

Het gebruik van schaalfactoren als hulpmiddel voor een  
snelle convergentie bij niet-lineaire vereffening

PH.TH. Stol

BIBLIOTHEEK DE HAAG

Droevendaalsesteeg 3a

6708 PB Wageningen

1. Inleiding

Bij het vereffenen van functies waarin de parameters die berekend moeten worden in een niet-lineaire vorm voorkomen worden veelal iteratieve bewerkingsmethoden toegepast.

In Nota 138 getiteld: "Een meetkundige beschouwing over de niet-lineaire vereffening" werden de begrippen en werkwijzen uit de niet-lineaire vereffening nader toegelicht. Hierbij werden de iteratieve methode van Gauss-Newton en een gemodificeerde methode (HARTLEY, 1961), welke laatste als het aftasten van het parameteroppervlak werd beschreven, onderling vergeleken.

In deze Nota zal aangegeven worden op welke wijze een snellere convergentie naar de eindoplossing bereikt kan worden door het in rekening brengen van schaalfactoren.

2. Kort overzicht over de vereffeningprocedure

In par. 5 van Nota 138 (pag. 7 e.v.) werd de theoretische afleiding gegeven van de formules die bij de niet-lineaire vereffening gebruikt worden. In het volgende worden, voornamelijk met het doel de in deze Nota te gebruiken symbolen te definiëren, de afgeleide formuleringen kort samengevat.

Stel dat er sprake is van een functie waarin twee parameters voorkomen bijvoorbeeld

$$y = f(x; a, b) \quad (2.1)$$

Een parametervoorstelling van een dergelijke functie voor het geval een drietal x-waarden beschikbaar zijn heeft bijvoorbeeld de gedaante als wordt weergegeven in figuur 1. De figuur is een afbeelding van het driedimensionale vlak.

$$\varphi(a, b) = 0 \quad (2.2)$$



1787063

Figuur 1

De vorm van het vlak  $\varphi(a, b)$  wordt bij gegeven (2.1) bepaald door de drie x-waarden die bijvoorbeeld uit waarnemingen afkomstig kunnen zijn. Op  $\varphi(a, b)$  ligt een stelsel parameterlijnen respectievelijk een stelsel voor constante waarden van b en een stelsel voor constante waarden van a.

Wanneer a en b een paar waarden aanneemt wordt op  $\varphi(a, b)$  een punt  $P(a, b)$  aangewezen waaruit volgt dat de parameterlijnen als kromlijnige coördinaten opgevat kunnen worden.

Wanneer de vector  $\underline{y}$  de voorstelling is van een reeks waarnemingsuitkomsten wordt gevraagd naar die waarden  $a^*$  en  $b^*$  van de parameters a en b die het voetpunt aanwijzen van de loodlijn uit het eindpunt van  $\underline{y}$  op  $\varphi(a, b)$ . Het voetpunt kan aangeduid worden met het punt

$$P^* \stackrel{\text{def}}{=} P(a^*, b^*) \quad (2.3)$$

Dit voetpunt kan met de oorsprong O verbonden gedacht worden waarmee de vector  $\underline{y}^*$  verkregen wordt.

Als uitgangspunt voor het berekenen van het punt  $P^*$  (in fig. 1 niet zichtbaar) wordt een punt  $P_0$  gebruikt dat vermoedelijk een goede benadering van  $P^*$  is en gedefinieerd wordt door het waardenpaar

$$P_0 \stackrel{\text{def}}{=} P(a_0, b_0)$$

Het punt  $P_0$  kan weer met de oorsprong O verbonden gedacht worden, waarmee de vector  $\underline{y}_0$  verkregen is,

Uit deze eerste schatting wordt een  $P_1$  berekend dat dichter bij de oplossing  $P^*$  ligt dan  $P_0$ , waarna de bewerking herhaald wordt tot het punt  $P^*$  zelf bereikt is. Bij de berekening wordt gebruikgemaakt van het raakvlak aan  $\varphi(a, b)$  ter plaatse van  $P_0$ .

Stel dat de loodlijn uit  $\underline{y}$  op  $\varphi(a, b)$  het raakvlak (fig.1) snijdt in Q. Het voetpunt van de loodlijn uit  $\underline{y}$  op het raakvlak wordt voorgesteld door het punt R. Indien het raakvlak ontwikkeld zou worden voor  $P_0 = P^*$  dan zouden alle voetpunten samenvallen met  $P^*$  zodat dan

$$R = Q = P^*$$

Het met R aangeduide voetpunt in het raakvlak kan berekend worden uit een lineaire regressie van  $(\underline{y} - \underline{y}_0)$  op het raakvlak. Een uitdrukking voor R wordt verkregen door de parameterlijnen in  $P_0$  te lineariseren en deze als

coördinaten van het raakvlak te gebruiken, wat in figuur 2 nader staat aangegeven.

Figuur 2

De richtingen van de coördinaatassen zijn dan

$$\frac{\partial \underline{y}_o}{\partial a} \quad \text{en} \quad \frac{\partial \underline{y}_o}{\partial b}$$

waarin de index aangeeft dat de afgeleiden naar  $P_o$  ontwikkeld moeten worden.

Samengevat tot kolomvectoren in een matrix, met indices voor partiële afgeleiden, wordt de basis van het raakvlak

$$M_o \stackrel{\text{def}}{=} (\underline{y}_a, \underline{y}_b)_o$$

In het algemeen zullen  $\underline{y}_a$  en  $\underline{y}_b$  weer van de parameters a en b afhankelijk zijn, zodat  $M_o$  en daarmee het raakvlak, wat richting en coördina-  
tenstelsel betreft, van  $P_o$  afhankelijk is.

De lengte van de vectoren langs de parameterlijnen is dus respectievelijk

$$\left| \frac{\partial \underline{y}}{\partial a} \right| \quad \text{en} \quad \left| \frac{\partial \underline{y}}{\partial b} \right| \quad (2.4)$$

en worden aangeduid met de symbolen  $h_a$  en  $h_b$ .

Indien  $\underline{e}_1$  en  $\underline{e}_2$  eenheidsvectoren in de overeenkomstige richtingen zijn kan geschreven worden

$$\begin{aligned} \underline{y}_a &= h_a \underline{e}_1 \\ \underline{y}_b &= h_b \underline{e}_2 \end{aligned} \quad (2.5)$$

waarin  $h_a$  en  $h_b$  de schaalfactoren genoemd worden (SPIEGEL, 1958 pag.135).

Ook deze schaalfactoren zijn in het algemeen weer afhankelijk van a en b en dus van  $P_o$ . De schaalfactoren geven dus aan hoe groot een schaal-eenheid in de richting van a, respectievelijk van b is.

Gemakshalve wordt nog de matrix van schaalfactoren ingevoerd

$$H = \begin{pmatrix} h_a & 0 \\ 0 & h_b \end{pmatrix} \quad (2.6)$$

Opgemerkt wordt dat de kwadraten van de schaalfactoren voorkomen op de hoofddiagonaal van de symmetrische matrix  ${}^t M_o M_o$ .

De regressiecoëfficiënten worden met  $\Delta a$  en  $\Delta b$  aangeduid en komen in figuur 2 ook alleen als coëfficiënten voor.

Een van de  $P_o$  onafhankelijke voorstelling van de parameterlijnen wordt gegeven in de zogenaamde parameterruimte waarin twee loodrechte assen met gelijke schalen gebruikt worden zoals in figuur 3 wordt weergegeven.

Figuur 3

In deze figuur is het coördinatenstelsel een afbeelding van het parameternet van  $\varphi(a, b)$ . In de parameterruimte kan een waardenpaar  $(a, b)$  als vector opgevat worden zodat bijvoorbeeld

$$\underline{\theta}_o \stackrel{\text{def}}{=} \begin{pmatrix} a_o \\ b_o \end{pmatrix}$$

De regressiecoëfficiënten  $\Delta a$  en  $\Delta b$  hebben tevens de betekenis van correcties (Nota 138 pag, 10) om welke reden de gevolgde methode in de literatuur ook wel als die van de differentiaal-correcties wordt aangeduid.

