

NN31545.0138

INSTITUUT VOOR CULTUURTECHNIEK EN WATERHUISHOUDING

NOTA No.138 d.d.20 juli 1962

Een meetkundige beschouwing over de niet-lineaire vereffening

Ph.Th.Stol

BIBLIOTHEEK DE HAAG

	Droevendaalsesteeg 3a 6708 PB Wageningen	pag.
1. Inleiding		1
2. Meetkundige voorstelling		1
3. Vectorvoorstelling van een functie met één parameter		2
4. Vectorvoorstelling van een functie met meer parameters		5
5. Benadering van de vector Δy		7
6. Nadere beschouwing van de linearisering		10
7. De invloed van de schaalfactoren		12
8. De berekening van de booglengte langs een parameterkromme		15
9. Het minimaliseren van de kwadraatsom		16
10. Berekening en afbeelding van de kwadraatsom		19
11. Het gedrag van de parameters		19
12. Enkele bijzondere gevallen		20
Betrekkingen tussen parameters		20
Parameters met grenswaarden		21
Schaalfactoren van de afhankelijke variabele		21
13. Slotopmerkingen		22
14. Literatuurlijst		23

1787062



1. Inleiding

Bij het aanpassen van een serie waarnemingsuitkomsten aan een gegeven functie wordt gewoonlijk gebruikgemaakt van de methode van de kleinste kwadraten. Dit wil zeggen dat de waarden van de parameters in de functie zo gekozen moeten worden dat de som van kwadraten van het verschil tussen de waarnemingsuitkomsten en de berekende functiewaarden zo klein mogelijk is. De oplossing van dit probleem wordt verkregen uit de oplossing van een stelsel normaalvergelijkingen en is exact indien de parameters lineair in de functie voorkomen. De variabelen die de meetuitkomsten symboliseren (x, y, z, \dots) mogen in een niet-lineaire vorm voorkomen wat door Kamil in nota 134 nader is uiteengezet.

Voor het geval de parameters in een niet-lineaire vorm in de functie opgenomen zijn wordt met de gebruikelijke oplossing met normaalvergelijkingen slechts een benadering van de werkelijke oplossing gevonden. Door een iteratief rekenproces te volgen kan de eindoplossing verkregen worden.

Teneinde uiteen te zetten op welke moeilijkheden men bij niet-lineaire vereffening stuit en in welke richting het overwinnen daarvan gezocht moet worden zal in deze nota de niet-lineaire vereffening aan een meetkundige beschouwing worden onderworpen. In nota 139 zal nader op het probleem van de convergentie naar de eindoplossing worden ingegaan.

2. Meetkundige voorstelling

Stel gegeven een functie van x waarin één parameter voorkomt. Algemeen geldt dan

$$y = f(x;a) \quad (1)$$

Afhankelijk van de waarde die a aanneemt zal er een bepaalde relatie tussen y en x bestaan. Zijn van y en x waarnemingsuitkomsten beschikbaar dan kunnen deze op de gebruikelijke wijze als een stippenzwerm worden weergegeven. In plaats van het tweedimensionale verband tussen x en y kan (1) als een driedimensionale betrekking tussen x, y en a opgevat worden, voorgesteld door

$$F(x, y; a) = 0 \quad (2)$$

FIGUUR 1

De doorsnijdingen van $F = 0$ met vlakken evenwijdig aan het x, y -vlak vormen curven die de niveaulijnen voor verschillende waarden van a voorstellen. De gebruikelijke voorstelling van een dergelijke bundel niveaulijnen in het x, y -vlak volgens (1) wordt gegeven in figuur 2.

FIGUUR 2

Het vereffenen op de parameter a komt nu hierop neer dat in figuur 1 de som van kwadraten van de afstanden d_i minimaal moet zijn wat bereikt wordt door een geschikte waarde van a te kiezen (te berekenen). In de gegeven voorstelling betekent een vereffening dat het vlak V waarin zich de gegevens bevinden, zo ten opzichte van het oppervlak $F = 0$ wordt verschoven dat $\sum d_i^2$ minimaal geworden is.

De x, y -coördinaten van de gegevens die in V uitgezet zijn, blijven dus steeds dezelfde. De volledige voorstelling is dan deze dat de waarnemingsuitkomsten (gegevens) voorgesteld worden door rechten evenwijdig aan de a -as. De stippen in V zijn dan de doorsnijdingen van V met deze lijnen.

De gebruikelijke berekeningswijze is dan deze dat een beginwaarde a_0 gekozen wordt. Vervolgens wordt het oppervlak $F = 0$ in de richting van de parameter gelineariseerd, dat wil zeggen dat de helling in die richting door een reeksontwikkeling volgens Taylor benaderd wordt. Nu kan een regressieberekening ten opzichte van deze eerstegraads benadering uitgevoerd worden waarbij een correctie op a_0 gevonden wordt. Hiermee is de eerste ronde van de bewerking afgesloten. Met de nieuwe waarde ($a_0 +$ correctie) wordt de berekening herhaald net zo lang tot een correctie = 0 gevonden wordt en a een stabiele waarde heeft aangenomen.

Het effect van de linearisering zal in deze nota nader worden beschouwd. Inplaats echter van de voorstellingswijze uit figuur 1 en figuur 2 zal een representatie met vectoren worden toegepast. Het voordeel hiervan wordt duidelijk door te bedenken dat alle afstandjes d_i door één meetkundige figuur, een vector, kan worden voorgesteld en dat de som van kwadraten ($\sum d_i^2$) het kwadraat van de lengte van die vector is.

3. Vectorvoorstelling van een functie met één parameter

Het probleem van het minimaliseren van een som van kwadraten gaat in vectorvoorstelling over in het bepalen van een kortste afstand of het vaststellen van het voetpunt van een loodlijn.

Teneinde in een aantal, overigens veelal hypothetische, voorbeelden een en ander nader te kunnen toelichten wordt vooraf voor een willekeurig gekozen functie toe te passen vectorvoorstelling toegelicht.

Stel gegeven de betrekking

$$y = x^a \quad (3.1)$$

als bijzonder geval van

$$y = f(x; a)$$

Worden een aantal x -waarden gemeten dan wordt volgens de gegeven functie aan elke x_i een y_i toegevoegd. In kolommen gerangschikt kan gezegd worden dat de kolomvector \underline{x} overgaat in de kolomvector \underline{y} .

Is de kolomvector \underline{x} twee-dimensionaal bijvoorbeeld

$$\underline{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$$

dan ontstaat het volgende overzicht

x_i	y_i indien a is			
	0	1	2	3
2	1	2	4	8
3	1	3	9	27

De kentallen van de vector \underline{y} hangen dus af van de parameter a zodat algemeen geldt

$$\underline{y} = \begin{pmatrix} y_1(a) \\ y_2(a) \end{pmatrix} \quad (3.2)$$

en in het gekozen voorbeeld

$$\underline{y} = \begin{pmatrix} 2^a \\ 3^a \end{pmatrix} \quad (3.3)$$

Worden y_1 en y_2 op coördinaatassen uitgezet dan geeft (3.2) de meetkundige plaats $\varphi(a)$ wanneer a alle mogelijke waarden doorloopt.

FIGUUR 3

Een vector \underline{y}^* zal zijn eindpunt op de curve $\varphi(a)$ hebben indien voor alle y_i voldaan is aan (3.2) respectievelijk (3.3).

De voorstelling in een plat vlak kan dus slechts gegeven worden indien de vector \underline{y} twee kentallen heeft. Bij drie kentallen (drie gegevens of waarnemingsuitkomsten) is een perspectivische voorstelling nog mogelijk bijvoorbeeld in de vorm zoals die in nota 139 wordt gegeven. Voor meer kentallen moet de beschouwingwijze analoog vervolgd worden, een volledige meetkundige voorstelling is dan niet zonder meer mogelijk.

Langs $\varphi(a)$ kan een becijfering voor de opvolgende waarden van a aangebracht worden, waaruit blijkt dat $\varphi(a)$ een parameterkromme is.

Bij figuur 3 wordt nog opgemerkt dat bij een andere reeks x -waarden zowel de kromming van φ als de parameterschaal verandering kunnen ondergaan. Zo is bijvoorbeeld in figuur 9 dezelfde functie voorgesteld doch met

$$\underline{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ e \end{pmatrix}$$

Nu blijkt $\varphi(a)$ geen kromming te hebben.

Elke vector \underline{y}^* kan als plaatsvector van $\varphi(a)$ opgevat worden: voor een gegeven a is zijn plaats op $\varphi(a)$ bepaald.

Stel vervolgens gegeven een vector \underline{y} van waarnemingsuitkomsten van de afhankelijk variabele y behorend bij de meetuitkomsten

$$\underline{x} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$$

Deze vector \underline{y} zal over het algemeen niet zijn eindpunt op $\varphi(a)$ hebben (zie fig. 3). De vector is stochastisch en de verwachtingswaarde

$$E(\underline{y}) = \underline{y}^*$$

zal die \underline{y}^* zijn waarvoor de lengte van de verschilvector

$$|\underline{y} - \underline{y}^*| \tag{3.4}$$

minimaal is.

