	ABDCE (-0.8,1.4) .00	DECAB (-9.2,-44.2) .07

Het derde gedeelte van het voorbeeld is om te kijken waar ideaalpunten gaan liggen die een *niet-passende* voorkeur hebben; dat wil zeggen deze voorkeuren komen niet voor in ons model. De drie willekeurige gekozen voorkeuren zijn CEBAD, EBDCA en CBAED. en in Tabel 6 zijn de resultaten weergegeven na toepassing van ADDIDEAL. Ook hier komen geen horizonpunten voor. Het verlies per ideaalpunt is hoger dan bij ideaalpunten met een passende voorkeur. Het programma ADDIDEAL geeft, bij de verschillende aanvangsconfiguraties, dezelfde ideaalpunten en verlies .

Tabel 6 Resultaten ADDIDEAL voor drie voorkeuren. Hier wordt weergegeven het gebied waarin het ideaalpunt ligt, de coördinaten van het ideaalpunt en het verlies.

voorkeur	beginconfiguratie				
	(1) standaard beginconf	(2) resultaat beginconf	(3) ABCDE (1.0,3.0)	(4) BEADC (-4.0,3.0)	(5) ABDEC (-1.5,0.5)
CEBAD	DABEC (-0.9,0.4)	CDAEB (4.7,-5.3) .86	CDAEB (5.0,-5.3) .86	CDAEB (4.9,-5.3) .86	CDAEB (4.7,-5.3) .86
EBDCA	DEBAC (-3.0,0.6)	EBDAC (-7.0,0.9) .43	EBDAC (-7.0,0.9) .43	EBDAC (-7.0,0.9) .43	EBDAC (-7.0,0.9) .43
CBAED	CBAED (-0.2,0.9)	CADBE (6.3,0.9) .77	CADBE (6.3,0.8) .77	CADBE (6.4,0.8) .77	CADBE (6.2,0.9) .77

De resultaten van het programma MONIDEAL voor deze niet-passende voorkeuren staan in Tabel 7. Er komen horizonpunten voor. De verschillende aanvangsconfiguraties geven verschillende resultaten. Bij de standaard beginconfiguratie en de beginconfiguratie van kolom (5), is het punt, behorende bij de tweede voorkeur EBDCA, in beide gevallen een ideaalpunt (kolom (2)), en geen horizonpunt. Het verlies van dit punt is ook hoger dan bij de twee horizonpunten. Voor de punten behorende bij de eerste en derde voorkeur zien we in de resultaten dat dit allemaal horizonpunten zijn. Het verlies voor de verschillende horizonpunten is laag.

Tabel 7 Resultaten MONIDEAL voor drie voorkeuren. Hier wordt weergegeven het gebied waarin het ideaalpunt ligt, de coördinaten van het ideaalpunt en het verlies.

voorkeur	beginconfiguratie				
	(1) standaard beginconf	(2) resultaat beginconf	(3) ABCDE (1.0,3.0)	(4) BEADC (-4.0,3.0)	(5) ABDEC (-1.5,0.5)
CEBAD	DABEC (-0.9,0.4)	CDAEB (-19.2,-26.9) .08	ABCDE (17.6,44.0) .07	BACED (7.9,49.7) .06	CDAEB (-19.1,26.7) .08
EBDCA	DEBAC (-3.0,0.6)	EBDAC (-5.1,0.2) .29	BACED (11.9,49.0) .07	EDBAC (-16.0,-5.7) .08	EBDAC (-5.1,0.2) .29
CBAED	CBAED (-0.2,0.9)	CDAEB (19.5,-26.7) .08	ABCDE (11.4,18.9) .05	ABCDE (16.5,35.2) .03	CDAEB (19.2,26.4) .08

6.1.2 MONIDEAL en SMIDEAL

Bij deze vergelijking wordt in de programma's de standaard beginconfiguratie gebruikt. Twee punten zijn hier essentieel. Het eerste punt is of horizonpunten worden vermeden. Het tweede punt is of SMIDEAL met betrekking tot de ideaalpunten dezelfde resultaten geeft als MONIDEAL.

Als Tabel 3 en onderstaande Tabel 8 vergeleken worden is het volgende te zien. In de eerst genoemde tabel is te zien dat de ideaalpunten, behorende bij de drie eerste voorkeuren, perfect zijn weergegeven en dus een verlies van 0 hebben. De resultaten in Tabel 8 zien er minder goed uit voor deze ideaalpunten: alleen het derde ideaalpunt is perfect weergegeven. Het ideaalpunt behorende bij de vierde voorkeur is daarentegen beter terecht gekomen omdat de eerste twee stimuluspunten in de juiste volgorde van voorkeur staan.

Tabel 8 Resultaat SMIDEAL voor vier voorkeuren. Hier wordt weergegeven het gebied waarin het ideaalpunt ligt (kolom 2), de coördinaten van het ideaalpunt (kolom 3) en het verlies (kolom 4).

voorkeur	resultaat standaard beginconfiguratie		
BAEDC	BADEC	(-2.6,1.9)	.02
BEADC	BEDAC	(-2.9,1.7)	.01
BEDAC	BEDAC	(-3.2,1.7)	.00
EDCBA	EDBAC	(-4.5,-1.2)	.37

De resultaten in Tabel 5 en 9 geven eenzelfde beeld. In Tabel 5 is te zien dat drie ideaalpunten perfect zijn weergegeven, terwijl na toepassing van SMIDEAL is er maar een ideaalpunt perfect weergegeven. De coördinaten van dit punt zijn verschillend in de beide programma's: bij MONIDEAL ligt het punt bijna op de lijn AB, terwijl bij SMIDEAL deze in het gebied liggen en niet vlak langs de lijn. Het derde ideaalpunt is in beide gevallen niet perfect weergegeven: het verlies bij SMIDEAL (0.27) en bij MONIDEAL (0.03).