De correcties kunnen eveneens als een vector voorgesteld worden namelijk

$$\underline{\Delta\theta} = \begin{pmatrix} \Delta a \\ \Delta b \end{pmatrix} \quad (2.7)$$

De berekening van de vector (2.7) volgt uit de normaalvergelijkingen

$${}^t M_o M_o \underline{\Delta\theta} = {}^t M_o (\underline{y} - \underline{y}_o) \quad (2.8)$$

waarvan de oplossing is

$$\underline{\Delta\theta} = ({}^t M_o M_o)^{-1} {}^t M_o (\underline{y} - \underline{y}_o) \quad (2.9)$$

Gebruikte uitgangswaarden van de parameters kunnen met een aantal malen de correctie (2.9) vermeerderd worden waarmee bijvoorbeeld een punt  $P_1(a_1, b_1)$  gevonden wordt dat kan dienen als beginpunt van een volgende berekening.

Algemeen kan voor dit proces geschreven worden

$$\underline{\theta}_{-i+1} = \underline{\theta}_{-i} + \epsilon \underline{\Delta\theta}_{-i} \quad (2.10)$$

waarin  $\Delta \theta_i$  de regressie-coëfficiënt (correctietermen) zijn van de uitgangswaarde  $\theta_i$  berekend volgens (2.9). In (2.10) komt  $\varepsilon$  als parameter voor waarmee tevens is aangegeven dat in de parameterruimte (fig. 3) de vergelijking (2.10), onafhankelijk van de dimensie van  $\theta$ , een rechte voorstelt (EIJL en SALM, 1958, pag. 12 en 19). De richtingsvector van de rechte is  $\Delta \theta_0$  die uit de normaalvergelijkingen berekend wordt, Omgekeerd kan gesteld worden dat met het oplossen van de normaalvergelijkingen nagestreefd wordt een nieuwe richting te bepalen waarin een volgende benadering van  $P^*$  zal liggen.

In figuur 3 tenslotte is de afbeelding van  $P_0$  aangegeven met  $P'_0$  en van  $Q$  met  $Q'$ .

De minimale som van kwadraten van de afwijkingen van de waarnemingen voorgesteld door  $\underline{y}$  ten opzichte van  $\varphi(a, b)$  wordt voorgesteld door de lengte in het kwadraat van de vector  $(\underline{y} - \underline{y}^*)$  doch voor een willekeurig punt  $P_i$  op  $\varphi(a, b)$  met vector  $\underline{y}_i$  geldt voor de som van kwadraten

$$S_i^2 = (\underline{y} - \underline{y}_i) \cdot (\underline{y} - \underline{y}_i) \quad (2.11)$$

Er geldt dus dat  $S^2$  bij een gegeven waarnemingsvector  $\underline{y}$  alleen nog afhankelijk is van de vector  $\underline{y}_i$  en daarmee van het waardenpaar  $(a, b)$  zodat  $S^2 = S^2(a, b)$ .

De meetkundige plaats van het voetpunt van alle vectoren met gelijke lengte en dus gelijke  $S^2$  uit (2.11) zal wanneer  $\varphi(a, b)$  in alle richtingen continu is, een gesloten figuur zijn, Ook deze kan in de parameterruimte afgebeeld worden, doch de vorm wordt nu mede door de richting van en de schaalverandering langs de parameterlijnen bepaald (fig.4).

Figuur 4

Uit de figuur valt af te leiden dat in het algemeen aan  $\varepsilon$  uit (2.10) een waarde gegeven kan worden die  $S^2$  in de richting  $\Delta \theta_0$  kleiner maakt, doch ook dat deze waarde niet noodzakelijkerwijs gelijk aan 1 behoeft te zijn. In figuur 5 worden doorsnijdingen met  $S^2(a, b)$  in verschillende richtingen uit figuur 4 schematisch weergegeven.

Figuur 5

Volledigheidshalve zal nog nagegaan worden in welke richting de sterkste afneming van de kwadraatsom vanaf bijvoorbeeld  $P_0$  zal plaatsvinden. Hiertoe wordt in de eerste plaats opgemerkt dat (2.11) een scalaire functie is zodat ook geschreven kan worden met een  $\Sigma$ -teken, aangevend dat over alle waarnemingen gesommeerd

$$S^2(a,b) = \Sigma (y-y_0)^2 \quad (2.11)a$$

De uitwendig gerichte normaal op deze functie volgt uit de gradiënt van  $S^2$  vblgens (SPIEGEL; 1959, pag.61)

$$\nabla S^2 \quad \text{met } \nabla = \begin{pmatrix} \frac{\partial}{\partial a} \\ \frac{\partial}{\partial b} \end{pmatrix}$$

waardoor de vector verkregen wordt in de richting van toenemende  $S^2$ -waarden met kentallen

$$\begin{pmatrix} -2 \Sigma (y-y_0) \left( \frac{\partial y_0}{\partial a} \right) \\ -2 \Sigma (y-y_0) \left( \frac{\partial y_0}{\partial b} \right) \end{pmatrix}$$

De richting waarin  $S^2$  afneemt is hieraan tegengesteld terwijl bovendien door de factor 2 gedeeld kan worden, De richting van de vector verandert door deze deling niet en er ontstaat, nu weer in vectornotatie

$$\begin{pmatrix} (\underline{y}-\underline{y}_i) \cdot \underline{y}_a \\ (\underline{y}-\underline{y}_i) \cdot \underline{y}_b \end{pmatrix}_0 = t_{M_0} (\underline{y}-\underline{y}_0) \quad (2.12)$$

Hiermee is een uitdrukking gegeven voor de richting van de steilste daling van  $S^2$  vanaf een punt  $P_0$  of vanuit  $P_0'$  uit figuur 4. Het rechterlid van (2.12) is gelijk aan dat van (2.8) en stelt de kolomvector van bekende termen uit de normaalvergelijkingen voor. Deze is dus  $-\frac{1}{2}$  x de normaal op  $S^2(a,b)$ .

De lengte van de normaal  $\underline{n}$  volgt nog uit (2.12) en kan geschreven worden in de vorm

$$|\underline{n}| = 2 \sqrt{\Sigma^2 (y-y_0) (y_a) + \Sigma^2 (y-y_0) (y_b)}$$

De totale differentiaal voor  $S^2$  volgt uit (2.11)a en luidt

$$dS^2 = - 2 \sum (y-y_0) y_a da - 2 \sum (y-y_0) y_b db$$

Met kleine veranderingen van de variabelen, aangeduid door de toevoeging  $\Delta$ , en gebruikmakend van (2.12) kan geschreven worden

$$\Delta S^2 = - 2 {}^t M_0 (\underline{y}-\underline{y}_0) \Delta \underline{\theta} \quad (2.13)$$

Tenslotte geldt nog voor de hoek  $\alpha$  tussen de richting van de parameterlijn a en de verschilvector tussen  $\underline{y}$  en het punt  $P_0$  - dus  $(\underline{y}-\underline{y}_0)$  voorgesteld in figuur 1 -

$$\cos \alpha = \frac{(\underline{y}-\underline{y}_0) \cdot \underline{y}_a}{|\underline{y}-\underline{y}_0| |\underline{y}_a|} \quad (2.14)$$

en in verband met (2.5) en (2.11)

$$= \frac{(\underline{y}-\underline{y}_0) \cdot \underline{y}_a}{S_0 h_a}$$

Analoog kan  $\cos \beta$  berekend worden.

Wordt dus reeds bij het opstellen van de normaalvergelijkingen overgegaan op eenheidsvectoren  $\underline{y}_a$  en  $\underline{y}_b$  en tevens  $(y-y_0) e_j$  dan wordt de vorm van de symmetrische matrix:

$${}^t M_0 M_0 = \begin{pmatrix} 1 & \underline{y}_a \cdot \underline{y}_b \\ \underline{y}_a \cdot \underline{y}_b & 1 \end{pmatrix} \quad (2.15)$$

wat de correlatiematrix is van wat in de regressieberekening (fig. 2) de onafhankelijk variabelen genoemd worden, De hoek tussen de richtingen van de parameterlijnen a en b wordt nu gevonden uit (2.15) namelijk

$$\cos(a, b) = \underline{y}_a \cdot \underline{y}_b$$

Het rechterlid van de normaalvergelijkingen (2.8) wordt nu door de overgang op (2.15) een kolomvector waarvan de kentallen richtingscosinus-  
sen zijn van de hoeken tussen de vector  $(\underline{y}-\underline{y}_0)$  met de achtereenvolgende parameterlijnen.