De kentallen van (3.4) worden (algemeen met n waarnemingen) gevormd door het verschil van elke waarneming \underline{y}_i met de berekende waarde \underline{y}_i^* en zijn dus

$$\begin{pmatrix} y_1 - y_1^* \\ y_2 - y_2^* \\ \vdots \\ y_n - y_n^* \end{pmatrix} = \begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{pmatrix} \tag{3.5}$$

De lengte in het kwadraat van deze vector is $\sum d_i^2$ overeenkomstig het in de laatste zin van paragraaf 2 gestelde.

Het minimaliseren van de som van kwadraten betekent dus dat in het voorbeeld van figuur 4 een waarde a^* van de parameter a gevonden moet worden zodanig dat de vector

$$(\underline{y} - \underline{y}^*) \perp \varphi(a)$$

FIGUUR 4

De berekening kan dusdanig opgezet worden dat ergens op de parameter-curve $\varphi(a)$ een geschatte beginwaarde a_0 als uitgangspunt gekozen wordt. Bij deze a_0 wordt de lengte van de vector $(\underline{y} - \underline{y}^*)$ berekend. Vervolgens kan, door de bewerking voor andere waarden van a te herhalen nagegaan worden in welke richting de verschilvector $(\underline{y} - \underline{y}^*)$ kleiner wordt tot in a^* het minimum bereikt wordt.

Bij deze bewerking behoeven dus geen normaalvergelijkingen opgelost te worden doch wordt steeds bij de aangenomen a , met de gegevens x , de kentallen van de vector \underline{y}^* berekend volgens (3.3). Uit het verschil met de waarnemingen y_i (de kentallen van de vector \underline{y}) volgt dan uit (3.5) de kwadraatsom van de afwijkingen.

4. Vectorvoorstelling van een functie met meer parameters

Wanneer in een functie twee parameters voorkomen dan kan algemeen geschreven worden

$$y = f(x; a, b)$$

Wordt evenals in het vorige geval, nu drie-dimensionaal, een parameter (vector)-voorstelling gegeven dan ontstaat de schematische voorstelling van figuur 5.

FIGUUR 5

Er ontstaat nu een oppervlak met een stelsel kromlijnige coördinaten (a, b) de zogenaamde parameterlijnen zoals in de detailtekening wordt aangegeven. In nota 139 wordt van een dergelijk stelsel een numeriek voorbeeld gegeven.

De loodrechte projectie van het eindpunt van \underline{y} op $\varphi(a, b)$ wordt vastgesteld door coördinaten op het oppervlak namelijk de coördinaten van het

eindpunt van \underline{y}^* bijvoorbeeld

$$(a^*, b^*)$$

Het parameternet kan men zich afgebeeld denken op een platvlak met rechthoekige lineaire coördinaten (fig. 6). Elk punt in dit vlak correspondeert met een punt op het oppervlak $\varphi(a, b)$ uit figuur 5.

FIGUUR 6

Ook nu wordt opgemerkt dat de vorm van het oppervlak en het parameternet bij verandering van de waarden van x_1 , x_2 of x_3 wijzigingen ondergaan. Zijn de x_i 's standvastig dan wordt bij elk paar waarden (a, b) een punt op het oppervlak $\varphi(a, b)$ bepaald. Een paar waarden (a, b) kan in de afbeelding van het parameternet als een vector opgevat worden bijvoorbeeld

$$\underline{\varrho} = \begin{pmatrix} a \\ b \end{pmatrix}$$

Vergelijk figuur 5 (detail) en 6 onderling.

De vraag het voetpunt van de loodlijn uit het eindpunt van \underline{y} op het oppervlak $\varphi(a, b)$ te bepalen komt dan overeen met het bepalen van een vector $\underline{\varrho}^*$ (in fig. 6) waarvoor (in fig. 5)

$$(\underline{y} - \underline{y}^*)$$

een minimale lengte heeft.

Stel dat als eerste stap in de berekening het eindpunt van \underline{y}_0 gebruikt wordt, bepaald door de parameterwaarden (zie fig. 5 en 6)

$$\underline{\varrho}_0 = \begin{pmatrix} a_0 \\ b_0 \end{pmatrix}$$

De verschilvector $(\underline{y} - \underline{y}_0)$ heeft nog een grote lengte. Door voor andere vectoren \underline{y}_0 de berekening van de lengte van de verschilvector te herhalen, door dus als het ware het oppervlak $\varphi(a, b)$ af te tasten, kan nagegaan worden in welke richting de kwadraatsom afneemt. Wordt als volgend punt

$$\underline{\varrho}_1 = \begin{pmatrix} a_1 \\ b_1 \end{pmatrix}$$

gebruikt dan is de verschilvector reeds kleiner. Het eindpunt van de bere-

kening ligt bij $\underline{\theta}^*$. Het verband tussen de (y_1, y_2, y_3) coördinaten en de vectoren $\underline{\theta}$ volgt uit onderlinge vergelijking van de figuren 5 en 6.

In het bovenstaande is de suggestie gewekt dat de berekeningen alleen zouden behoeven te bestaan uit het bepalen van de lengte van de vectoren $(\underline{y} - \underline{y}_i)$. Anders dan met een functie met één parameter moeten voor het volgen van deze werkwijze nu twee parameters van waarde veranderen en vele combinaties van a en b geprobeerd worden om een reeks \underline{y}_i vectoren te verkrijgen.

Het aftasten van het oppervlak kan echter versneld worden door als hulpbewerking een berekening uit te voeren naar de richting waarin bijvoorbeeld vanuit \underline{y}_0 het eindpunt van de vector \underline{y}^* vermoedelijk zal liggen. Het berekenen van een dergelijke richting is het onderwerp van de volgende paragrafen.

5. Benadering van de vector $\Delta \underline{y}$ (fig. 7)

De verschilvector tussen de uitgangsvector \underline{y}_0 en de vector \underline{y}^* wordt voorgesteld door $\Delta \underline{y}$.

FIGUUR 7

Is deze $\Delta \underline{y}$ bekend, of bij benadering te berekenen, dan is de vector

$$\underline{y}_0 + \Delta \underline{y} = \underline{y}_1$$

een goed uitgangspunt voor een tweede berekening. Een benadering van $\Delta \underline{y}$ kan als volgt verkregen worden.

Voor kleine waarden van $\Delta \underline{y}$ geldt volgens de Taylor-ontwikkeling (zie ook Timman, 1959)

$$\begin{aligned} \Delta \underline{y} &= \underline{y}(a + \Delta a, b + \Delta b) - \underline{y}(a, b) & (5.0) \\ &= \left(\Delta a \frac{d}{da} + \Delta b \frac{d}{db} \right) \underline{y} \end{aligned}$$

of in matrixnotatie met indices voor partiële afgeleiden

$$\begin{aligned} \Delta \underline{y} &= (\underline{y}_a, \underline{y}_b) \begin{pmatrix} \Delta a \\ \Delta b \end{pmatrix} \\ &= (\underline{y}_a, \underline{y}_b) \Delta \underline{\theta} & (5.1) \end{aligned}$$

en korter, met

$$M = (\underline{y}_a, \underline{y}_b) \quad (5.2)$$

$$\underline{y} = M \Delta \underline{\theta}$$

waarin M dus de matrix van partiële afgeleiden voorstelt.

Opgemerkt wordt dat de vectoren

$$\underline{y}_a = \frac{dy}{da} \quad \text{en} \quad \underline{y}_b = \frac{dy}{db} \quad (5.3)$$

raakvectoren zijn respectievelijk aan de a-coördinaten en aan de b-coördinaten ter plaatse van het punt (a_i, b_i) waarvoor de berekening wordt uitgevoerd.

Deze raakvectoren zijn geen eenheidsvectoren doch hebben respectievelijk een lengte

$$\begin{aligned} h_1 &= |\underline{y}_a| \\ h_2 &= |\underline{y}_b| \end{aligned} \quad (5.4a)$$

De grootheden h worden de schaalfactoren van de bijbehorende raakvectoren genoemd (Timman 1959).

Uit de betrekkingen volgt dus dat

$$h_1^2 = \underline{y}_a \cdot \underline{y}_a \quad (5.4b)$$

en

$$h_2^2 = \underline{y}_b \cdot \underline{y}_b$$

wat de diagonaal-elementen zijn van de symmetrische matrix tMM , zoals uit (5.2) volgt, waarin tM de getransponeerde van M is.

De beide raakvectoren (5.3) spannen het raakvlak aan $\varphi(a, b)$ op in het punt waarvoor de partiële afgeleiden zijn ontwikkeld, bijvoorbeeld P in figuur 8.