Tabel 9 Resultaat SMIDEAL voor vier voorkeuren. Hier wordt weergegeven het gebied waarin het ideaalpunt ligt (kolom 2), de coördinaten van het ideaalpunt (kolom 3) en het verlies (kolom 4).

voorkeur	resultaat standaard beginconfiguratie		
DABCE	DAEBC	(-1.3,-0.6)	.09
ABDEC	ABDEC	(-1.0,1.9)	.00
DEBCA	DEBAC	(-2.6,-0.7)	.27
ADBEC	ABDEC	(-0.8,1.5)	.02

Tabel 10 Resultaat SMIDEAL voor vier voorkeuren. Hier wordt weergegeven het gebied waarin het ideaalpunt ligt (kolom 2), de coördinaten van het ideaalpunt (kolom 3) en het verlies (kolom 4).

voorkeur	resultaat standaard beginconfiguratie		
CEBAD	CDABE	(-2.9,-2.4)	.43
EBDCA	EDBAC	(-4.4,0.4)	.32
CBAED	CADBE	(2.5,-1.5)	.49

De resultaten in Tabel 10 zien er veelbelovend uit als deze worden vergeleken met die in Tabel 7. De twee horizonpunten zijn hier verdwenen na toepassing van SMIDEAL. Het tweede ideaalpunt ligt na MONIDEAL toegepast te hebben in gebied EBDAC, met een verlies van 0.29. In Tabel 10 is te zien dat dit tweede ideaalpunt in een ander gebied ligt en een hoger verlies heeft (0.32).

6.2 Tweede voorbeeld

In dit tweede voorbeeld wordt data gebruikt uit een onderzoek van Vonk (1988). Hier wordt onderzoek gedaan naar impliciete persoonlijkheidstheorieën. De data zijn preferentiedata van 23 proefpersonen. Van elk van deze proefpersonen is nagegaan welke persoonlijkheidstrekken hij/zij belangrijk vindt. Dus van alle 23 proefpersonen is, door middel van MDS, een individuele representatie van de impliciete persoonlijkheids theorie (IPT) verkregen. Met andere woorden, er zijn 23 verschillende targetconfiguraties. Vervolgens worden de 23 proefpersonen gevraagd elk 6 personen (3 onbekenden en 3 bekenden) op deze persoonlijkheidstrekken te beoordelen op een schaal van 1 t/m 6, waarbij de waarde 1 aangeeft dat die persoonlijkheidstrek zeker van toepassing is op de beoordeelde en de waarde 6 als dit zeker niet het geval is. De proefpersonen zijn toegestaan gelijke beoordelingen te geven; er komen in deze data 'ties' voor. Elk van deze zes personen moet als punt worden afgebeeld in de bijbehorende targetconfiguratie. Dus per proefpersoon moeten er zes ideaalpunten worden geplaatst in ieder van de 23 targetconfiguraties. Deze zes ideaalpunten zijn onafhankelijk van elkaar. Kortom, de data staan in 23 matrices met zes rijen en een verschillend aantal kolommen (niet ieder proefpersoon heeft een zelfde aantal persoonlijkheidstrekken).

In dit onderzoek wordt de data van een aantal proefpersonen onderzocht, namelijk de proefpersonen 1 t/m 6. Dit is een willekeurige keuze uit de hoeveelheid data. Op deze data worden de programma's ADDIDEAL en MONIDEAL toegepast. We zijn er met name in geïnteresseerd of in dit voorbeeld horizonpunten optreden. In deze data komen 'ties' (gelijke dissimilarities) voor, vandaar dat in geval van MONIDEAL beide benaderingen (secondary en primary approach) worden toegepast. Van iedere proefpersoon wordt in de Tabellen 11 tot en met 13 (Appendix B) de coördinaten van het ideaalpunt en het verlies per ideaalpunt weergegeven. In Tabel 11 staan de resultaten van ADDIDEAL en in Tabel 12 van MONIDEAL met de optie primary approach, voortaan PMONIDEAL genoemd. Vervolgens staan in Tabel 13 de resultaten van MONIDEAL met de optie secondary approach, voortaan SMONIDEAL genoemd.

Als we de resultaten van ADDIDEAL in Tabel 11 bekijken, zien we dat er horizonpunten voorkomen bij de proefpersonen 1 en 3. Bij proefpersoon 1 is het tweede ideaalpunt een horizonpunt. Het verlies van het horizonpunt (1.1) is niet opvallend hoog of laag, in tegenstelling tot proefpersoon 3. Bij de laatste is het vijfde ideaalpunt een horizonpunt en dit heeft een hoog verlies (2.1). In het eerste voorbeeld kwamen geen horizonpunten voor, wat kan liggen aan het voorbeeld omdat er geen ties in zaten.

Bij toepassing van PMONIDEAL komen er horizonpunten voor bij proefpersonen 3 en 4. Het vijfde ideaalpunt is een horizonpunt bij de eerstgenoemde. Het verlies bij dit horizonpunt is laag (0.21). Proefpersoon 4 heeft twee horizonpunten, het derde en het vijfde punt. Ook hier is het bijbehorende verlies laag (respectievelijk 0.15 en 0.18).

De meeste horizonpunten komen voor bij toepassing van SMONIDEAL. Proefpersoon 1 heeft het tweede ideaalpunt een horizonpunt met een verlies van 0.16. De derde proefpersoon heeft vier horizonpunten, de eerste twee en de laatste twee. Alleen de middelste twee zijn ideaalpunten. Alle vier horizonpunten hebben een laag verlies (0.16, 0.17, 0.14, 0.20 respectievelijk). De vierde proefpersoon heeft twee horizonpunten, namelijk de vierde en de vijfde. Ook hier is er sprake van een laag verlies (respectievelijk 0.15 en 0.14).