Uitgedrukt in de oorspronkelijke onbekenden (parameters) ziet het

schema van de normaalvergelijkingen er nu als volgt uit

$$\left( \begin{array}{cc|c} h_a \Delta a & h_b \Delta b & = S_o \\ \hline 1 & \frac{Y_a \cdot Y_b}{h_a \cdot h_b} & \vdots (Y - Y_o)_e \cdot \frac{Y_a}{h_a} \\ \frac{Y_a \cdot Y_b}{h_a \cdot h_b} & 1 & \vdots (Y - Y_o)_e \cdot \frac{Y_b}{h_b} \end{array} \right)$$

Deze notatie houdt in dat gelezen moet worden dat

$$\Delta a + \frac{Y_a \cdot Y_b}{h_a} h_b \Delta b = \frac{S_o}{h_a} (Y - Y_o)_e \cdot \frac{Y_a}{h_a}$$

Indien nu het punt  $P^*$  volgens (2.3) gevonden is zodat gesteld kan worden dat alle afgeleiden nu ontwikkeld worden voor

$$P_o = P^*$$

dan zullen de volgende kenmerken gelden

$$S^2 = \text{minimaal}$$

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

$${}^t M^* (\underline{y} - \underline{y}^*) = \underline{0}$$

De betekenis van deze laatste uitdrukking is dat de normaal op het oppervlak  $S(a,b)$  geen componenten meer heeft in het grondvlak  $(a, b)$ , zoals bijvoorbeeld in figuur 4 in  $Q'$  het geval is, maar ook dat de verschilvector  $(\underline{y} - \underline{y}^*)$  loodrecht op de parameterlijnen staat.

De multipele correlatie-coëfficiënt tenslotte is de cosinus van de hoek tussen de vectoren  $\underline{y}$  en  $\underline{y}^*$  en is dus de gewone correlatie tussen de gemeten en de berekende waarden.

Hiermee is van de belangrijkste rekenkundige grootheden de meetkundige betekenis nog eens toegelicht.

### 3. De invloed van de schaalfactoren

Tussen de uitkomst van de regressieberekening in het raakvlak (2.8) en (2.9) en het nieuwe punt  $P_1(a_1, b_1)$  op  $\varphi(a, b)$  verkregen uit

$$P_1(a_1, b_1) = P(a_o + \varepsilon \Delta a, b_o + \varepsilon \Delta b) \tag{3.1}$$



bestaat geen onmiddellijke meetkundige betrekking. De relatie is die welke bestaat tussen beeld en afbeelding. Extreem gesteld zou gezegd kunnen worden dat na de regressieberekening het nieuwe punt op  $\varphi(a, b)$  zonder meer opgezocht wordt. Het punt  $P_1$  zal dan afhankelijk van de schaalfactoren van de parameterlijnen ergens op  $\varphi(a, b)$  terecht komen. Zouden de schaalfactoren vanuit bijvoorbeeld  $P_0$  toenemen, het parameternet vanuit  $P_0$  als het ware steeds wijder worden, dan zou de correctie  $\Delta a$  respectievelijk  $\Delta b$  een overbeoordeling inhouden en kan  $P_1$  op  $\varphi(a, b)$  een veel "slechter" tweede punt blijken te zijn dan het punt R in het raakvlak (fig, 1) doet vermoeden. Het punt  $P_1$  kan hierdoor verder van  $P^*$  verwijderd liggen - een grotere  $S^2$  hebben - dan het uitgangspunt  $P_0$ . In Nota 138 werd hiervan in figuur 9 en 10 een voorbeeld gegeven; in het volgende zal een en ander nog verder uitgewerkt worden,

Het wijder of nauwer worden van het parameternet komt tot uiting in de verandering van de schaalfactoren. Algemeen zijn deze weer functies van de parameters zelf en dus zal gelden

$$\frac{\partial h_a}{\partial a} = f_1(a, b) \quad (3.2)a$$

$$\frac{\partial h_b}{\partial b} = f_2(a, b) \quad (3.2)b$$

Voor een lineaire vereffening gaat dit over in het bijzondere geval

$$\frac{\partial h_a}{\partial a} = \frac{\partial h_b}{\partial b} = 0$$

zodat dan over het gehele parametervlak dezelfde eenheden blijven gelden.

Voor die gevallen waarin bijvoorbeeld (3.2)a positief is zal het parameternet met toenemende waarden van  $a$  wijder worden en zal de nieuwe waarde  $a_0 + \Delta a$  te groot uitvallen. Te lage waarden zullen ontstaan wanneer (3.2) een negatief teken heeft.

In het volgende zal aangegeven worden op welke wijze de over- respectievelijk onderbeoordeling van differentiaal-correcties op de parameters tengevolge van het schaalverloop met behulp van het in rekening brengen van de schaalfactoren voorkomen kan worden. Het resultaat van het in rekening brengen van de schaalfactoren bestaat uit het sneller convergeren naar de eindoplossing tijdens de vereffeningsprocedure.

4. Voorbeeld van het toepassen van schaalfactoren

Als voorbeeld is gekozen het geval dat

$$y = (x; a, b) = a e^{-bx^2} \quad (4.1)$$

met tabel van waarnemingsuitkomsten

i	$x_i$	$y_i$	
1	0,3	2,50	dus $\underline{y} = \begin{pmatrix} 2,50 \\ 3,80 \\ 1,50 \end{pmatrix}$
2	0,1	3,80	
3	0,5	1,50	

(4.2)

Gevraagd wordt de parameters a en b volgens kleinste kwadraten aan te passen aan (4.2).

-Algemeen geldt voor de kentallen van een vector  $\underline{y}(a, b)$  uit (4.2) en (4.1)

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} ae^{-0,09b} \\ ae^{-0,01b} \\ ae^{-0,25b} \end{pmatrix} \quad (4.3)$$

De parametervoorstelling voor dit oppervlak  $\phi(a, b)$  is een driedimensionale figuur met assen respectievelijk voor  $y_1$ ,  $y_2$  en  $y_3$ ; In figuur 6 is een perspectief tekening van dit oppervlak gegeven, De vector in deze figuur is de waarnemingsvector  $\underline{y}$  uit (4.2). Het eindpunt van  $\underline{y}$  ligt niet op  $\phi(a, b)$ .

Figuur 6

Het voetpunt van de loodlijn uit het eindpunt van  $\underline{y}$  op het oppervlak  $\phi(a, b)$  is door een vierkantje aangegeven.

De matrix van afgeleiden is nu algemeen

$$M_o = \begin{pmatrix} e^{-0,09b} & -0,09a & e^{-0,09b} \\ e^{-0,01b} & -0,01a & e^{-0,01b} \\ e^{-0,25b} & -0,25a & e^{-0,25b} \end{pmatrix} \quad (4.4)$$

waaruit verder volgt

$$h_a = \sqrt{e^{-0,18b} + e^{-0,02b} + e^{-0,50b}} \quad (4.5)$$

en

$$h_b = \sqrt{0,0081 a^2 e^{-0,18b} + \dots + 0,0625 a^2 e^{-0,50b}}$$

Tenslotte geldt nog

$$\frac{\partial h_a}{\partial a} = 0 \quad (\text{voor elke } a \text{ en } b)$$

$$\frac{\partial h_b}{\partial b} = f(a, b)$$
(4.6)

Wanneer in (4.5) de parameter a alle waarden doorloopt bij een constante b worden hiermee op  $\varphi(a, b)$  parameterlijnen voor a verkregen die met a-waarden becijferd kunnen worden, Op overeenkomstige wijze kan een stelsel parameterlijnen voor b verkregen worden, Hierdoor ontstaat een parameternet dat gedeeltelijk ook in figuur 6 is weergegeven.

Uit (4.3) volgt nog dat met  $b = \text{constant}$  een lineaire betrekking ontstaat die geschreven kan worden

$$y = a \begin{pmatrix} e^{-0,09b} \\ e^{-0,01b} \\ e^{-0,25b} \end{pmatrix}, \quad b \text{ constant}$$

De parameterlijnen voor a zijn rechten met een lineaire verdeling, wat uit (4.6) volgt, die echter afhankelijk is van de (constante) waarde die b aanneemt, wat uit (4.5) volgt. Uit dezelfde vergelijkingen valt af te leiden dat de parameterlijnen voor b curven zijn met een niet lineaire verdeling die zowel van a als b afhangt.