Het raakvlak in $P(a_o, b_o)$ met plaatsvector \underline{y}_o wordt in vectorvorm gegeven door

$$\underline{x} = \underline{y}_o + \alpha \underline{y}_a + \beta \underline{y}_b \quad (5.5)$$

De linearisering heeft tot gevolg dat het parameternet van het oppervlak $\varphi(a, b)$ op het raakvlak overgaat in een scheefhoekig assenstelsel. De assen hiervan zijn lineair en hebben de richting van respectievelijk \underline{y}_a en \underline{y}_b . Het nulpunt is het punt P, de schaalverdelingen langs de assen verlopen weliswaar lineair doch de schaalfactoren zijn respectievelijk h_1 en h_2 volgens (5.4) en zullen dus in het algemeen niet gelijk zijn.

FIGUUR 8

De benadering van $\Delta \underline{y}$ wordt nu verkregen door uit het eindpunt van \underline{y} een loodlijn op het raakvlak neer te laten. In figuur 8 stelt P het punt voor van $\varphi(a, b)$ dat als eerste schatting wordt gebruikt en waarmee \underline{y}_0 uit de aangenomen a_0 en b_0 berekend wordt.

De oplossing komt neer op het berekenen van een lineaire regressie van $(\underline{y} - \underline{y}_0)$ op de basisvectoren van het raakvlak. De basisvectoren zijn de naar het punt P ontwikkelde partiële afgeleiden en worden volgens (5.2) voorgesteld door de matrix M_0 . De vector \underline{y}_R is de vector die zijn eindpunt in het raakvlak heeft ter plaatse van het voetpunt van de loodlijn uit het eindpunt van $(\underline{y}$ en $\underline{y}_0)$ op dat raakvlak. De gebruikelijke afleiding van de normaalvergelijkingen volgt uit de voorwaarde dat de vector

$$(\underline{y} - \underline{y}_R)$$

loodrecht moet staan op de basisvectoren van het raakvlak, dus loodrecht moet staan op de kolomvectoren van M_0 zodat

$${}^t M_0 (\underline{y} - \underline{y}_R) = 0$$

of, na invoeren van de vector \underline{y}_0

$${}^t M_0 (\underline{y} - \underline{y}_0) - {}^t M_0 (\underline{y}_R - \underline{y}_0) = 0 \quad (5.6)$$

waarvan de laatste verschilvector in het raakvlak ligt (fig. 8) en een lineaire combinatie is van de basisvectoren zodat, overeenkomstig (5.1) en (5.2).

$$(\underline{y}_R - \underline{y}_0) = M_0 \Delta \underline{\theta} \quad (5.7)$$

Wordt (5.7) gesubstitueerd in (5.6) dan ontstaan de normaalvergelijkingen

$${}^t M_0 M_0 \Delta \underline{\theta} = {}^t M_0 (\underline{y} - \underline{y}_0)$$

waarvan de oplossing is

$$\underline{\Delta\theta} = ({}^t M_o M_o)^{-1} {}^t M_o (y - y_o) \quad (5.5)$$

(zie fig.8a in de tekst, deze figuur is een detail van fig.8)(pag.13).

Hiermee is een correctie op de beginschattingen van de parameters verkregen zodat (zie ook (5.0))

$$\underline{\theta}_1 = \underline{\theta}_o + \underline{\Delta\theta} = \begin{pmatrix} a_o \\ b_o \end{pmatrix} + \begin{pmatrix} \Delta a \\ \Delta b \end{pmatrix} = \begin{pmatrix} a_o + \Delta a \\ b_o + \Delta b \end{pmatrix} = \begin{pmatrix} a_1 \\ b_1 \end{pmatrix}$$

Een volgende berekening kan nu met het punt P op $\varphi(a,b)$ uitgevoerd worden waarvoor geldt dat

$$P = P(a_1, b_1)$$

In principe kan nu iteratief te werk gegaan worden door achtereenvolgens te berekenen

$$\underline{\theta}_{i+1} = \underline{\theta}_i + \underline{\Delta\theta}$$

De bewerking kan gestopt worden indien een $\underline{\Delta\theta}$ verkregen wordt die verwaarloosbaar klein geacht kan worden. Theoretisch moet dan

$$\underline{\Delta\theta} = \begin{pmatrix} \Delta a \\ \Delta b \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \underline{0}$$

In dit geval zegt men dat convergentie naar $\underline{\theta}^*$ is opgetreden daar elke volgende bewerking opnieuw $\underline{\Delta\theta} = \underline{0}$ tot uitkomst zal hebben waardoor $\underline{\theta}^*$ niet meer van kantallen verandert.

In het geval van divergentie zal bijvoorbeeld Δa bij volgende iteraties een grotere waarde krijgen zodat geen stabiele vector $\underline{\theta}^*$ gevonden wordt.

De gevolgde procedure kan aangeduid worden als een iteratief vereffeningsproces.

6. Nadere beschouwing van de linearisering

Door de functie

$$y = f(x; a, b)$$

in de richting van de parameters te lineariseren worden benaderingen toegepast voor:

1. de richting van het vlak $\varphi(a,b)$
2. de richting van de parameterkrommen
3. de schaalverandering langs de parameterkrommen

Als bezwaar van de linearisering wordt veelal punt 1 genoemd. Met een voorbeeld zal toegelicht worden dat ook de punten 2 en 3 complicaties bij de vereffening kunnen veroorzaken en een reden kunnen zijn van het niet-convergeren wanneer iteratief vereffend wordt.

Uitgegaan wordt van (3.1) met

$$\underline{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ e \end{pmatrix} \text{ en } \underline{y} = \begin{pmatrix} 1^a \\ e^a \end{pmatrix} \quad (\text{zie fig. 9})$$

De vector $\underline{\theta}$ heeft slechts één kental namelijk a . De matrix M bestaat dus slechts uit één kolomvector en is

$$M = \frac{d}{da} \underline{y} = \begin{pmatrix} 0 \\ e^a \end{pmatrix}$$

De schaalfactor is vervolgens

$$h_1 = \sqrt{0^2 + (e^a)^2} = e^a$$

Nu wordt berekend

$$({}^t_{MM})^{-1} {}^t_M = \frac{1}{e^{2a}} (0, e^a) = (0, e^{-a})$$

zodat algemeen geldt volgens (5.5)

$$\Delta^a = (0, e^{-a_0})(\underline{y} - \underline{y}_0)$$

Met als beginwaarde

$$a_0 = 1$$

wordt

$$\underline{y}_0 = \begin{pmatrix} 1 \\ 2,72 \end{pmatrix}$$

zodat

$$\Delta a = \frac{(10 - 2,72)}{2,72} = 2,7$$

wat de regressiecoëfficiënt is van $(\underline{y} - \underline{y}_0)$ op $\varphi(a)$

FIGUUR 9

Gevonden wordt

$$a_1 = a_0 + \Delta a = 3,7$$

met

$$\underline{y}_1 = \begin{pmatrix} 1 \\ 40,4 \end{pmatrix}$$

Bij benadering geldt voor de lengten

$$|\underline{y} - \underline{y}_0| = 3,8 \quad (6.1)$$

en

$$|\underline{y} - \underline{y}_1| = 15 \quad (6.2)$$

De kwadraatsom van de afwijkingen (de lengte-kwadraat van de verschilvector) is dus ongeveer 4x zo groot geworden waarbij opgemerkt wordt dat de meetkundige plaats $\varphi(a)$ voor het gestelde geval lineair is (zie fig.9) en dat het voetpunt van de loodlijn uit het eindpunt van $(\underline{y} - \underline{y}_0)$ op de "benaderde" richting $\frac{d}{da} \underline{y}$ exact op de goede plaats op $\varphi(a)$ ligt.

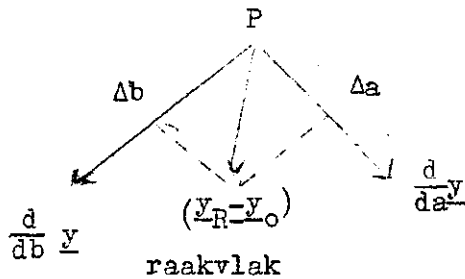
7. De invloed van de schaalfactoren

Het grote verschil tussen de uitkomst (6.1) en (6.2) houdt in dat na de correctie op a_0 een grotere verschilvector wordt gevonden en dat de correctie als het ware zijn doel is voorbijgeschoten. De oorzaak ligt in het feit dat er geen meetkundige betrekking tussen de correctie (regressiecoëfficiënt) Δa en het nieuwe punt $a_1 = a_0 + \Delta a$ bestaat.

De correctie Δa wordt immers uit (5.5) als een regressiecoëfficiënt berekend. Het nieuwe punt a_1 wordt verkregen door deze coëfficiënt bij a_0 op te tellen zodat $a_0 + \Delta a = a_1$ en dit punt op de meetkundige plaats $\varphi(a)$ op te zoeken. Het hangt dan mede van de schaalverandering langs $\varphi(a)$ af of het nieuwe punt a_1 dichter bij de oplossing a^* ligt dan het punt a_0 .