7. Bespreking van de resultaten

In de vorige sectie zijn de drie programma's toegepast om een viertal punten te onderzoeken, namelijk of de ideaalpunten in de bijbehorende isotone gebieden terecht komen (1), of er horizonpunten optreden en in welke situaties (2), of ADDIDEAL en MONIDEAL last hebben van locale minima (3) en in hoeverre de resultaten van MONIDEAL en SMIDEAL van elkaar verschillen (4). Allereerst worden hier de resultaten besproken van het eerste voorbeeld.

Als met de standaard beginconfiguratie wordt begonnen komen de ideaalpunten niet altijd in het bijbehorende gebied. MONIDEAL geeft wat dat betreft betere resultaten; er komen perfecte ideaalpunten voor, terwijl dat in ADDIDEAL niet het geval is. De eisen in MONIDEAL zijn ook lager: er hoeft alleen rekening te worden gehouden met de rangordening. In Tabel 2 is daarom het verlies behorende bij ideaalpunt 2 niet 0 omdat alleen aan de rangorde eis is voldaan. Het is moeilijker om een punt te vinden die aan een ratio meetniveau voldoet.

De aard van de horizonpunten, bij gebruik van de standaard beginconfiguratie, is niet van belang. Bij toepassing van de beide programma's ADDIDEAL en MONIDEAL komen alleen horizonpunten voor als er sprake is van een niet-passende voorkeur en in het eerste voorbeeld alleen na toepassing van MONIDEAL. In dit geval bestaat er geen ideaalpunt in de targetconfiguratie die aan de eisen voldoet. De horizonpunten gaan gepaard met een zeer laag verlies. Als een punt in de configuratie niet bestaat en toch moet het ideaalpunt worden ingepast, dan probeert het algoritme een zodanig ideaalpunt te vinden dat de pseudo-afstanden tot alle stimuluspunten gelijk wordt gemaakt. In dit geval betekent het dat de gewichten niet meer van elkaar verschillen. Alleen het gewicht behorende bij de kleinste dissimilariteit differentieert nog wel het verlies in WORK; de afstand van een horizonpunt tot het stimuluspunt van grootste voorkeur is de kleinste en het bijbehorende gewicht verschilt van de andere gewichten die aan elkaar gelijk zijn. In dit geval gaat het om een gedegeneerde oplossing, maar omdat het de kleinste dissimilarity nog wel wordt onderscheiden is het verlies niet gelijk aan 0.

De beginconfiguratie is in het eerste voorbeeld gevariëerd. Uit de resultaten hiervan kunnen twee zaken worden geconcludeerd. Ten eerste blijkt een willekeurig gekozen beginconfiguratie in het geval van ADDIDEAL niet veel uit te maken: de ideaalpunten komen, op een uitzondering na, telkens in hetzelfde gebied terecht en het verlies is, op een uitzondering na, hetzelfde. Voor het programma MONIDEAL zijn de resultaten anders. Als de beginconfiguratie verandert, veranderen de resultaten. Ten tweede is het belangrijk voor het programma MONIDEAL een juiste beginconfiguratie te creëren om onnodige horizonpunten te voorkomen. De standaard beginconfiguratie is hiervoor een goed uitgangspunt.

Voor wat betreft het vierde punt is gebleken dat de horizonpunten door 'smooth' monotone regressie te gebruiken, kunnen worden voorkomen. De verkregen resultaten met betrekking tot de ideaalpunten verschillen echter met de ideaalpunten na toepassing van MONIDEAL. Het verlies bij de resultaten van SMIDEAL is slechter. Dit komt omdat er strengere eisen zijn met betrekking tot de afstanden. Er kunnen geen gelijke pseudo afstanden tot de stimuluspunten voorkomen omdat geeïst wordt dat er verschil is bij verschillende dissimilarities. In MONIDEAL wordt dit niet geeïst, want een ideaalpunt kan op een middelloodlijn terechtkomen, zoals we eerder in hoofdstuk 6 zijn tegen gekomen; het algoritme heeft dus meer vrijheid dan SMIDEAL.

Voor wat betreft het tweede voorbeeld blijken er horizonpunten voor te komen, zowel bij ADDIDEAL als bij MONIDEAL. De mogelijke oorzaak hiervan is dat in het eerste voorbeeld geen gelijke dissimilarities voorkomen. De horizonpunten komen het meeste voor na toepassing van SMONIDEAL. Het verschil met PMONIDEAL is dat gelijke dissimilarities gelijke pseudo-afstanden moeten geven. Er is minder vrijheid om het punt ergens neer te leggen. Het resultaat is dat meerdere horizonpunten ontstaan, omdat het lastiger is een ideaalpunt te vinden die gelijk afstand heeft tot een aantal stimuluspunten. Om dit ideaalpunt te vinden zoekt het algoritme een punt waar de afstanden tot de stimuluspunten gelijk aan elkaar zijn, zowel binnen de gelijke dissimilarities als tussen de verschillende dissimilarities.

8. Conclusie

In de inleiding zijn twee punten genoemd om horizonpunten te onderzoeken, namelijk waar komen horizonpunten vandaan en het tweede punt is een oplossing proberen te vinden om horizonpunten te vermijden. In dit laatste hoofdstuk worden deze twee punten besproken.