Als voorbeeld van de bewerkingsmogelijkheden zullen achtereenvolgens behandeld worden

- a. de iteratieve methode
- b. het aftasten van het oppervlak
- c. het toepassen van schaalfactoren

terwijl tot slot nog vermeld wordt

- d. de steilste-weg-methode

Als beginpunt van de berekeningen zal steeds gebruikt worden (zie fig. 6)

$$P(a_0, b_0) = (3, 10)$$

of identiek hiermee

$$\underline{\theta}_0 = \begin{pmatrix} 3 \\ 10 \end{pmatrix} \quad (4.7)$$

De resultaten van de berekeningen worden in de belangrijkste decimalen, afgerond weergegeven.

De algemene vormen gaan nu over in

$$\underline{y}_0 = \begin{pmatrix} 1,22 \\ 2,71 \\ 0,25 \end{pmatrix} \quad (4.8)$$

de verschilvector wordt

$$(\underline{y} - \underline{y}_0) = \begin{pmatrix} 1,28 \\ 1,09 \\ 1,25 \end{pmatrix} \quad (4.9)$$

De matrix van afgeleiden, ontwikkeld naar  $P_0$  wordt

$$M_0 = \begin{pmatrix} 0,407 & -0,110 \\ 0,905 & -0,027 \\ 0,082 & -0,062 \end{pmatrix} \quad (4.10)$$

waaruit de symmetrische matrix ontstaat

$${}^t M_0 M_0 = \begin{pmatrix} 0,991 & -0,074 \\ -0,074 & 0,017 \end{pmatrix} \quad (4.11)$$

met schaalfactorenmatrix volgens 2.6) en de opmerking daaronder:

$$H_0 = \begin{pmatrix} 0,995 & 0 \\ 0 & 0,129 \end{pmatrix} \quad (4.12)$$

Opgemerkt wordt nog dat H een diagonale matrix is waarvoor eigenschappen gelden die met die van scalaire grootheden overeenkomen.

#### a. De Gauss-Newton of iteratieve methode

Wordt met (2.9) de correctie  $\Delta \underline{\theta}_0$  op de beginschattingen uitgerekend, dan wordt verkregen

$$\Delta \underline{\theta}_0 = \begin{pmatrix} 0,764 \\ -11,478 \end{pmatrix} \text{ waaruit } \underline{\theta}_1 = \begin{pmatrix} 3,76 \\ -1,48 \end{pmatrix}$$

volgens (2.10) met  $\epsilon = 1$ . Dit levert een volgend punt  $P_1 = P(a_1, b_1)$  op.

Voor de beginwaarden  $\begin{pmatrix} 3 \\ 10 \end{pmatrix}$  is het kwadraat van de lengte van de verschilvector  $(\underline{y} - \underline{y}_0)$  ten opzichte van  $\varphi(a, b)$  - vergelijk met figuur 1 - gelijk aan  $S_0^2 = 4,39$ .

Wordt  $P_1$  als punt gebruikt waarvoor de verschilvector wordt berekend, dan is  $S_1^2 = 18,78$  en dus groter dan bij de beginwaarden.

De tweede iteratie stap, uitgaande van  $P_1$  of  $\underline{\theta}_1$  heeft tot resultaat

$$\Delta \underline{\theta}_1 = \begin{pmatrix} -0,252 \\ 2,715 \end{pmatrix} \quad \text{en} \quad \underline{\theta}_2 = \begin{pmatrix} 3,51 \\ 1,23 \end{pmatrix} = \begin{pmatrix} a_2 \\ b_2 \end{pmatrix}$$

Voor het nu gevonden punt  $P_2 = P(a_2, b_2)$  geldt dat  $S_2^2 = 1,688$  en dus nog steeds groter is dan  $S_0^2$ .

In figuur 6 staan op het oppervlak de gekozen  $P_0$  en de, uit de iteraties verkregen,  $P_1$  en  $P_2$  aangegeven.

De curve die deze punten verbindt is het beeld van de lineaire betrekking

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

en

$$\underline{\theta} = \underline{\theta}_1 + \epsilon \Delta \underline{\theta}_1$$

In de figuur valt het dadelijk op dat de eerste iteratie een uitkomst ( $P_1$ ) geeft die slechter is dan de beginwaarde  $P_0$ . De volgende iteratie geeft een verbetering ten opzichte van  $P_1$  maar nog niet ten opzichte van  $P_0$ . Na een aantal iteraties zal het eindpunt van de berekeningen bereikt worden. Samengevat ontstaat het overzicht van tabel 1.

Tabel 1

Enkele uitkomsten van de iteratieve berekeningen

Index i	a	$b_i$	$\Delta a$	$\Delta b$	$S_i^2$
0	3	10	0,76	-11,48	4,39
1	3,76	-1,48	-0,25	2,72	18,78
2	3,51	1,23			1,69
.					.
.					.
*	3,86	4,08			0,05

b. De gemodificeerde Gauss-Newton methode of de methode van behoud van de kleinste kwadraatsom

De in Nota 138 besproken methode van het aftasten van het oppervlak  $\varphi(a, b)$ , de gemodificeerde Gauss-Newton methode (HARTLEY, 1961) is de tweede methode die op het gekozen voorbeeld werd toegepast.

Met de betrekking (2.10) werd  $\epsilon$  bepaald die de som van kwadraten langs de weg (fig. 6) van  $P_0$  naar  $P_1$  zo klein mogelijk maakt. In figuur 7 is de curve I de grafische voorstelling van het verband tussen  $\epsilon$  en  $S^2$ .

Figuur 7

Het resultaat van de berekeningen staat in tabel 2 samengevat, en werd in figuur 6 in beeld gebracht onder bijvermelding van de parameterwaarden voor  $\epsilon$  langs de afgelegde weg in de richting van  $P_1$ .

Tabel 2

Het aftasten van  $\varphi(a, b)$  in de eerste richting

	$\Delta \theta_0$	$\epsilon$					
		0	0,2	0,4	0,6	0,8	1,0
a	0,76	3	3,15	3,31	3,46	3,61	3,76
b	-11,48	10	7,70	5,41	3,11	0,82	-1,48
$S^2$		4,39	2,05	1,07	0,22	2,85	18,78

Uit deze tabel blijkt dat de differentiaal correctie  $0,6 \Delta \theta_0$  een betere uitgangswaarde voor een volgende berekening oplevert dan de waarde  $\epsilon = 1$  die bij de vorige methode is toegepast (tabel 1).

De volgende berekening bestaat weer uit een lineaire regressie, nu in het raakvlak ter plaatse van

$$\underline{\theta} = \underline{\theta}_0 + 0,6 \Delta \underline{\theta}_0 = \begin{pmatrix} 3,46 \\ 3,11 \end{pmatrix}$$

In feite betekent dit dat een nieuwe richting, een nieuwe  $\Delta \underline{\theta}$ , aangeduid met de index 1, bepaald wordt waarin het oppervlak  $\varphi(a, b)$  verder afgetast kan worden zoals met (2.10) en de beschouwing daarna werd aangegeven. In tabel 3 staan de resultaten vermeld.

Tabel 3

Het aftasten van  $\varphi(a,b)$  in de tweede richting

	$\Delta \theta_1$	$\epsilon$						
		0	0,2	0,4	0,6	0,8	1,0	1,2
a	0,393	3,46	3,54	3,62	3,70	3,77	3,85	4,07
b	0,949	3,11	3,30	3,49	3,68	3,87	4,06	4,25
$s^2$		0,220	0,160	0,113	0,080	0,061	0,052	0,096

Nu kan  $\epsilon = 1$  genomen worden, waarmee de kwadraatsom van 0,220 tot 0,052 daalt.

Ook de in tabel 3 berekende weg over  $\varphi(a, b)$  is in figuur 6 uitgezet.

De laatste bewerking, na het vaststellen van een nieuwe richting  $\Delta \theta_2$  heeft tenslotte tot resultaat:

Tabel 4

Het aftasten van  $\varphi(a,b)$  in de derde richting

	$\Delta \theta_2$	$\epsilon$		
		0	0,5	1,0
a	0,0197	3,85	3,86	3,87
b	0,0486	4,06	4,08	4,11
$s^2$		0,0516	0,0516	0,0518

Hiermede is het eindpunt van de berekening praktisch bereikt, Voor het verkrijgen van een eindoplossing die een nog kleinere  $s^2$  oplevert, kan een van de besproken methoden toegepast worden daar de lineaire benadering nu zal voldoen. Voor het gestelde doel, het geven van een numeriek voorbeeld, werd met deze uitkomst volstaan.