In figuur 8 bijvoorbeeld worden de correcties gevonden uit de ontbinding van $(y_R - y_0)$ op de basisvectoren die het regressievlak opspannen (zie fig. 8a). Dit levert de regressiecoëfficiënt Δa en Δb , die samen de vector

$$\underline{\Delta\theta} = \begin{pmatrix} \Delta a \\ \Delta b \end{pmatrix}$$



Figuur 8a

vormen.

In figuur 8 wordt het nieuwe punt P dus gevonden door op het oppervlak $\varphi(a, b)$ met behulp van het parameternet het punt

$$\underline{\theta} = \underline{\theta}_0 + \underline{\Delta\theta}$$

op te zoeken. Afhankelijk van het beloop van de parameterkrommen en de schaalfactoren zal een punt op $\varphi(a, b)$ gevonden worden dat dichtbij of verder verwijderd van de gevraagde oplossing zal liggen. Een numeriek voorbeeld van een drie-dimensionale vectorvoorstelling wordt in nota 139 behandeld.

Het voorbeeld van figuur 9 geeft nog aanleiding tot de volgende beschouwing.

De schaalfactor h van y_a in de richting van $\varphi(a)$ wordt berekend volgens (5.4b) en wel

$$h = \sqrt{\frac{dy}{da} \cdot \frac{dy}{da}}$$

wat voor het voorbeeld wordt

$$h = e^a \quad \left\{ \begin{array}{l} = 2,72 \text{ voor } a_0 = 1 \text{ (fig.9)} \\ = 40,45 \text{ voor } a_1 = 3,7 \end{array} \right.$$

De verandering van de schaalfactor volgt uit de afgeleide naar de parameter, in dit geval

$$\frac{dh}{da} = e^a$$

Heeft deze afgeleide, zoals hier, een positief teken dan is bij de kleine beginwaarde de mogelijkheid steeds aanwezig dat het gecorrigeerde punt, gevonden na de eerste iteratie, verder van de oplossing afligt dan deze beginwaarde (zie fig.9). Bij een te grote beginwaarde zal wel convergentie optreden doch deze zal zeer langzaam verlopen. Is de afgeleide negatief van teken dan verloopt deze eigenschap tegengesteld; is de afgeleide 0 dan verandert de schaalfactor niet en zal het al of niet convergeren uitsluitend van de kromming van $\varphi(a)$ afhangen.

Het bovenstaande kan bovendien nog afhankelijk zijn van overige parameters daar algemeen de schaalfactoren functies zullen zijn van alle parameters en de afgeleiden dus eveneens bijvoorbeeld

$$\frac{dh}{da} = \Psi_1(a,b)$$

$$\frac{dh}{db} = \Psi_2(a,b)$$

FIGUUR 10

Man kan zich nu het volgende voorstellen.

Heeft het parameternet een vorm als in figuur 10 wordt aangegeven dan valt vanuit θ_0 in de richting langs a_0 convergentie te verwachten. In de richting langs b_0 is de mogelijkheid van divergentie aanwezig door het veronderstelde toenemen van de schaalfactoren. Stel er wordt na een eerste iteratie het punt θ_1 verkregen. Nu zijn door de gewijzigde vorm van het parameternet ter plaatse de rollen omgekeerd en bestaat de mogelijkheid dat a sterk divergeert en b convergeert zodat bijvoorbeeld θ_2 bereikt wordt. Opgemerkt wordt dat bij de berekening in θ_1 , een ander raakvlak gebruikt wordt zodat in een en ander tevens de kromming van het vlak $\varphi(a,b)$ een rol speelt.

Wanneer geldt dat bijvoorbeeld de parameters die in de verschillende functies voorkomen achtereenvolgens zijn

$$y = y(a,b)$$

$$y_a = y_a(b)$$

$$h = \int y_a \cdot y_a = f(b)$$

$$\frac{dh}{da} = 0 \tag{7.1}$$

$$Y_{ab} = Y_{ab}(b) \quad (7.2)$$

dan verandert de schaalfactor niet door (7.1) doch zal de van b afhankelijke kromming welke in (7.2) tot uiting komt (Timman, 1959) het verloop van het vereffeningsproces bepalen. Uit deze overwegingen kunnen eigenschappen van het vlak $\varphi(a, b)$ met behulp van de vectoranalyse afgeleid worden.

8. De berekening van de booglengte langs een parameterkromme

In eenvoudige gevallen kan een snellere convergentie tot stand komen door de volgende berekeningen uit te voeren welke geïllustreerd wordt aan de hand van het voorbeeld van figuur 9.

De afgeleide vector $\frac{d}{da}Y$ heeft voor $a_0 = 1$ een lengte van 2,718. Na de vereffening ontstaat langs $\varphi(a)$ een vector met lengte (fig. 9):

$$\sqrt{(2,718e^{a_0})^2} = 7,389 \quad (8.1)$$

Voor de booglengte s langs $\varphi(a)$ tussen a_0 en a_1 geldt

$$s = \int_{a_0}^{a_1} \sqrt{Y_a \cdot Y_a} da \quad (8.2)$$

waarvoor na enige herleiding en met gebruik van (8.1) volgt

$$s = [e^a]_{a_0}^{a_1} = 7,389$$

De ondergrens is $a_0 = 1$ waaruit voor de bovengrens volgt

$$a_1 = 2,31^5 \quad (8.3)$$

De werkelijke waarde is 2,303 zodat hiermee wel een goed uitgangspunt voor een volgende iteratiestap is verkregen.

Het oplossen van de integraal (8.2) zal veelal niet mogelijk zijn, waarbij bovendien bedacht moet worden dat hier slechts het geval met één parameter behandeld is.

Een andere mogelijkheid in het gegeven voorbeeld is nog de berekening van

$$\begin{aligned} \underline{y}^* &\approx \underline{y}_0 + M\Delta a \\ &= \begin{pmatrix} 1 \\ 2,27 \end{pmatrix} + \begin{pmatrix} 0 \\ e \end{pmatrix} 2,72 = \begin{pmatrix} 1 \\ 11,65 \end{pmatrix} \end{aligned}$$

waarvoor geldt

$$a_1 = 2,45$$

Deze uitkomst is weer wat minder goed dan (8.3). Bovendien is de bewerking slechts uitvoerbaar wanneer

$$(\underline{y}^* - \underline{y}^0) \text{ in } \varphi(a) \text{ ligt}$$

Wat zich slechts in zeer bijzondere gevallen zal voordoen.

9. Het minimaliseren van de kwadraatsom

De nadelen verbonden aan het lineariseren van de functie $\varphi(a,b)$ kunnen gedeeltelijk opgeheven worden door na een eerste vereffening dat is een regressie-berekening in een raakvlak, de kwadraatsom systematisch kleiner te maken door het oppervlak $\varphi(a,b)$ in de richting van $\Delta\theta$ af te tasten.

Met de beginschatting (zie fig.12)

$$\underline{\theta}_0 \begin{pmatrix} a_0 \\ b_0 \end{pmatrix}$$

en de verkregen waarde

$$\underline{\theta}_1 = \underline{\theta}_0 + \Delta\underline{\theta} = \begin{pmatrix} a_1 \\ b_1 \end{pmatrix}$$

Worden respectievelijk de bijbehorende kwadraatsommen S_0 en S_1 bepaald als afwijkingen ten opzichte van het oppervlak $\varphi(a,b)$.

Vervolgens wordt, nog steeds ten opzichte van $\varphi(a,b)$ een tussenliggende waarde berekend bijvoorbeeld $S_{1/2}$ voor

$$\underline{\theta} = \underline{\theta}_0 + \frac{1}{2}\Delta\underline{\theta}$$

Volgende berekeningen kunnen dienen voor het nader vastleggen van het punt waar S minimaal wordt.

Algemeen

$$\underline{\theta}_i = \underline{\theta}_0 + \epsilon \Delta \underline{\theta} \quad (9.1)$$

Als tegenhanger van het geval met één parameter, toegelicht aan het einde van par.3, kan een figuur getekend worden waarin S en ϵ als variabelen voorkomen (fig.11a).

FIGUUR 11a

Al naar de verkregen uitkomsten kan geïnterpolceerd respectievelijk geëxtrapoleerd worden. Door Booth and Peterson (1960) wordt deze werkwijze systematisch toegepast door steeds met halve respectievelijk dubbele afstanden te werken. In de figuur wordt met $i = 0$ aangegeven dat de kwadraat-son van het punt $\underline{\theta}_0$ is berekend, $i = 1$ voor $\underline{\theta}_1$ enz. Tenslotte wordt onder de voorwaarden bij figuur 11a vermeld een tweedegraads kromme door de laatste drie punten berekend, waaruit het minimum bepaald wordt.

