TECHNISCHE HOGESCHOOL EINDHOVEN
Afdeling Algemene Wetenschappen
Onderafdeling der Wiskunde

NUMERIEKE ALGORITHMEN

voor

NIET-LINEAIRE OPTIMALISERINGSPROBLEMEN

door

Dr. Ir. J. L. de Jong

Voorjaarssemester 1976

Eindhoven University of Technology

Department of Mathematics

Numerieke algorithmen voor niet-lineaire optimaliseringsproblemen

door

J.L. de Jong

TECHNISCHE HOGESCHOOL EINDHOVEN

Onderafdeling der Wiskunde

Numerieke algorithmen voor niet-lineaire optimaliseringsproblemen

door

J.L. de Jong

Syllabus bij het College Capita Optimaliseringsmethoden

Voorjaarssemester 1976

Inhoudsopgave

Lijst van veelvuldig gebruikte symbolen

a_i : normaal vector in $\mathbb{R}^n$ corresponderend met i-de (lineaire) beperking

A := $[a_1, a_2, \ldots, a_m]$ n×m-matrix met als kolommen de normalen van de (lineaire) beperkingen

$B^{(k)}$:= k-de benadering van de Hessiaan $G(x^{(k)})$ van de object functie

$B(x,r)$: barrièrefunctie met parameter r

$c(x)$: m-vectorfunctie van gelijkheidsbeperkingen

$d^{(k)}$: zoekrichting in de (k+1)-e iteratie

$f(x)$: objectfunctie ($f: \mathbb{R}^n \to \mathbb{R}^1$)

$g(x)$: m-vectorfunctie van ongelijkheidsbeperkingen

$g^{(k)} := \nabla f(x^{(k)}) := \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n}\right)^T$: gradiënt. (kolom!)vector

$G(x^{(k)}) := \left[\frac{\partial^2 f}{\partial x_i \partial x_1}(x^{(k)})\right]$: Hessiaan van de objectfunctie

$h(x)$: vectorfunctie van gelijkheidsbeperkingen

$H^{(k)}$: k-de benadering van de inverse van de Hessiaan $[G(x^{(k)})]^{-1}$ van de objectfunctie

$J^{(k)} := J(x^{(k)})$: functionaalmatrix of Jacobiaan in $x^{(k)}$ van de vectorfunctie $f(x)$ (bij kleinste kwadraten problemen)

$L(x,\lambda)$ (of $\mathcal{L}(x,\lambda)$): Lagrange functie

$n_i^{(k)}$: i-de kolom van de matrix $N^{(k)}$

$N^{(k)}$: matrix met als kolommen de lineair onafhankelijke normalen van de actieve beperkingen in $x^{(k)}$

$P(x,r)$: boetefunctie met parameter r

Q (of A) : symmetrische n×n-matrix waardoor kwadratische vorm wordt gedefinieerd

$Q(x,\lambda,\rho)$: aangevulde Lagrangefunctie

$s^{(k)} := x^{(k+1)} - x^{(k)}$: stap in de (k+1)-de iteratie

$x^{(k)}$:= k-de benadering voor het optimale punt

$\hat{x}$ (of x^*): optimale punt

$y^{(k)} := g^{(k+1)} - g^{(k)}$: verschil tussen de gradienten van de objectfunctie in $x^{(k+1)}$ en $x^{(k)}$

$z_i^{(k)}$: i-de kolom van de matrix $Z^{(k)}$

$Z^{(k)}$: matrix met als kolommen de basisvectoren van het orthogonale complement van de deelruimte opgespannen door de kolommen van de matrix $N^{(k)}$

$\alpha^{(k)}$: stapgrootte-factor in de (k+1)-de iteratie stap

$\lambda^{(k)}$: k-de benadering van de Lagrange multiplicatorenvector $\hat{\lambda}$

$\hat{\lambda}$ (of: λ^*) : vector van Lagrange multiplicatoren in het optimale punt

1. INLEIDING

1.1. Algemene opmerkingen.

Probleemformuleringen

1.1.1. Deze syllabus heeft tot onderwerp numerieke methoden voor het bepalen van de oplossing van niet-lineaire optimaliseringsproblemen. De aandacht richt zich daarbij in het bijzonder op de volgende drie probleem typen:

UMP: Niet-lineaire programmeringsproblemen zonder nevenvoorwaarden
(= Unconstrained Minimization Problems).

CMP: Niet-lineaire programmeringsproblemen met nevenvoorwaarden
(= Constrained Minimization Problems).

OCP: Optimale-besturingsproblemen
(= Optimal Control Problems).