Aan de hand van hoofdstuk 2 is duidelijk geworden dat horizonpunten twee oorzaken kunnen hebben. Een eerste oorzaak is dat er geen punt bestaat in de configuratie waarvan de grootte van de afstanden tot de stimuluspunten overeenkomen met een bepaalde voorkeur. De tweede oorzaak is dat het een punt betreft dat ligt in een isotoon gebied dat ver van het convexe omhulsel aflight. Naar aanleiding van de resultaten in hoofdstuk 6 kan geconcludeerd worden dat de horizonpunten voorkomen in het geval van een niet-passende voorkeur. Het omgekeerde echter geldt niet: als er een niet-passende voorkeur is

dan worden ideaalpunten niet automatisch horizonpunten. Als in de data veel gelijke dissimilarities voorkomen treden er ook horizonpunten op.

Voor het tweede onderzoekspunt, een oplossing proberen te vinden voor het vermijden van horizonpunten, is de verliesfunctie WORK en de 'smooth' monotone regressie geïntroduceerd. Het doel van de verliesfunctie WORK is om differentiatie aan te brengen in het wegen van het verlies, maar dit blijkt niet te zijn opgetreden. Dit hangt samen met de oorzaak van de horizonpunten waarbij de pseudo-afstanden, en dus de gewichten, tot de stimuluspunten gelijk zijn. Hierdoor werkt de verliesfunctie niet, waarin juist de nadruk gelegd wordt op de verschillende waarden die de gewichten kunnen aannemen.

Een tweede methode, voor ordinaal meetniveau, is de beperkte 'smooth' monotone regressie. Door middel van deze methode is een verbetering opgetreden voor wat betreft de horizonpunten; deze worden vermeden. Zoals te verwachten is, zijn de resultaten van SMIDEAL met betrekking tot de ideaalpunten slechter in vergelijking met de ideaalpunten verkregen door MONIDEAL. In het kader van deze studie is de conclusie dat de beperkte 'smooth' monotone regressie een goede methode is om horizonpunten te voorkomen.

Als laatste twee opmerkingen. De eerste opmerking heeft betrekking op het programma SMIDEAL. Het is de moeite waard om SMIDEAL ook geschikt te maken voor data waarin gelijke beoordelingen in voorkomen, zodat het programma volledig kan worden uitgetest. De laatste opmerking heeft betrekking op de vraag of de 'smooth' monotone regressie benadering ook zonder WORK, in dit externe ontvouwingsmodel, horizonpunten kan voorkomen.

APPENDIX A

Het 'smooth' monotone regressie probleem is beschreven door Heiser (1985). Hier wordt zijn methode gevolgd om een algoritme af te leiden.

Het 'smooth' monotone regressie probleem kan worden geformuleerd als het vinden van een toegestane pseudo-afstand, zodanig dat de volgende functie wordt geminimaliseerd:

$$\tau(\gamma) = \|\gamma - d\|_W^2 = \text{tr}(\gamma - d)'W(\gamma - d) \quad (1)$$

W is een diagonale $m \times m$ matrix die de niet-negatieve gewichten bevat. De vergelijking (1) kan worden herschreven in de volgende algemene kwadratische vorm

$$\max_{D\gamma \geq 0} \quad 2\gamma'Wd - \gamma'W\gamma \quad (2)$$

waarbij D een $(3m-1) \times m$ coëfficiënten matrix is waarbij de eerste m rijen worden gedefinieerd door de monotoniceits eisen en de rest van de rijen wordt gedefinieerd door de 'smoothness' condities. De Kuhn-Tucker condities voor probleem (2) worden afgeleid door de Lagrangean op te zetten. Dit impliceert dat een k -vector met niet-negatieve waarden λ moet bestaan zodat de oplossing λd^* kan worden geschreven als

$$d^* = d + W^{-1}D'\lambda \quad (3)$$

Als E en f als volgt worden gedefinieerd

$$E = W^{-1/2}D'$$
$$f = -W^{1/2}d$$

dan wordt het duale probleem, het vinden van de juiste λ , als volgt

$$\min_{\lambda \geq 0} \quad \|E\lambda - f\|^2 \quad (4)$$

Het probleem (3) kan worden herleid tot een kleinste kwadraten probleem (4). Door de laatste op te lossen kan het antwoord op (2) worden verkregen door (3).

Vergelijking (4) herschreven krijgt men

$$\min_{\lambda \geq 0} \quad \|W^{1/2}d + W^{-1/2}D'\lambda\|^2 \quad (4a)$$

Dit is hetzelfde als

$$\min_{\lambda \geq 0} \quad \| (\mathbf{W}^{1/2} \mathbf{d} + \mathbf{W}^{-1/2} \sum_{j \neq i} \lambda_j \mathbf{d}_j) + \mathbf{W}^{-1/2} \lambda_i \mathbf{d}_i \|^2 \quad (4b)$$

Waarbij \mathbf{d}_i en \mathbf{d}_j de kolommen zijn van matrix \mathbf{D}' .

Stel $\mathbf{W}^{1/2} \mathbf{d}^* = \mathbf{W}^{1/2} \mathbf{d} + \mathbf{W}^{-1/2} \sum \lambda_j \mathbf{d}_j$ dan ziet vergelijking (4b) er als volgt uit:

$$\min_{\lambda \geq 0} \quad \| \mathbf{W}^{1/2} \mathbf{d}^* + \mathbf{W}^{-1/2} \lambda_i \mathbf{d}_i \|^2 \quad (4c)$$

Als (4c) uitgewerkt wordt dan wordt de volgende vergelijking verkregen:

$$\min_{\lambda \geq 0} \quad \mathbf{d}^{*T} \mathbf{W} \mathbf{d}^* + 2 \mathbf{d}^{*T} \lambda_i \mathbf{d}_i + \lambda_i \mathbf{d}_i^T \mathbf{W}^{-1} \mathbf{d}_i \lambda_i \quad (5)$$

Dit probleem wordt opgelost door

$$\lambda_i = -\mathbf{d}^{*T} \mathbf{d}_i / \mathbf{d}_i^T \mathbf{W}^{-1} \mathbf{d}_i \quad (6)$$

Dit probleem wordt iteratief opgelost, waarbij in elke iteratie een λ_i wordt berekend. De volgende notatie wordt gehandhaafd: λ^{new} is de λ die wordt berekend in een bepaalde iteratie en λ^{old} is de in de vorige iteratie berekende λ .