Tot slot geldt dus

$$\underline{\theta} = \underline{\theta}_0 + 0,6\Delta\underline{\theta}_0 + \Delta\underline{\theta}_1 + 0,5\Delta\underline{\theta}_2$$

of

$$\underline{\theta} = \begin{pmatrix} 3 \\ 10 \end{pmatrix} + \begin{pmatrix} 0,46 \\ -6,89 \end{pmatrix} + \begin{pmatrix} 0,39 \\ 0,95 \end{pmatrix} + \begin{pmatrix} 0,01 \\ 0,02 \end{pmatrix} = \begin{pmatrix} 3,86 \\ 4,08 \end{pmatrix}$$

In figuur 8 staat de parameterruimte afgebeeld met niveaulijnen voor gelijke waarden van  $s^2$ . De kromme in figuur 6 van  $P_0$  naar  $P_1$  staat hier als  $1 \times$

de vector  $\Delta\theta_0$  voorgesteld als eerste richting zonder correctie op schaalverloop, Ter plaatse van de stip (.) werd de kleinste  $S^2$  bepaald (tabel 2) vanuit welk punt zoals in het bovenstaande werd uiteengezet, een nieuwe richting werd berekend met (2.8) en (2.9). Deze is niet meer in figuur 8 ingetekend.

Figuur 8

c. Het toepassen van schaalfactoren

Uit figuur 6 blijkt uit de langs de curve in de eerste richting aangegeven schaalverdeling dat voor grotere  $\epsilon$  met afnemende waarden van  $b$  de onderlinge afstand van de schaal eenheden van  $b$  steeds groter wordt. Dit komt tot uiting in de omstandigheid dat in de eerste richting gaande bij  $\epsilon = 1$  een goede tweede benadering van  $a$  verkregen wordt maar dat bij deze zelfde waarde van  $\epsilon$  de correctie op de parameter  $b$  nu overbeoordeeld wordt. Numeriek volgt dit uit tabel 2 door te bedenken dat de eindoplossing ongeveer is  $(a^*; b^*) = (3,86; 4,08)$ .

Het bovenomschreven effect van het schaalverloop komt tot uiting in de schaalfactoren die voor verschillende  $(a, b)$ -paren verschillende waarden aannemen. Over de afgelegde weg van  $P_0 \rightarrow P_1$  in figuur 6 zijn de schaalfactoren als aangegeven in tabel 5. Tevens zijn in deze tabel de bijbehorende waarden van de partiële afgeleiden (aangeduid met accenten) volgens (3.2)a opgegeven.

Tabel 5

De schaalfactoren en hun afgeleiden voor verschillende  $(a, b)$ -waarden langs de eerste richting (volgens tabel 2)

	$\epsilon$					
	0	0,2	0,4	0,6	0,8	1,0
a	3	3,15	3,31	3,46	3,61	3,76
b	10	7,70	5,41	3,11	0,82	-1,48
-----	-----	-----	-----	-----	-----	-----
$h_a$	0,995	1,062	1,159	1,312	1,584	2,105
$h_b$	0,129	0,185	0,283	0,463	0,796	1,415
-----	-----	-----	-----	-----	-----	-----
$h'_a$	0	0	0	0	0	0
$h'_b$	-0,016	-0,028	-0,051	-0,096	-0,180	-0,337



De schaalfactoren zijn dus mede afhankelijk van de waarden die a en b aannemen doch tevens blijkt dat de verandering van  $h_a$  bij constante b gelijk aan nul is zodat a lineair verdeeld is langs de parameterlijn  $b = \text{constant}$  zoals reeds in (4.6) algemeen is aangetoond.

Vanuit een gegeven punt kan de verandering van  $h_b$  voorgesteld worden door

$$h(b+\Delta b) = h(b) + \frac{\partial h(b)}{\partial b} \Delta b + \frac{1}{2!} \frac{\partial^2 h(b)}{\partial b^2} \Delta^2 b + \dots$$

Verwaarlozing van hogere afgeleiden heeft tot resultaat dat de verhouding ten opzichte van de schaalfactor van het gegeven punt wordt (met accenten voor partiële afgeleiden):

$$\lambda_b \stackrel{\text{def}}{=} \frac{h_b}{h_b + h'_b \Delta b} \quad (4.13)$$

De verhoudingen  $\lambda_a$  en  $\lambda_b$  worden dus voor het gekozen voorbeeld in  $P_0$  (3.10)

$$\lambda_a = \frac{0,995}{0,995 + h'_a(0,76)} = 1, \quad (\text{alle } h'_a = 0)$$

$$\lambda_b = \frac{0,129}{0,129 + h'_b(-11,48)} = 0,412, \quad (\text{met } h'_b = -0,016)$$

Opgenomen in een matrix wordt gedefinieerd

$$\underline{\lambda} = \begin{pmatrix} \lambda_a & 0 \\ 0 & \lambda_b \end{pmatrix} \quad (4.14)$$

zodat ontstaat

$$\underline{\lambda} = \begin{pmatrix} 1 & 0 \\ 0 & 0,412 \end{pmatrix} \quad (4.14)a$$

Op het schaalverloop wordt nu gecorrigeerd door de verhouding van de schaalfactoren volgens (4.13) in rekening te brengen volgens

$$\underline{\lambda} \Delta \theta_0 = \Delta \theta_0' \quad (4.15)$$

Hiermee wordt dus van de richting  $\Delta \theta_0$  afgeweken en een nieuwe richting  $\Delta \theta_0'$  berekend die een correctie op het schaalverloop inhoudt.

Het oppervlak  $\varphi(a, b)$  kan nu weer afgetast worden volgens

$$\underline{\theta}_1 = \underline{\theta}_0 + \varepsilon \Delta \underline{\theta}_0' \quad (4.16)$$

Voor het gekozen voorbeeld werden (4.13), (4.14) en (4.15)

$$\begin{pmatrix} 1 & 0 \\ 0 & 0,412 \end{pmatrix} \begin{pmatrix} 0,764 \\ -11,478 \end{pmatrix} = \begin{pmatrix} 0,764 \\ -4,729 \end{pmatrix}$$

De volgende berekening bestaat weer uit het aftasten van  $\varphi(a, b)$  in de nieuwe richting, samengevat in tabel 6.

Tabel 6

Het aftasten van  $\varphi(a, b)$  in de eerste richting met correctie op schaalverloop

		$\varepsilon$							
		0	0,2	0,4	0,6	0,8	1,0	1,2	1,3
a	0,76	3	3,15	3,31	3,46	3,61	3,76	3,92	3,99
b	-4,73	10	9,05	8,11	7,16	6,22	5,28	4,32	3,85
$S^2$		4,39	3,45	2,51	1,65	0,90	0,326	0,055	0,106

Het beste resultaat wordt nu bereikt bij  $\varepsilon = 1,2$ . De bereikte kwadraat-som van 0,055 is kleiner dan in de eerste richting zonder correctie (tabel 2) die toen de waarde 0,22 had.

In figuur 8 staat deze eerste richting eveneens aangegeven, hierin komt tot uiting dat een richting verkregen is die als het ware dieper door het dal gaat dan wanneer geen correctie op schaalverloop wordt toegepast. In figuur 7 staat de verticale doorsnijing van  $S^2(a, b)$  weergegeven als curve II., terwijl tenslotte in figuur 9 de in de nieuwe richting afgelegde weg op  $\varphi(a, b)$  staat ingetekend.