Voor het interpolatie-geval geldt dan

$$\epsilon = \frac{3}{4} \left(\frac{1}{2}\right)^i \frac{4S_i - 5S_{i-1} + S_{i-2}}{2S_i - 3S_{i-1} + S_{i-2}} \quad (9.2)$$

en voor het extrapolatie-geval

$$\epsilon = \frac{3}{8} (2)^i \frac{4S_{i-2} - 5S_{i-1} + S_i}{2S_{i-2} - 3S_{i-1} + S_i} \quad (9.3)$$

Een wat eenvoudiger werkwijze wordt besproken door Hartley (1961). Hier toe worden rechts berekend de kwadraatssommen voor $\epsilon = 0$, $\epsilon = 1/2$ en $\epsilon = 1$. Het minimum wordt nu gevonden uit

$$\epsilon_{\min} = \frac{1}{2} + \frac{1}{4} \frac{S_0 - S_1}{S_1 - 2S_{1/2} + S_0} \quad (9.4)$$

Het aantal S-waarden dat berekend moet worden blijft nu tot een drietal beperkt. Het gevonden minimum zal echter een benadering zijn van dat welke met (9.2) of (9.3) gevonden.

Opgemerkt wordt dat bij de bovenomschreven berekeningen geen vereffening meer is toegepast, dat wil zeggen dat niet opnieuw een stel normaalvergelijkingen moet worden opgelost.

Nadat echter de ligging van het minimum van S volgens (9.2), (9.3) of

(9.4) is bepaald kan met de bijbehorende ϵ volgens (9.1) een nieuwe, betere beginschatting voor een volgende regressieberekening in het raakvlak (fig. 8) gevonden worden.

Hiermee wordt dan een nieuwe richting bepaald waarin het oppervlak φ (a,b) verder afgetast kan worden. De beschreven procedure van interpolatie en extrapolatie kan nu weer leiden tot een verdere verlaging van de kwadraatsom. Belangrijk is dat niet eerder hoeft te worden vereffend dan nadat de kwadraatsom verlaagd is.

Zo de gevolgde werkwijze een naam behoeft naast de vereffeningprocedure zou deze aangeduid kunnen worden als de methode van behoud van de kleinste kwadraatsom.

Figuur 11a geeft een indruk van de samenhang tussen S en ϵ . Na een volgende iteratie zal het minimum wat lager komen zoals in figuur 11b wordt weergegeven en verschuiven naar de waarde $\epsilon = 1$. Bij de laatste iteraties zal S zijn minimum waarde bereiken en zal $\Delta\theta \rightarrow 0$ zodat in (9.1) de uitkomst onafhankelijk van ϵ is. Dit wordt in figuur 11b door de horizontale rechte tot uitdrukking gebracht.

FIGUUR 11b

In de voorstellingswijze die het parameternet op gewone coördinaten afbeeldt (fig. 12) kan het verschil tussen beide werkwijzen als volgt schematisch worden weergegeven.

Stel dat het voetpunt van de normaal uit het eindpunt van \underline{y} afgebeeld wordt in $\underline{\theta}^*$. Stel vervolgens dat een beginschatting voorgesteld wordt door de vector $\underline{\theta}_0$. Na vereffening wordt een $\Delta\underline{\theta}_0$ bepaald.

Zou iteratief met $\underline{\theta}_1 = \underline{\theta}_0 + \Delta\underline{\theta}_0$ verder vereffend worden dan zouden bijvoorbeeld achtereenvolgens de punten i verkregen kunnen worden, uiteindelijk wordt dan, indien convergentie optreedt $\underline{\theta}^*$ bereikt.

FIGUUR 12

Volgens de zojuist besproken methode van behoud van de kleinste kwadraatsom wordt het oppervlak in de richting van $\Delta\underline{\theta}_0$ afgetast. Ongeacht het aantal parameters zal steeds

$$\underline{\theta} = \underline{\theta}_1 + \epsilon \Delta\underline{\theta}_1 \quad (9.5)$$

een rechte voorstellen met richting ΔO_1 gaande door het punt θ_1 . Door verschillende waarden aan ϵ toe te kennen wordt de parameterruimte volgens de rechte (9.5) afgetast. In figuur 12 wordt deze 1e bewerking voorgesteld door de rechte in de 1e richting. Is het punt op deze rechte bereikt waarvoor S minimaal is dan wordt na een regressieberekening gevonden dat in de 2e richting verder gewerkt moet worden. Opgemerkt wordt dat de rechten (9.5) op het oppervlak $\varphi(a,b)$ als krommen afgebeeld worden zoals in nota 139 met een voorbeeld zal worden toegelicht.

10. Berekening en afbeelding van de kwadraatsom

De som van de kwadraten van de afwijkingen kan berekend worden volgens de in I.C.W.nota 113 (1961) besproken werkwijze. Hier kan zonodig rekening gehouden worden met een vooraf vastgestelde richting van middelen.

De kwadraatsom kan voorgesteld worden door een vector met lengte gelijk aan de wortel uit deze som. Wordt de vector waarvoor de kwadraatsom minimaal is op poolcoördinaten uitgezet, $(\underline{y}-\underline{y}^*)$ van figuur 4, dan kan een indruk verkregen worden van de wijze waarop het oppervlak de minimale afstand tot het eindpunt van \underline{y} bereikt, gaande van een willekeurige vector \underline{y}_0 naar een andere willekeurige vector \underline{y}_1 .

FIGUUR 13

Hiertoe wordt de hoek die de berekende vectoren $(\underline{y}-\underline{y}_0)$ en $(\underline{y}-\underline{y}_1)$ maakt met $(\underline{y}-\underline{y}^*)$ bepaald. Zijn deze hoeken φ_1 en φ_2 dan zal van niet te veraf gelegen vectoren een afbeelding van onderlinge ligging en daarmee van de vorm van het oppervlak $\varphi(a,b)$ verkregen kunnen worden, wanneer tevens alle lengten bekend zijn.

Is de hoek tussen \underline{y}_0 en \underline{y}_1 gelijk aan φ_3 dan zal ongeveer

$$\varphi_1 + \varphi_2 = \varphi_3$$

moeten zijn.

11. Het gedrag van de parameters

Met behulp van de vectoranalyse zouden de eigenschappen van het $\varphi(a,b)$ -vlak nader bestudeerd kunnen worden. Ook op numerieke wijze kan

volgens de vorige paragraaf een indruk van het gedrag van elk van de parameters afzonderlijk verkregen worden.

Wordt aangenomen dat voor kleine veranderingen van een parameter, bijvoorbeeld a bij constante b , een goede afbeelding in het platte vlak kan worden verkregen dan kan een figuur met poolcoördinaten getekend worden. Uitgaande van de eindoplossing wordt de waarde van één parameter iets verhoogd en vervolgens iets verlaagd. De bijbehorende verschilvectoren worden bepaald en er wordt bijvoorbeeld de toestand van figuur 14a verkregen.

FIGUREN 14a EN b

De cirkel geeft de lengte aan van de loodlijn uit het eindpunt van \underline{y} op $\varphi(a,b)$ en vertegenwoordigt de minimale kwadraatsom. Het middelpunt is het eindpunt van \underline{y} (zie ook fig. 13).

De parameter a volgt langs de parameterkromme $b=\text{constant}$ een weg met geringe kromming, zodat in dat geval op a lineair vereffend kan worden. De beginwaarde hoeft niet al te nauwkeurig te zijn.

Een ander geval wordt aanwezig geacht voor de parameter b indien a constant gehouden wordt. In de richting van b is het oppervlak $\varphi(a,b)$ sterk gekromd. De methode van het aftasten van het oppervlak zal in deze richting de aangewezen methode zijn om de kwadraatsom te verlagen.

12. Enkele bijzondere gevallen

Betrekkingen tussen parameters

Bestaat tussen een aantal parameters een zekere betrekking bijvoorbeeld zo dat

$$b = g(a) \quad (11.1)$$

dan komt dit op $\varphi(a,b)$ tot uiting doordat op dit oppervlak een curve ligt waarvan alle punten aan (11.1) voldoen (fig.15).

FIGUUR 15

Door het in rekening brengen van de betrekking (11.1) zal de lengte van de verschilvector $(\underline{y}-\underline{y}^*)$ toenemen zoals uit een meetkundige beschouwing dadelijk blijkt. Bij de vereffening kan reeds vanaf het begin de gegeven betrekking opgenomen worden. Ook is het mogelijk a en b als afzonderlijke parameter op te vatten en daarna (11.1) in de bewerking op te

op te nemen. Dit vindt plaats door op $g(a)$ twee punten te kiezen en het minimum van de kwadraatsom volgens (9.2) respectievelijk (9.3) vast te stellen. Tenslotte kan dan nog een volgende vereffening toegepast worden.