Het algoritme bestaat uit de volgende stappen:

1. λ^{old} wordt gelijkgesteld aan 0 en \mathbf{d}^* aan δ .
2. In deze stap worden de opeenvolgende elementen van λ_i^{new} , van λ^{new} , berekend met de kolommen \mathbf{d}_i van \mathbf{D}' . Allereerst wordt gecontroleerd of $\lambda_i^{old} = 0$. Zo ja, dan wordt overgegaan naar stap 2A, anders naar stap 2B.

2A. als $\lambda_i^{old} = 0$, bereken λ_i^{new} door middel van (6).

Is $\lambda_i^{new} < 0$ dan wordt $\lambda_i^{new} = 0$

na deze stap is gedaan volgt stap 3.

2B. als $\lambda_i^{old} > 0$, dan definiëren we $\lambda_i^{new} = \lambda_i^{old} + \lambda_i^*$, waaruit volgt als geldt

$\lambda_i^{new} \geq 0$, dat $\lambda_i^* \geq -\lambda_i^{old}$, anders wordt $\lambda_i^* = -\lambda_i^{old}$

Het resultaat van λ_i^{old} eerst teniet worden gedaan. Dit gaat als volgt:

$\mathbf{d}^* = \mathbf{d} - \mathbf{W}^{-1} \mathbf{d}_i \lambda_i^{old}$.

In deze stap wordt λ_i^* berekend in plaats van λ_i^{new} , door middel van (6).

Als $\lambda_i^* < -\lambda_i^{old}$ dan wordt $\lambda_i^* = -\lambda_i^{old}$.

Ga naar stap 3.

3. In deze stap wordt de verandering aangebracht. De 'update' vindt als volgt plaats:

$$d^* = d + W^{-1}d_i \lambda_i^{\text{new}}.$$

In geval van λ_i^* , wordt de 'update' natuurlijk $d^* = d + W^{-1}d_i \lambda_i^*$

4. Zet $\lambda_i^{\text{old}} = \lambda_i^{\text{new}}$ en, in geval van λ_i^* , is dit $\lambda_i^{\text{old}} = \lambda_i^*$
5. Ga terug naar stap 2 net zolang totdat $\|d^*\|_W^2$ is een gekozen convergentie criterium heeft bereikt.

APPENDIX B

Appendix B bevat drie tabellen, 11 tot en met 13, waarin de resultaten van de programma's ADDIDEAL, MONIDEAL en SMIDEAL staan.

Tabel 11 Resultaten ADDIDEAL : coördinaten ideaalpunten en bijbehorend verlies van zes proefpersonen.

Proefpersoon 1

<i>coördinaten</i>		<i>verlies</i>
-0.09656	0.02803	0.91323
-1.52618	-1.08632	1.13479
-0.10644	0.01349	1.04383
0.05448	0.10803	1.66669
-0.09357	-0.04445	1.46866
-0.05235	-0.04892	1.97315

Proefpersoon 2

<i>coördinaten</i>		<i>verlies</i>
-0.11025	-0.05266	1.70436
0.06487	0.09357	2.28224
0.04014	0.14023	1.09741
-0.00856	0.12296	1.72366
0.11588	-0.05149	1.26510
0.10418	-0.07440	1.46130

Proefpersoon 3

<i>coördinaten</i>		<i>verlies</i>
-0.24779	-0.15199	1.65390
-0.03263	0.06589	1.63556
0.03066	-0.01211	1.87997
-0.10219	-0.09740	1.67256
-0.59720	-1.14565	2.10048
-0.06560	0.10126	1.46459

Proefpersoon 4

<i>coördinaten</i>		<i>verlies</i>
-0.00039	0.04950	1.46283
0.10256	0.07927	1.00378
0.59638	0.25204	1.36364
-0.04737	-0.08253	1.40211
0.05829	-0.19896	1.55134
0.07975	0.01768	1.85091

Proefpersoon 5

<i>coördinaten</i>		<i>verlies</i>
0.07277	0.13529	1.66324
-0.10190	0.02227	1.16793
0.04819	0.39751	1.79630
-0.05650	-0.07670	1.34184
0.01154	-0.03737	1.38885
-0.05116	0.04933	1.26189

Proefpersoon 6

<i>coördinaten</i>		<i>verlies</i>
-0.08591	0.01929	1.05354
-0.13307	0.05196	1.22031
-0.10172	0.01261	1.42170
-0.00549	-0.02543	1.52754
-0.04766	0.02858	1.31802
0.08941	-0.06911	1.67312

Tabel 12 Resultaten MONIDEAL : coördinaten ideaalpunten en bijbehorend verlies van zes proefpersonen.