Figuur 9

Met de waarde  $\varepsilon = 1,2$  uit tabel 6 wordt een nieuw uitgangspunt bepaald en een nieuwe regressie berekening uitgevoerd. Hiermee wordt dus een nieuwe richting bepaald waarin  $\varphi(a, b)$  verder afgetast kan worden. Deze richting is

$$\Delta \underline{\theta}_1 = \begin{pmatrix} -0,043 \\ -0,199 \end{pmatrix}$$

In het punt  $(a;b) = (3,92; 4,32)$  uit tabel 6 is  $h_b = 0,412$  en  $h_b' = -0,080$  waaruit volgt volgens (4.14)

$$\lambda = \begin{pmatrix} 1 & 0 \\ 0 & 0,963 \end{pmatrix}$$

waarna (4.15) wordt

$$\Delta\theta_1' = \begin{pmatrix} -0,043 \\ -0,192 \end{pmatrix}$$

Het resultaat, werkend in de laatste richting, wordt nu als volgt

Tabel 7

Het aftasten van  $\varphi(a,b)$  in de tweede richting met correctie op schaalverloop

	$\Delta\theta_1'$	$\varepsilon$			
		0	1,0	1,2	1,3
a	-0,04	3,92	3,88	3,87	3,86
b	-0,19	4,32	4,13	4,09	4,07
$s^2$		0,0553	0,0505	0,0504	0,0511

Deze oplossing geeft reeds na 2 bewerkingen een waardenpaar voor  $(a, b)$  met een nog kleinere kwadraatsom dan in par. 4b met 3 bewerkingen (0,0516) werd bereikt. Een samenvatting overeenkomstig aan het einde van par. 4b werd gegeven wordt nu

$$\underline{\theta} = \underline{\theta}_0 + 1,2 \Delta\underline{\theta}'_0 + 1,2 \Delta\underline{\theta}'_1$$

of

$$\underline{a} = \begin{pmatrix} 3 \\ 10 \end{pmatrix} + \begin{pmatrix} 0,92 \\ -5,68 \end{pmatrix} + \begin{pmatrix} -0,05 \\ -0,23 \end{pmatrix} = \begin{pmatrix} 3,87 \\ 4,09 \end{pmatrix}$$

De bovenbeschreven methode heeft, wat het uit te voeren rekenwerk betreft, het voordeel dat de verhouding  $\lambda$  uit (4.13) berekend kan worden voordat tot het berekenen van een reeks kwadraatsommen wordt overgegaan. Dit houdt verband met het feit dat  $h_b'$ , verkregen uit (4.5) een eenvoudige betrekking oplevert,

Indien dit niet het geval is, kan de verhouding  $\lambda$  bijvoorbeeld bepaald worden uit de verhouding van een tweetal waarden van de schaalfac-

toren bijvoorbeeld

$$\lambda = H_0 H_1^{-1} \quad (4.17)$$

In het gestelde voorbeeld werd bij het minimum in de eerste richting (tabel 2) gevonden dat  $\varepsilon = 0,6$  met schaalfactoren (tabel 5)

$$h_a = 1,312$$

$$h_b = 0,463$$

zodat met

$$\lambda = H_0 H_{0,6}^{-1}$$

verkregen wordt

$$\lambda = \begin{pmatrix} 0,995 & 0 \\ 0 & 0,129 \end{pmatrix} \begin{pmatrix} 0,762 & 0 \\ 0 & 2,160 \end{pmatrix} = \begin{pmatrix} 0,758 & 0 \\ 0 & 0,279 \end{pmatrix}$$

wat evenredig is met

$$\lambda = \begin{pmatrix} 1 & 0 \\ 0 & 0,368 \end{pmatrix}$$

Deze uitkomst komt overeen met de berekening volgens (4.14)a. Ter besparing van rekenwerk zou bijvoorbeeld  $\lambda$  steeds berekend kunnen worden voor  $\varepsilon = 0$  en  $\varepsilon = 0,5$ , zonder  $S^2$  te berekenen, waarna  $\varphi(a, b)$  afgetast kan worden in de richting

$$\Delta \underline{\theta}'_0 = H_0 H_{0,5}^{-1} \Delta \underline{\theta}_0$$

en dus volgens de rechte in de parameterruimte

$$\underline{\theta} = \underline{\theta}_0 + \varepsilon H_0 H_{0,5}^{-1} \Delta \underline{\theta}_0 \quad (4.18)$$

Het verkrijgen van een goed resultaat is echter afhankelijk van de vraag of voor  $\varepsilon = 0,5$  de afstand over  $\varphi(a, b)$  niet te groot is.

#### d. De methode van de steilste weg

Volledigheidshalve wordt hier nog de methode van de steilste weg kort toegelicht.

Volgens (2.12) is de richting waarin  $S^2$  vanaf  $P_0$  het sterkst afneemt

$${}^t M_0(\underline{y} - \underline{y}_0) = \begin{pmatrix} 1,61 \\ -0,25 \end{pmatrix}, \quad \text{voor} \quad \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 3 \\ 10 \end{pmatrix}$$

De lengte van deze vector is

$$\sqrt{(1,61)^2 + (0,25)^2} = 1,63$$

In figuur 8 staat deze richting eveneens ingetekend. In figuur 10 staan nog de doorsnijdingen door  $S^2(a, b)$  ingetekend op dezelfde eenheden als waarmee figuur 8 is vervaardigd. De curve voor de steilste weg is met III aangeduid.

Figuur 10

Zoals uit de figuren 8 en 10 blijkt geeft de steilste - weg - methode weliswaar de sterkste afname van  $S^2$  per eenheid van afgelegde weg doch het laagste punt dat bereikt wordt ligt niet in de richting van  $P^*$ ; evenmin is de totale afname van  $S^2$  in deze richting het grootst. De convergentie zal met deze methode veelal minder snel verlopen dan met de andere besproken methoden het geval is (HARTLEY, 1961). Een uitzondering geldt wellicht voor het geval de steilste-weg richting samenvalt met de as van de niveau-curven voor  $S^2$ , indien deze as tenminste bij benadering een rechte is.

### 5. Slotbeschouwing

In het voorgaande werden een aantal methoden voor het bepalen van de parameters in een gegeven functie besproken, De snelheid waarmee de oplossing bereikt wordt of met andere woorden, het aantal iteraties dat uitgevoerd moet worden kan voor elk van de methoden verschillend zijn. Als ervaringsfeit wordt aangenomen dat de volgorde in methoden van langzaam naar snel convergeren meestal zal zijn (HARTLEY, 1961)

1. steilste-weg methode
2. iteratieve Gauss-Newton methode
3. gemodificeerde Gauss-Newton methode

Een mogelijkheid een te grote correctie op de uitgangswaarden van de parameters te dempen in afhankelijkheid van elk van de parameters afzonderlijk wordt besproken door LEVENBERG (1944). Bij de door hem gevolgde methode wordt ook het gebruik van de diagonaallementen van de matrix van normaalvergelijkingen aanbevolen. Uit de theoretische afleiding volgt echter niet noodzakelijkerwijs dat deze diagonaallementen gebruikt moeten worden. Het voordeel van zijn methode kan liggen in het feit dat de correcties

reeds in de normaalvergelijkingen aangebracht worden.

Een vergelijking van deze methode met de in deze Nota beschreven methode van correcties door het toepassen van schaalfactoren kan nog een afzonderlijk onderwerp van beschouwing uitmaken.

#### Literatuuroverzicht

BIJL J. en H.J.H.SALET, (1958) - Analytische meetkunde I. Delft(I.C.W.11/99)