Parameters met grenswaarden

FIGUUR 16

Stel dat een parameter bijvoorbeeld b een waarde van 100 niet mag overschrijden. In dat geval wordt eerst normaal vereffend met beginschatting θ_0 totdat blijkt dat voor een θ_1

$$b + \Delta b > 100$$

Nu wordt uit (9.1) een dusdanige waarde gegeven dat

$$b + \varepsilon \Delta b = 100$$

Nu wordt b constant gehouden op 100 en met θ_1 als nieuwe beginschatting wordt de minimum kwadraatsom door aftasten van het oppervlak langs de curve $b = 100$ bepaald. Tenslotte kan men b nog eenmaal een wat hogere respectievelijk lagere waarde geven en de bijbehorende kwadraatsom uitrekenen. In figuur 16 is bijvoorbeeld het geval geschetst dat tenslotte blijkt dat b toch iets kleiner is dan 100.

Schaalfactoren van de afhankelijke variabele

Zoals uit figuur 9 volgt moet er een waarde γ bestaan zodanig dat van γy het eindpunt op $\varphi(a)$ ligt en dus

$$y^* = \gamma y$$

Er is nu één waarde van a en γ die veroorzaakt dat aan $y = \frac{1}{\gamma} e^{ax}$ voldaan wordt.

Dit is in dit voorbeeld een gevolg van het feit dat nu twee vergelijkingen met twee onbekenden (a en γ) ontstaan. Met meer dimensies zal over het algemeen geen oplossing verkregen worden doch wel zal steeds nagegaan kunnen worden zoals in figuur 17 staat aangegeven welke factor γ de kwadraatsom het sterkst doet afnemen. Komt geen schaalfactor in de gegeven functie voor dan zou na de beëindigde vereffening zonodig achteraf nog nagegaan kunnen worden of een dergelijke factor γ aanwezig is, en wat de grootte ervan is.

FIGUUR 17

Ook in dit geval hoeft niet met een nieuwe vereffening begonnen te worden doch kan met interpolatie en extrapolatie de eindoplossing benaderd worden. Daarna kan met een regressieberekening in het raakvlak (fig.8) nagegaan worden of het eindpunt van de bewerking reeds is bereikt.

13. Slotopmerkingen

In het voorgaande werd uiteengezet op welke wijze het gedrag van de parameters van een functie geanalyseerd kan worden. Hiermee kan een nader inzicht verkregen worden in de wijze van convergeren naar een eindoplossing.

De complicaties die zich bij niet-lineaire vereffening voordoen te weten de meerwaardigheid van de functie, het voorkomen van asymptoten en de mate waarin convergentie optreedt lijken nu te kunnen worden overzien.

Evenwel kan nog een probleem genoemd worden dat het vinden van een oplossing bemoeilijkt, namelijk dat van het optreden van nevenoplossingen.

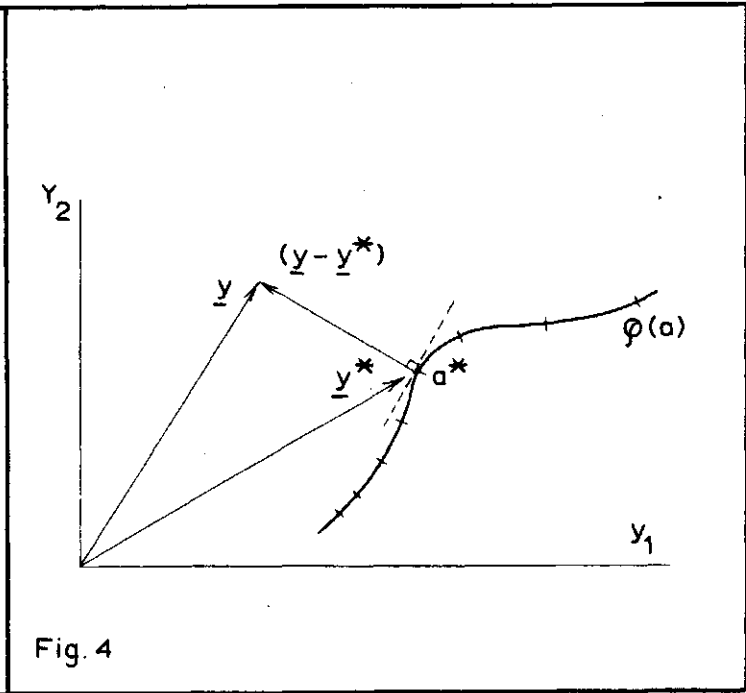
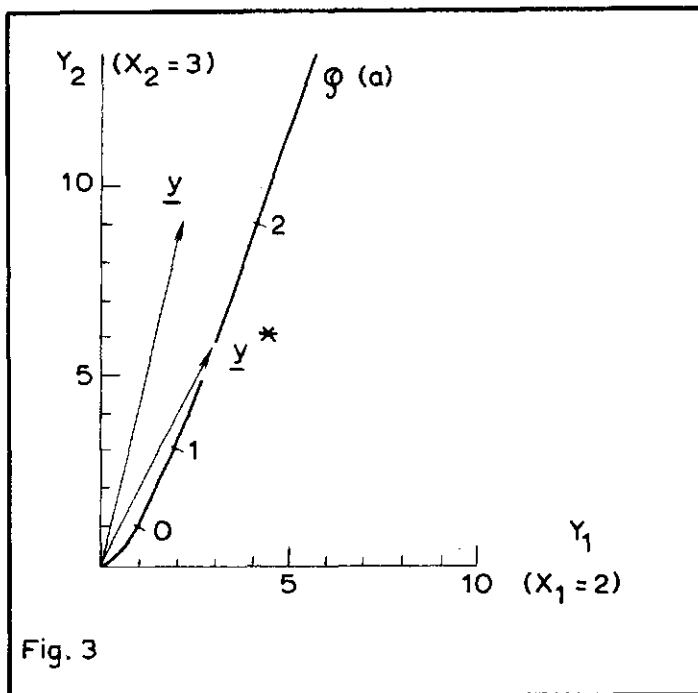
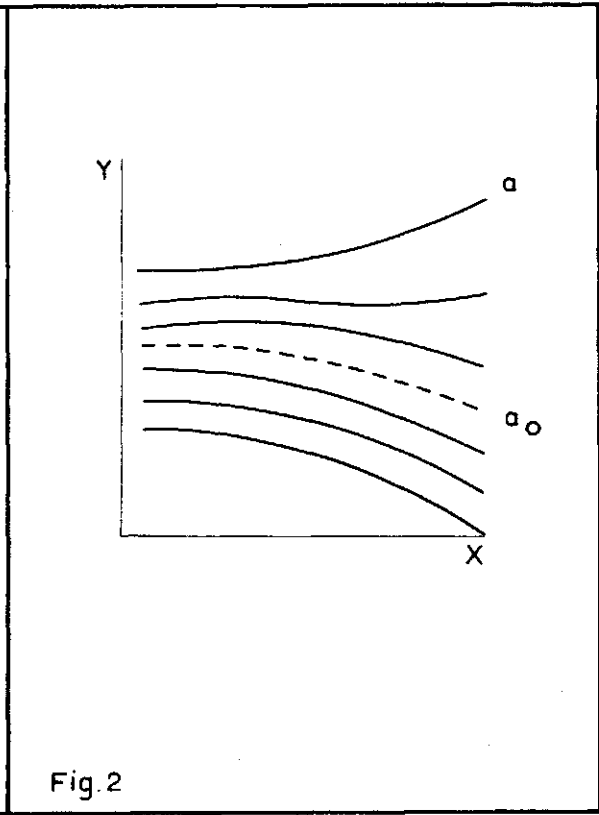
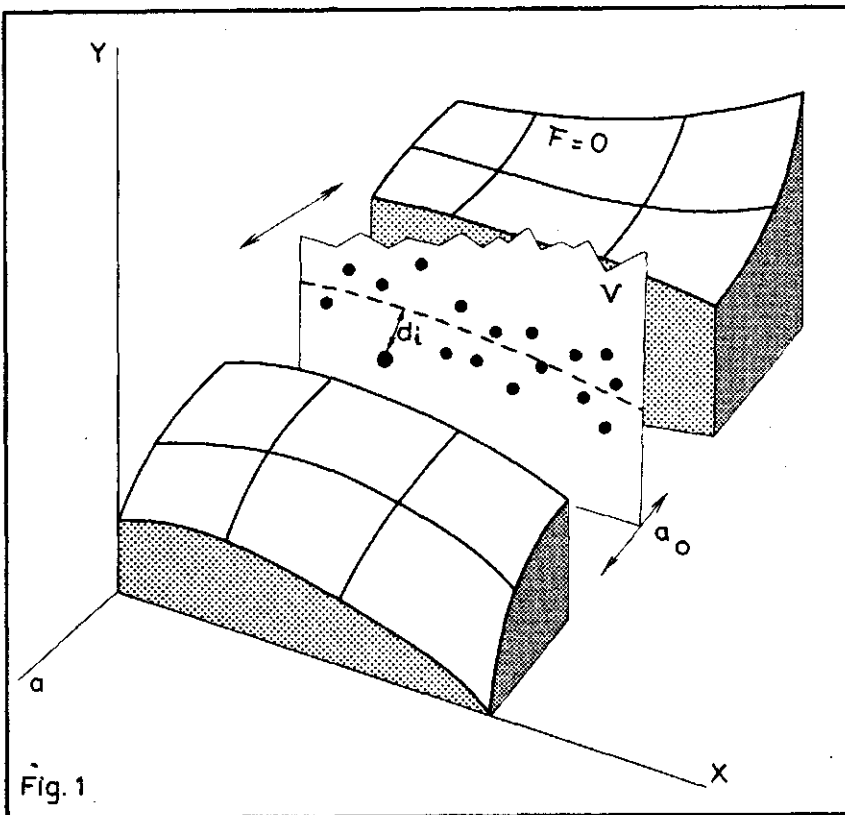
FIGUUR 18