Proefpersoon 1

<i>coördinaten</i>		<i>verlies</i>
-0.08023	0.06245	0.24416
-0.00878	0.11473	1.00474
-0.12429	0.02415	0.40590
0.00170	0.09329	0.69299
-0.04461	-0.08417	0.14638
0.09102	0.03173	0.85389

Proefpersoon 2

<i>coördinaten</i>		<i>verlies</i>
-0.13391	-0.05254	0.14880
-0.07044	0.05846	1.08983
0.12113	0.18807	0.46636
-0.03299	0.13481	0.82785
0.11288	-0.03437	0.38398
0.09048	-0.09875	0.63551

Proefpersoon 3

<i>coördinaten</i>		<i>verlies</i>
-0.05616	0.12705	0.86139
0.00779	0.00981	0.74832
0.01963	-0.01390	0.66876
-0.04481	0.00465	0.64778
-0.30946	2.36551	0.20510
-0.05442	0.13671	0.62390

Proefpersoon 4

<i>coördinaten</i>		<i>verlies</i>
-0.07894	0.05701	0.58375
0.09816	0.09076	0.49972
0.71390	1.71475	0.14582
-0.06927	0.03336	0.91511
0.21727	-1.96154	0.18544
0.15322	0.08744	0.99566

Proefpersoon 5

<i>coördinaten</i>		<i>verlies</i>
-0.03637	0.03392	0.89598
-0.10727	-0.03232	0.39401
-0.04594	0.15625	0.77903
-0.10927	-0.11819	0.32965
-0.08558	-0.04558	0.63381
-0.07793	0.04572	0.31018

Proefpersoon 6

<i>coördinaten</i>		<i>verlies</i>
-0.12513	-0.07187	0.31145
-0.37844	-0.15539	0.33558
-0.04248	-0.00070	0.58274
-0.01621	-0.00975	0.89312
-0.04327	0.11972	0.65783

Tabel 13 Resultaten SMIDEAL : coördinaten ideaalpunten en bijbehorend verlies van zes proefpersonen.

Proefpersoon 1

<i>coördinaten</i>		<i>verlies</i>
-0.09726	0.02337	0.87486
-3.01560	2.85233	0.15640
-0.12274	0.01044	0.90668
0.04624	0.12474	1.22758
-0.09806	-0.03150	1.12108
-0.07875	-0.03391	1.65503

Proefpersoon 2

<i>coördinaten</i>		<i>verlies</i>
-0.11617	-0.02628	0.53543
-0.07905	0.05034	1.09147
-0.00784	0.14612	0.97058
-0.07125	0.05999	1.07795
0.11486	-0.05203	1.16448
0.07327	-0.06939	1.30412

Proefpersoon 3

<i>coördinaten</i>		<i>verlies</i>
-1.22346	2.53911	0.16257
-0.39294	2.87735	0.17456
0.03661	-0.01233	1.19698
-0.00649	-0.03137	1.21716
0.28513	-3.42763	0.14356
-0.68668	2.37131	0.19761

Proefpersoon 4

coördinaten

0.09328 0.08790
0.11115 0.07676
0.16120 0.09823
0.31468 -3.19599
0.54682 -3.04455
0.14154 0.08347

verlies

1.19976
0.96412
0.99592
0.15425
0.14402
1.01529

Proefpersoon 5

coördinaten

-0.03260 0.02333
-0.07955 0.03388
-0.06450 0.05109
-0.08240 -0.02779
-0.00267 -0.02152
-0.09428 0.03931

verlies

0.92303
0.87514
0.97188
0.64197
0.93461
0.63174

Proefpersoon 6

coördinaten

-0.74837 5.15265
-0.70301 5.46513
-0.03328 0.96368
0.15973 -0.94600
-0.80856 5.33434
-0.77231 5.08789

verlies

0.10765
0.10167
0.58039
0.78481
0.10279
0.10895

APPENDIX C

De APL functies ADDIDEAL, ROWSCALE en UNDERFIT (Heiser, 1988).

```
▽X<DELTA ADDIDEAL TARG;A;B;W;P;R;EPS;DHAT;CRIT;DIST;SIGOLD;SIGNEW;ITER
[11] EPS<1E-10
[12] DHAT<DELTA
[13] CRIT<1E-6
[14] PR<(+/W<(1/1/DELTA)-DELTA)∘.x(ρTARG)[2]ρ1
[15] PX<(W+ .xTARG)÷R+R≤EPS
[16] 'UUL BEGINCONF IN'
[17] X<0
[18] Q<10 7▽SIGNEW<X UNDERFIT TARG
[19] ITER<0
[10] CYCLE:SIGOLD<SIGNEW
[11] DHAT<DELTA ROWSCALE DIST
[12] W<1÷DHAT*2
[13] A<Bx(WxDHAT)÷DIST+1-B<DIST>EPS
[14] P<(+/A)∘.x(ρTARG)[2]ρ1
[15] R<(+/W)∘.x(ρTARG)[2]ρ1
[16] X<((W+ .xTARG)+(PxX)-A+ .xTARG)÷R+R≤EPS
[17] Q<20 7▽SIGNEW<X UNDERFIT TARG
[18] →READYx1CRIT>|SIGOLD-SIGNEW
[19] →CYCLEx1500>ITER<ITER+1
[20] Q<'MAXIMUM NUMBER OF CYCLES EXCEEDED'
[21] READY:Q<'READY; NUMBER OF CYCLES IS:'
[22] Q<ITER
[23] Q<'THE MEAN LOSS IS:'
[24] Q<10 7▽SIGNEW
[25] Q<'THE IDEAL POINTS AND INDIVIDUAL LOSSES ARE:'
[26] Q<10 5▽X<X,[2](+/(1-DIST÷DHAT)*2)*0.5
[27] Q<'THE STANDARD DEVIATION IS:'
[28] Q<10 7▽SD<((+/X[; (ρX)[2]]-SIGNEW)*2)÷(ρX)[1]-1)*0.5
[29] Q<'THE COEFFICIENT OF VARIATION IS:'
[30] Q<10 7▽SD÷SIGNEW+1-SIGNEW>EPSxEPS
▽

▽ROWSCALE[Q]▽
▽DHAT<DELTA ROWSCALE DIST;ALF
[1] ALF<(+(DIST÷DELTA)*2)÷+/DIST÷DELTA
[2] DHAT<DELTAxq((ρDIST)[2],(ρDIST)[1])ρALF
▽
```

```

      ▽STRESS←X UNDERFIT TARG;NROW;NCO;ALF;BET
[1]  NROW←(ρX)[1]
[2]  NCOL←(ρTARG)[1]
[3]  ALF←(NCOL,NROW)ρ+X*2
[4]  BET←(NROW,NCOL)ρ+TARG*2
[5]  DIST←(ALF+BET-2X+.xTARG)*0.5
[6]  STRESS←(+/(+/(1-DIST÷DHAT)*2)*0.5)÷NROW