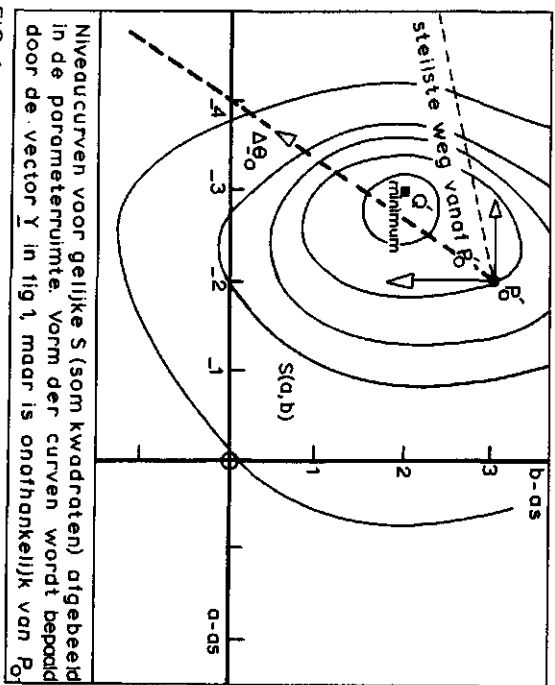
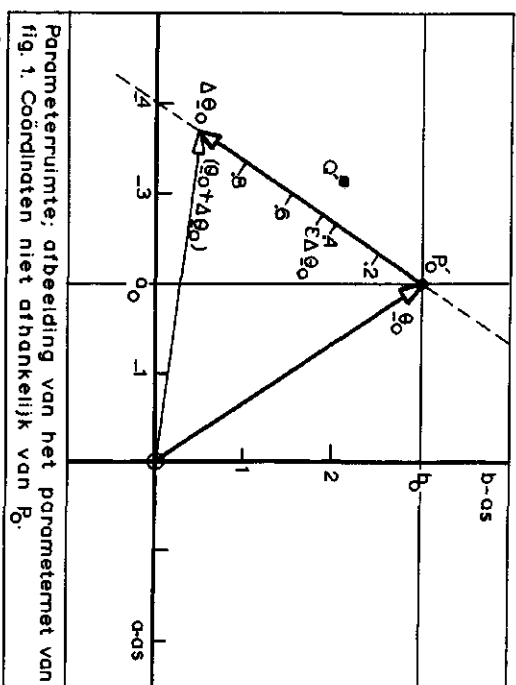
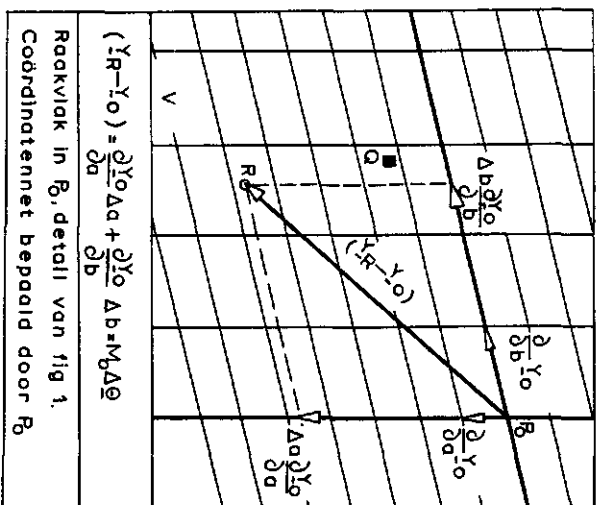
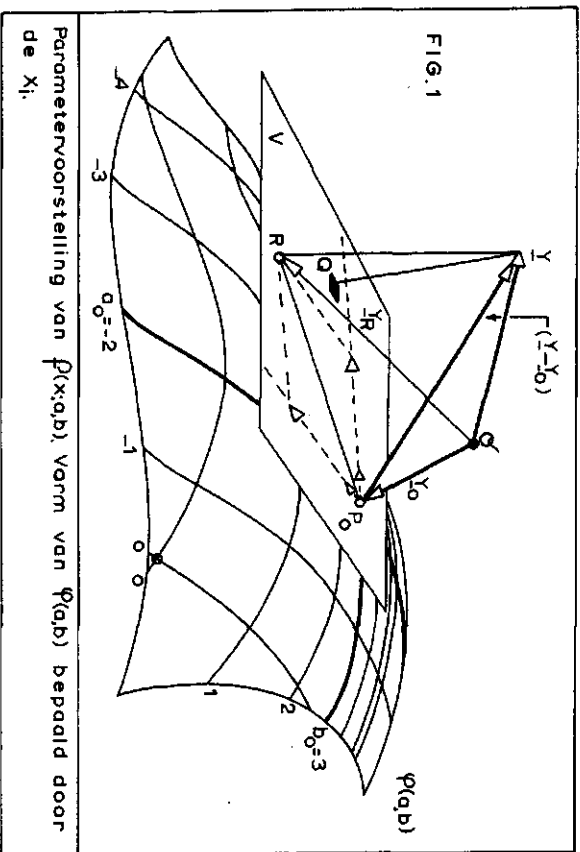
HARTLEY H.O., (1961) - The modified Gauss-Newton method for the fitting of non-linear regression functions by least squares. Technometrics, Vol. 3 No. 2

LEVENBERG K., (1944) - A method for the solution of certain non-linear problems in least squares. Quarterly of Applied Math. Vol. II No. 2

SPIEGEL M.R., (1959) - Theory and problems of vector analysis and an introduction to ten analysis. New York (I.C.W. 11/124)

STOL Ph.Th., (1962) - Een meetkundige beschouwing over de niet-lineaire vereffening. I.C.W.Nota 138

TIMMAN R., (1959) - Vectoranalyse, Delft (I.C.W. 11/93)



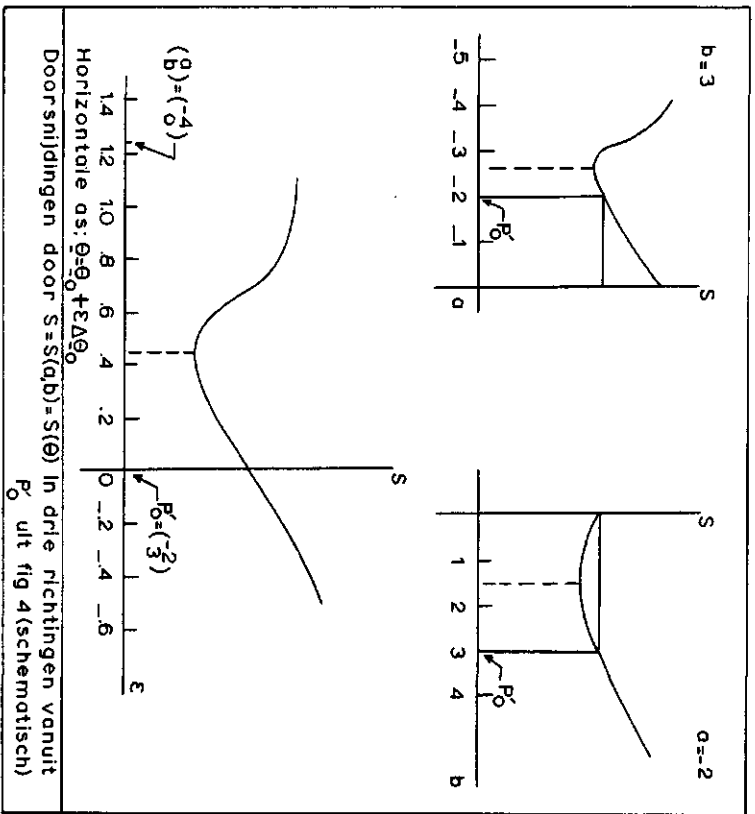
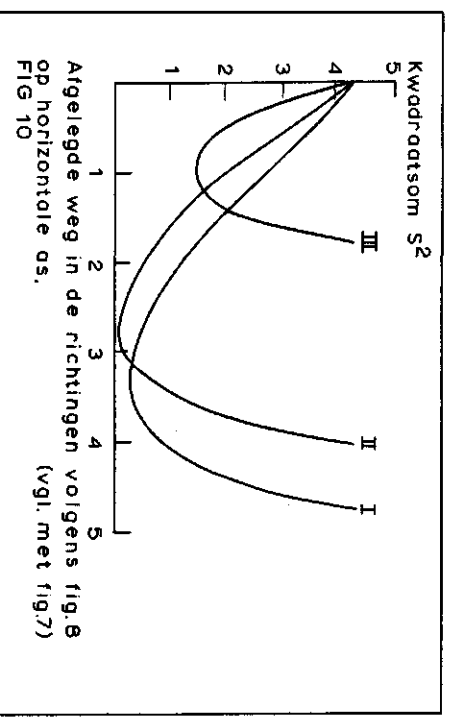
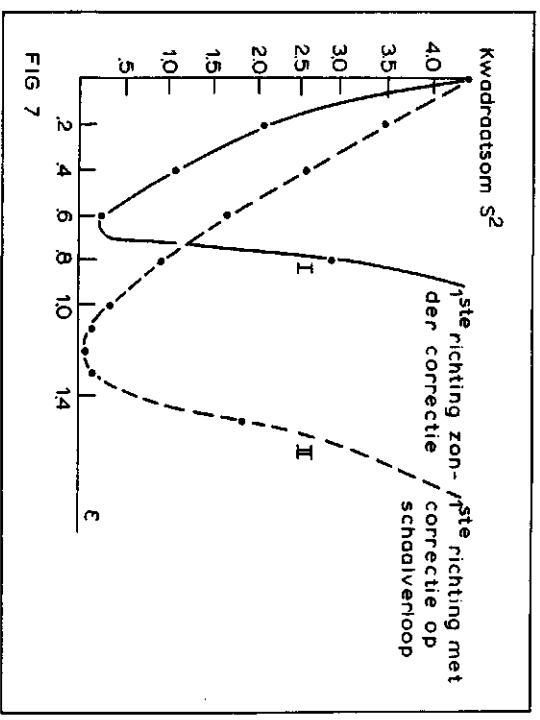