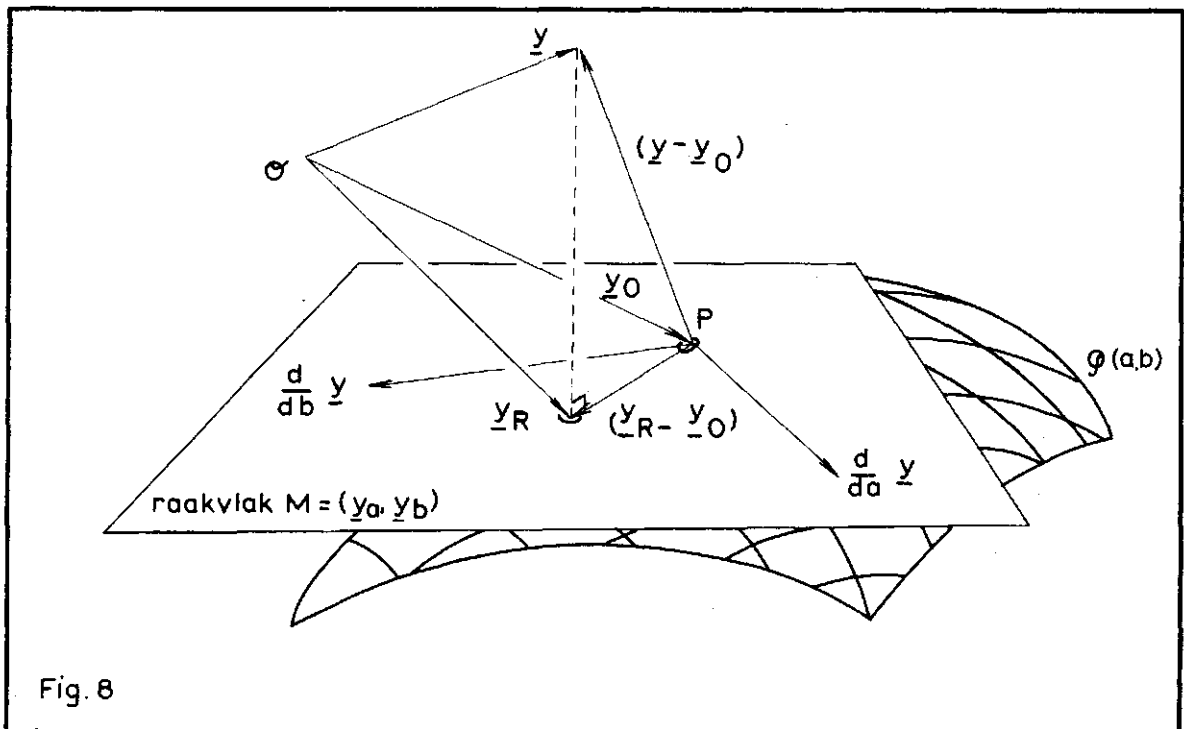
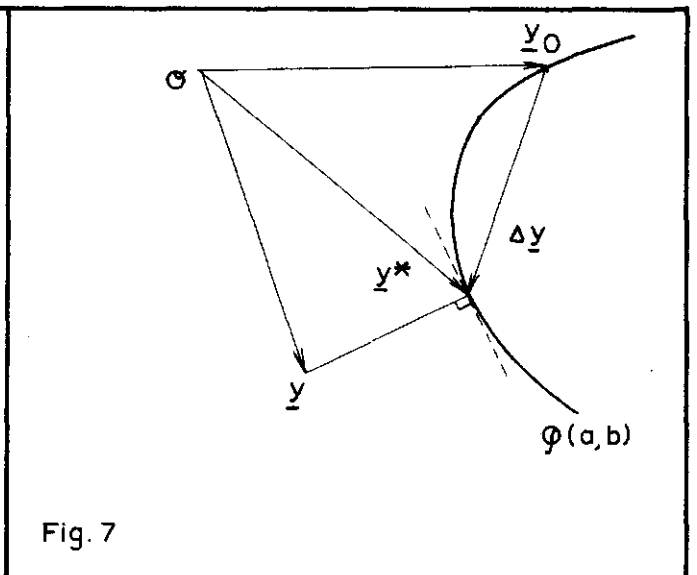
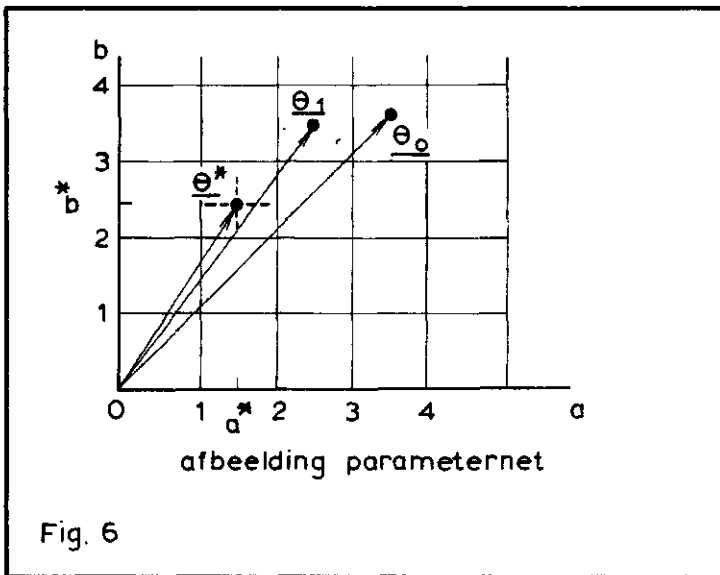
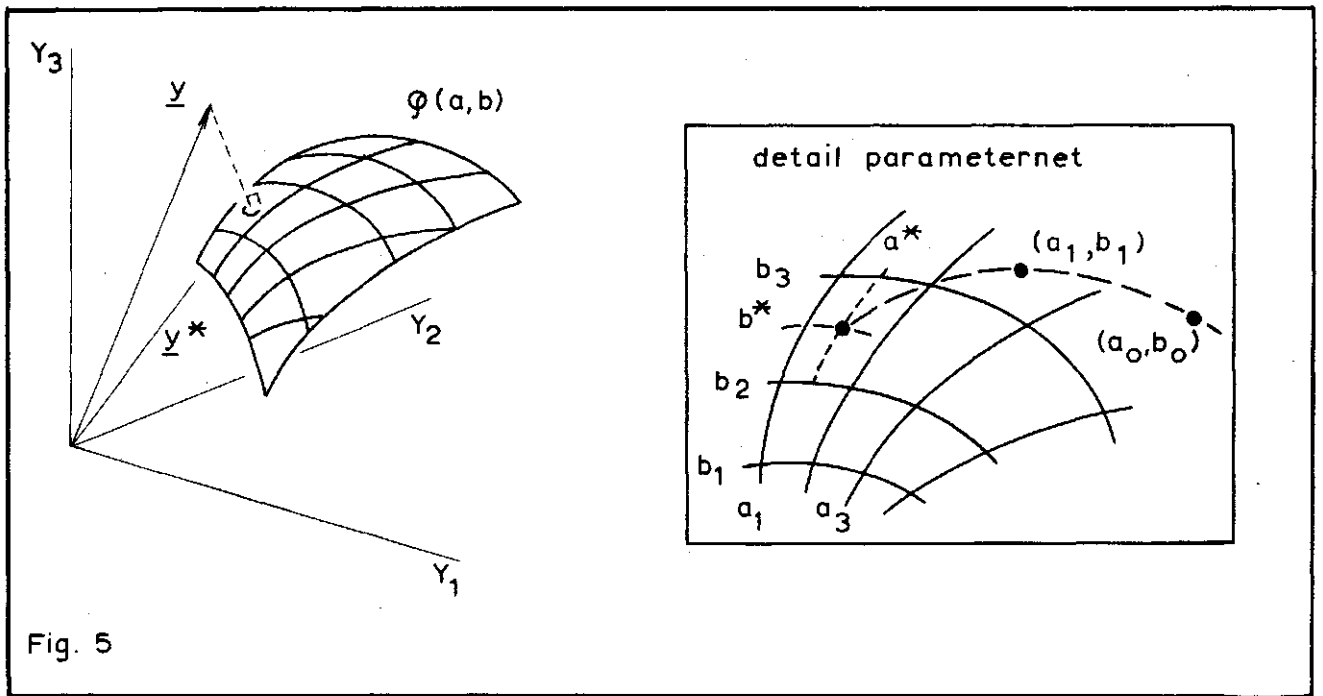
Uit figuur 18 volgt dat nevenoplossingen kunnen optreden afhankelijk van de vorm van $\varphi(a)$ respectievelijk $\varphi(a,b)$. Beide oplossingen zijn stabiel doch behoeven niet eenzelfde minimale kwadraatsom te bezitten.