```

De functies PMONIDEAL en SMONIDEAL verschillen van ADDIDEAL door regel 11. Hierin wordt respectievelijk PMONOREGR (primary approach) en SMONOREGR (secondary approach) opgeroepen.

```

      ▽PMONOREGR[[]]▽
      ▽DHAT←DELTA PMONOREGR DIST;DD;J;WW;HDELTA;HDIST
[1]  DHAT←DELTA
[2]  WW←DIST*2
[3]  J←0
[4]  R1:J←J+1
[5]  HDIST←1÷DIST
[6]  HDELTA←1÷DELTA
[7]  DD←(COL HDIST[J;],COL HDELTA[J;]
[8]  Z←WW[J;]WMOR DD
[9]  DHAT[J;]←1÷Z
[10] →(J>(ρDELTA)[1])÷READY1
[11] →R1
[12] READY1:→0
      ▽
      ▽SMONOREGR[[]]▽
      ▽DHAT←DELTA SMONOREGR DIST;DD;J;WW
[1]  DHAT←DIST
[2]  WW←DIST*2
[3]  J←0
[4]  R1:J←J+1
[5]  HDIST←1÷DIST
[6]  HDELTA←1÷DELTA
[7]  DD←(COL HDIST[J;],COL HDELTA[J;]
[8]  Z←WW[J;]WSOR DD
[9]  DHAT[J;]←1÷Z
[10] →(J>(ρDELTA)[1])÷READY1
[11] →R1
[12] READY1:→0

```

In de functies PMONOREGR en SMONOREGR worden respectievelijk WMOR en WSOR opgeroepen. Deze twee functies zijn geleend van de vakgroep Data Theorie (Universiteit Leiden)

```

◊WMOR[□]◊
◊Z◊WWW WMOR XY;L;I;IN;N;NORD;IDAT;ISOR;WWW
[1] X◊XY[;1]
[2] Y◊XY[;2]
[3] L◊((1↓I)≠1↓I◊Y[NORD◊Y]),1
[4] Z◊X[NORD]
[5] I◊0
[6] L1:→(◊/L)/L2
[7] Z[IN]◊IDAT[ISOR◊IDAT◊Z[IN◊I+1N◊L1]]
[8] NORD[IN]◊NORD[IN[ISOR]]
[9] I◊I+N
[10] →(0≠ρL◊N↓L)/L1
[11] L2:WWW◊WW[NORD]
[12] Z[NORD]◊WWW MONOR Z

```

```

◊WSOR[□]◊
◊Z◊WWW WSOR XY;L;I;IN;N;IDAT;NORD;ISOR;WWW
[1] X◊XY[;1]
[2] Y◊XY[;2]
[3] L◊((1↓I)≠1↓I◊Y[NORD◊Y]),1
[4] Z◊X[NORD]
[5] I◊0
[6] WWW◊WW[NORD]
[7] L1:→(◊/L)/L2
[8] IDAT◊Z[IN◊I+1N◊L1]
[9] IGEW◊WWW[IN]
[10] Z[IN]◊+/(IDAT◊IGEW)÷+/IGEW
[11] I◊I+N
[12] →(0≠ρL◊N↓L)/L1
[13] L2:Z[NORD]◊WWW MONOR Z

```

De functie SMIDEAL verschilt met ADDIDEAL in regel 13, waarin de functie SMOOTHREG wordt aangeroepen. Tevens kan de gebruiker hier de in hoofdstuk 5 beschreven D matrix invullen. Ook kan een keuze gemaakt worden om de standaard beginconfiguratie te gebruiken of de gebruiker kan zelf een beginconfiguratie aangeven.

```

▽X←DELTA SMIDEAL TARG;A;B;W;P;R;EPS;DHAT;CRIT;DIST;SIGOLD;SIGNEW;ITER;D
[1] EPS←1E-10
[2] 'MATRIX D IS'
[3] D←0
[4] DHAT←DELTA
[5] CRIT←0.001
[6] A←R←(+/W←(1/1/DELTA)-DELTA)•.x(ρTARG)[2]ρ1
[7] A←X←(W+.xTARG)÷R+R≤EPS
[8] 'BEGINCONF IS'
[9] X←0
[10] D←10 7▽SIGNEW←X UNDERFIT TARG
[11] ITER←0
[12] CYCLE:SIGOLD←SIGNEW
[13] DHAT←DELTA SMOOTHREG DIST
[14] W←1÷DHAT*2
[15] A←B←(W×DHAT)÷DIST+1-B←DIST>EPS
[16] P←(+/A)•.x(ρTARG)[2]ρ1
[17] R←(+/W)•.x(ρTARG)[2]ρ1
[18] X←((W+.xTARG)+(P×X)-A+.xTARG)÷R+R≤EPS
[19] D←20 7▽SIGNEW←X UNDERFIT TARG
[20] →READY×1CRIT>|SIGOLD-SIGNEW
[21] →CYCLE×1500?ITER←ITER+1
[22] D←'MAXIMUM NUMBER OF CYCLES EXCEEDED'
[23] READY:D←'READY; NUMBER OF CYCLES IS:'
[24] D←ITER
[25] D←'THE MEAN LOSS IS:'
[26] D←10 7▽SIGNEW
[27] D←'THE IDEAL POINTS AND INDIVIDUAL LOSSES ARE:'
[28] D←10 5▽X←X,[2](+/(1-DIST÷DHAT)*2)*0.5
[29] D←'THE STANDARD DEVIATION IS:'
[30] D←10 7▽SD←(((+/X[; (ρX)[2]]-SIGNEW)*2)÷(ρX)[1]-1)*0.5
[31] D←'THE COEFFICIENT OF VARIATION IS:'
[32] D←10 7▽SD÷SIGNEW+1-SIGNEW>EPS×EPS
▽