FIG. 5





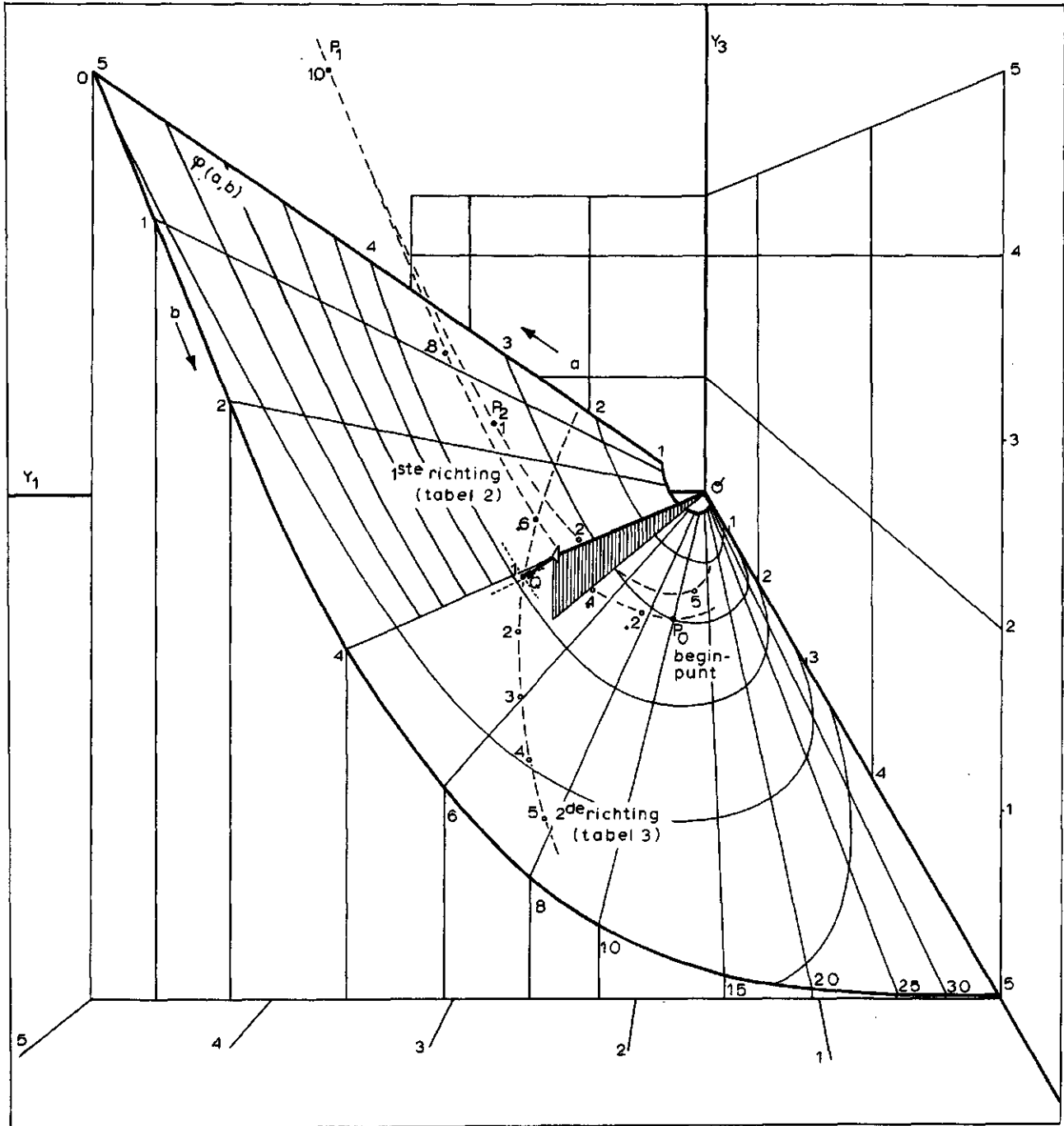


FIG. 6

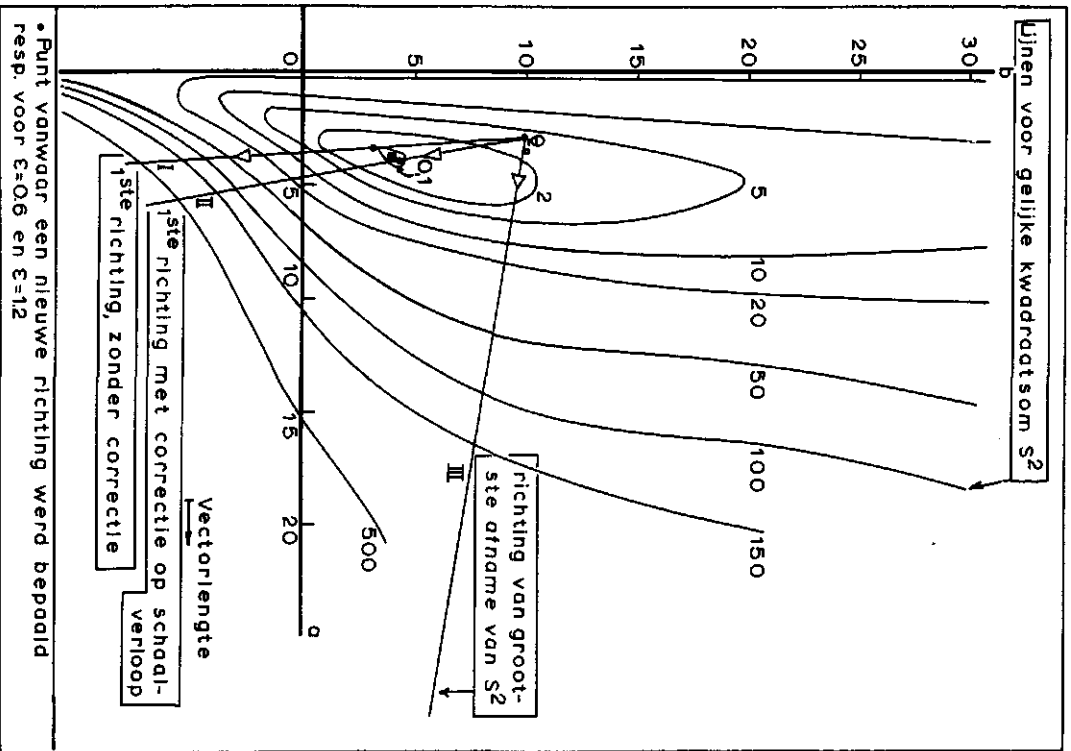


FIG 8

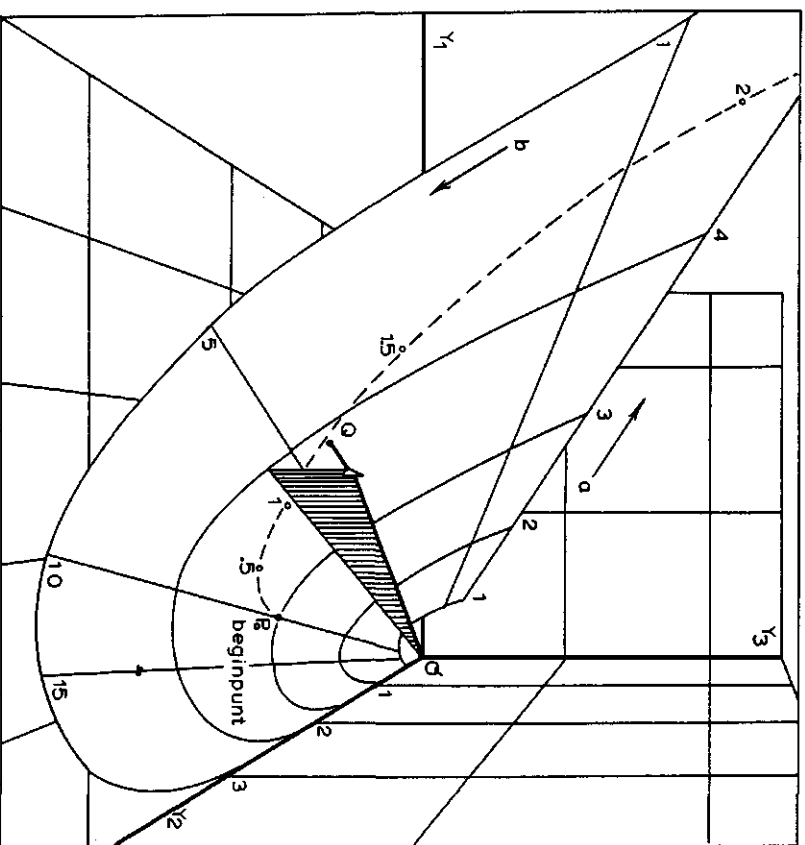


FIG 9