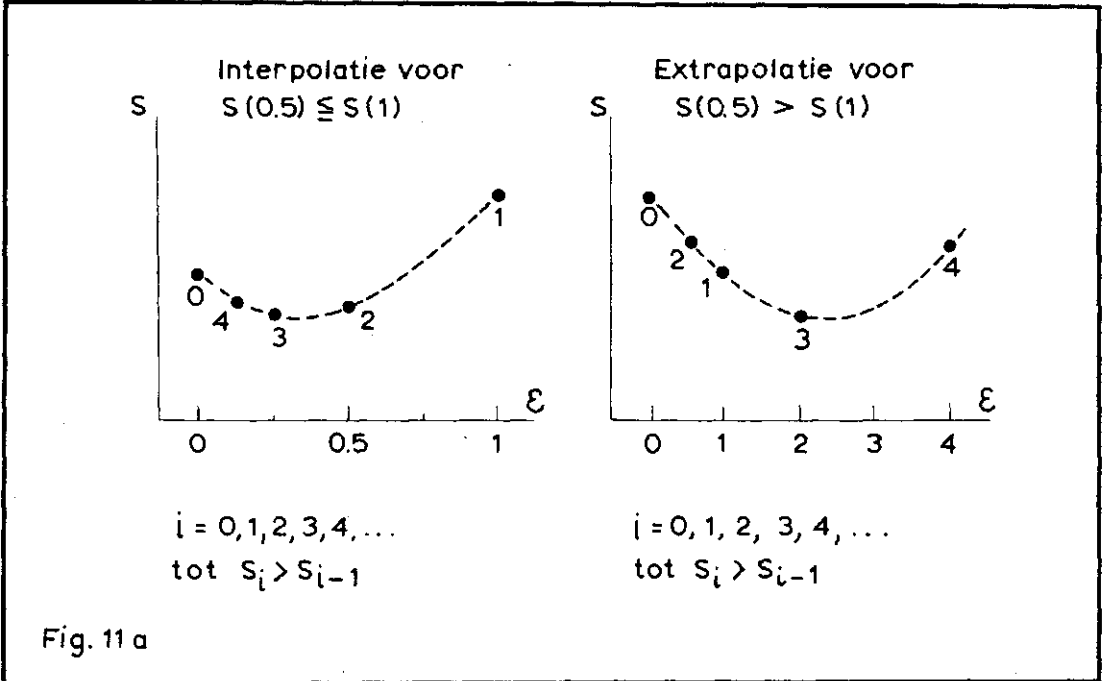
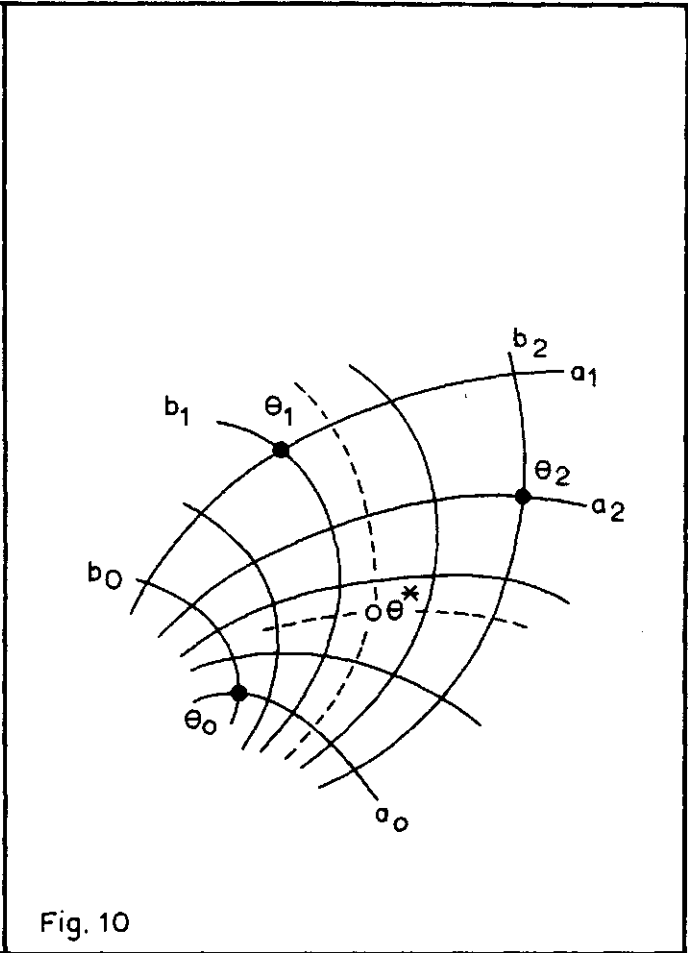
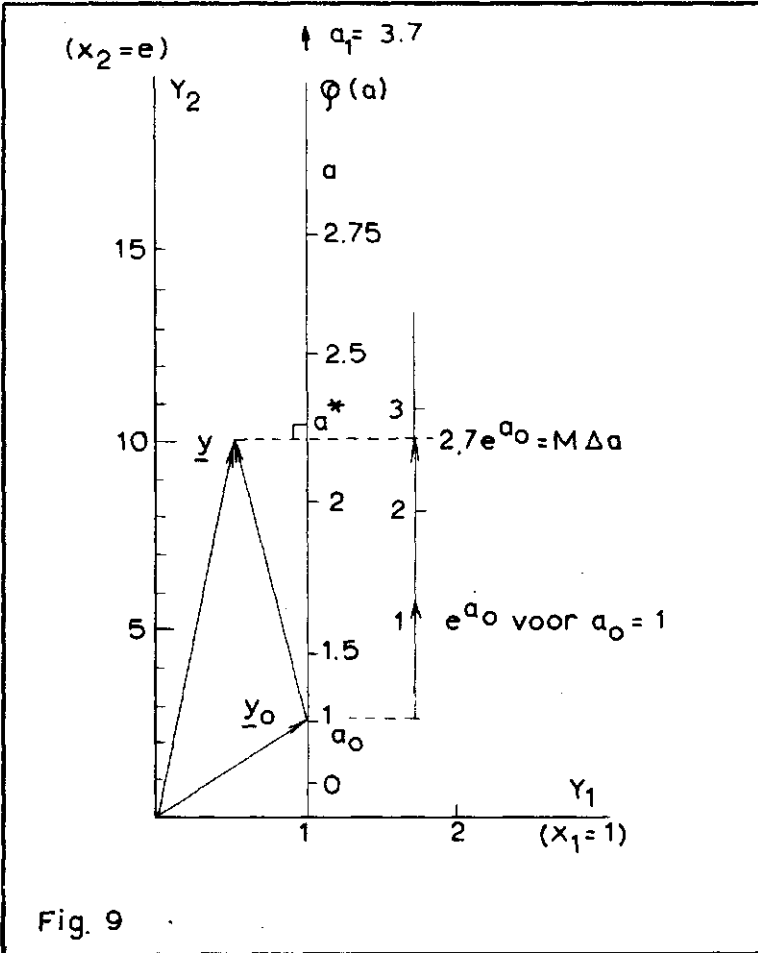
Het invoeren van een nieuwe afwijkende θ_0 zal kunnen dienen een nabijgelegen oplossing op het spoor te komen. Een en ander zal tot moeilijkheden bij de interpretatie leiden indien de oplossingen dicht bij elkaar liggen en fysisch niet te onderscheiden zijn.

Literatuurlijst

- BOOTH, G.W. and PETERSON, T.I (1960), I.B.M. 704 Program WL NLI "Non-Linear Estimation"
- HARTLEY, H.O. (1961) The Modified Gauss - Newton Method for the fitting of non-linear regression functions by least squares, Technometrics, Vol.3, No.2
- KAMIL, L.P. (1962) Lineaire regressie, I.C.W. Nota 134
- STOL, Ph.Th. (1961) Het vereffenen van een formule voor de pF-curve, I.C.W. Nota 113
- (1962) Het gebruik van schaalfactoren als hulpmiddel voor een snelle convergentie bij niet-lineaire vereffening, I.C.W. Nota 139
- TIMMAN, R. (1959) Vectoranalyse, Delft







(Fig. 11 b na fig. 18)