```

```

▽DHAT←DELTA SMOOTHREG DIST;W;T;XY;HDELTA;HDIST
[1] DHAT←DELTA
[2] W←DIST*2
[3] T←0
[4] HDIST←1÷DIST
[5] HDELTA←1÷DELTA

```

```

[6] ITER:T←T+1
[7] XY←(COL HDIST[T;]),COL HDELTA[T;]
[8] Z←W[T;]SMOOTH XY
[9] DHAT[T;]←1÷Z
[10] →(T≥(ρDELTA)[1])/READY
[11] →ITER
[12] READY:→0

```

```

    ▽Z←W SMOOTH XY;X;Y;NORD;WW
[1] X←XY[;1]
[2] Y←XY[;2]
[3] NORD←ϕY
[4] Z←X[NORD]
[5] WW←W[NORD]
[6] Z[NORD]←WW SMREG5 Z
    ▽

```

```

    ▽QDAK←W SMREG5 Q;LO;LN;LS;J;IT;S;K;AK;AR;SSQOLD;SSQNEW;LDO;LDN
[1] IT←1
[2] QDAK←Q
[3] SSQOLD←W+.x(QDAK)*2
[4] AR←(ρD)[2]
[5] AK←(ρD)[1]
[6] LDO←6 5ρ0
[7] LDN←LDO
[8] LS←LO←LN←6ρ0
[9] COLUMN:K←0
[10] S←0
[11] STEPS:S←S+1
[12] →(S>(ρD)[1])/CONU
[13] R2:J←1
[14] R3:K←K+1
[15] R1:→(LDO[K;]≠0)/L1
[16] LN[K]←-((bQDAK)+.xD[;K])÷((bD[;K])+.x(⊖DIA W)+.xD[;K])
[17] →(LN[K]>0)/UPDATE
[18] LN[K]←0
[19] →HOLD
[20] L1:QDAK←QDAK-LDO[K;]
[21] LS[K]←-((bQDAK)+.xD[;K])÷((bD[;K])+.x(⊖DIA W)+.xD[;K])
[22] →(LS[K]≥-LO[K])/CHANGE
[23] LS[K]←-LO[K]
[24] CHANGE:LN[K]←LS[K]
[25] UPDATE:LDN[K;]←LN[K]×((⊖DIA W)+.xD[;K])

```



```

[26] QDAK←QDAK+LDN[K;]
[27] →HOLD
[28] HOLD:LDO[K;]←LDN[K;]
[29] J←J+1
[30] →(AR=AK)/CONTR
[31] →(S=1)/EXTRA
[32] →(J>1)/STEPS
[33] →R3
[34] EXTRA:→(J>2)/STEPS
[35] →R3
[36] CONTR:→(J>(AK÷AR))/STEPS
[37] →R3
[38] CONU:SSQNEW←W+.x(QDAK)*2
[39] →((|SSQNEW-SSQOLD)≤0.001)/READY
[40] A QDAK
[41] IT←IT+1
[42] →(IT=300)/READY
[43] A |SSQNEW-SSQOLD
[44] SSQOLD←SSQNEW
[45] →COLUMN
[46] READY:'SMREG IS READY      ',IT

```

Literatuurlijst.

- Borg, I.(1981). *Anwendungsorientierte Multidimensionale Skalierung*. Heidelberg: Springer-Verlag.
- Heiser, W.J. & de Leeuw, J. (1979). *SMACOF-1*. Leiden: Department of Data Theory.
- Heiser, W.J.(1981). *Unfolding Analysis of Proximity Data*. (Ph. D. Thesis). Leiden: Department of Data Theory.
- Heiser, W.J.(1985). *Multidimensional scaling by optimizing goodness-of-fit to a smooth hypothesis*. (Internal Report RR-85-07). Leiden: Departement of Data Theory.
- Heiser, W.J.(1987). Joint ordination of species and sites: the unfolding technique. In: P. & L. Legendre (Ed.), *Developments in Numerical ecology* (pp 189-221). Heidelberg: Springer-Verlag.
- Heiser, W.J.(1988). Multidimensional scaling with least absolute residuals. In: H.H. Bock (Ed.), *Classification and Related Methods of Data Analyses* (pp. 455-462). North Holland: Elsevier Science Publishers B.V.
- Young, F. W. (1981). Quantitative analysis of Qualitative data. *Psychometrika*, **46**, 357-387.
- Kruskal, J.B.(1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, **29**, 1-27.
- de Leeuw, J., Young, F. W. & Takane Y. (1976). Additive structure in qualitative data: An alternating least squares method with optimal scaling features. *Psychometrika*, **41**, 471-503.
- de Leeuw, J. (1986). *Persoonlijke mededeling*.
- McGee, V.E.(1966). The multidimensional scaling of "elastic" distances. *The British Journal of Mathematical and Statistical Psychology*, **19**, 181-196.
- Meerling (1981). *Methoden en technieken van psychologisch onderzoek (deel II)*. Meppel: Boom.
- Vonk, R. & Heiser, W. J. (1988). *Implicit personality theory and social judgment: effects of familiarity and familiarization with a target person*. Unpublished manuscript.