Voor deze problemen wordt daarbij uitgegaan van de volgende standaard-probleem-formuleringen:

UMP:

$$\min\{f(x) \mid x \in \mathbb{R}^n\} \tag{1.1.1}$$

CMP:

$$\min\{f(x) \mid g_i(x) \geq 0,\ i \in I, h_j(x) = 0, j \in E,\ x \in \mathbb{R}^n\} \tag{1.1.2}$$

OCP:

$$\min\left\{\int_{t_b}^{t_f} \ell(x,u)dt \,\middle|\, \dot{x} = f(x,u),\ x(t_b) = x_b, \psi(x(t_f)) = 0,\ u \in U \subset \mathcal{L}_2^m[t_b,t_f]\right\} \tag{1.1.3}$$

1.1.2. Problemen van het type UMP (vgl (1.1.1)) komen niet alleen vaak voor in de praktijk, zij worden ook gegenereerd bij het zoeken van oplossingen van problemen van het type CMP (vgl (1.1.2)) met behulp van boete- (of penalty-) functie-methoden.
Het idee achter deze methoden is dat het niet voldoen aan de beperkingen "bestraft" wordt met een boete die opgeteld wordt bij de objectfunctie.
In plaats van het niet-lineair programmeringsprobleem met nevenvoorwaarden (1.1.2) beschouwt men dan voor successievelijk kleinere waarden van de iteratieparameter r een serie minimaliseringsproblemen van de vorm

$$\min\{P(x,r) \mid x \in \mathbb{R}^n\} \tag{1.1.4}$$

waarin, bijvoorbeeld,

$$P(x,r) = f(x) + r \sum_{i\in I_1} \{-\ell n[g_i(x)]\} + \frac{1}{r} \sum_{i\in I_2} \{\min[0,g_i(x)]\}^2 + \frac{1}{r} \sum_{j\in E} \{h_j(x)\}^2 \qquad (1.1.5)$$

De objectfuncties van deze nieuwe problemen bestaan telkens uit de oude objectfunctie plus een of meerdere (in dit geval logarithmische) barrière (of barrier-) functies en een of meerdere (in dit geval kwadratische) verlies- (of loss-) functies. Deze nieuwe minimaliseringsproblemen kennen geen beperkingen meer en voor de oplossing ervan kan men gebruik maken van een van de vele numerieke methoden voor de oplossing van problemen van het type UMP. Een meer gedetailleerde bespreking van deze boetefunctiemethoden vormt een onderdeel van hoofdstuk 3.

1.1.3. Een speciale en bijzonder belangrijke categorie van minimaliseringsproblemen van het type UMP wordt gevormd door de zogenaamde niet-lineaire kleinste-kwadraten problemen . Deze, in de praktijk meest veelvuldig van alle minimaliseringsproblemen voorkomende, problemen worden onder andere gegenereerd indien gezocht wordt naar een beste aanpassing in de zin van de kleinste kwadraten (minimum Euclidische norm) bij curve-fitting en niet-lineaire regressie problemen. Zij hebben dan veelal de vorm

$$\min\{ \sum_{t=1}^{\bar{m}} \{[y_t - m_t(x)]/\sigma_t\}^2 \,|\, x \in \mathbf{R}^n\} \qquad (1.1.6)$$

De problemen uit deze categorie, waartoe ook behoren de problemen gegenereerd door het toepassen van kleinste-kwadraten procedures voor het oplossen van stelsels niet-lineaire vergelijkingen, worden gekenmerkt door een speciale structuur die het mogelijk maakt speciale numerieke methoden toe te passen die vaak sneller zijn dan de meer algemene methoden. Meer details hierover worden gegeven in een speciale sectie in hoofdstuk 2.

1.1.4. Tussen de problemen van het type UMP en CMP enerzijds en de problemen van het type OCP bestaan een aantal fundamentele verschillen. De reden om in

deze syllabus desondanks aandacht te besteden aan problemen van het type OCP (optimale-besturings-problemen) komt voort uit de omstandigheid dat de numerieke oplossing van deze problemen kan worden gevonden met behulp van methoden die een sterke overeenkomst vertonen met de methoden voor het bepalen van de problemen van het type UMP en CMP. Deze numerieke methoden, die op zichzelf slechts een fractie uitmaken van de bekende numerieke methoden voor de oplossing van optimale-besturingsproblemen, vormen het hoofdbestanddeel van hoofdstuk 4.

Methoden:

1.1.5. Voor het numeriek oplossen van optimaliseringsproblemen onderscheidt men twee klassen van methoden: indirecte en directe methoden. Bij de indirecte methoden zoekt men naar punten die aan de noodzakelijke voorwaarden voor een optimum (bijvoorbeeld Kuhn-Tucker-condities) voldoen. Bij de directe methoden tracht men een rij benaderingsoplossingen te genereren, die (hopelijk) convergeren naar het optimum. In de praktijk is men bij niet-lineaire optimaliseringsproblemen in de meeste gevallen aangewezen op het gebruik van directe methoden. Het zijn nagenoeg uitsluitend deze laatste methoden die in deze syllabus ter sprake zullen komen.

1.1.6. De meeste directe methoden voor het (bij benadering) bepalen van de oplossing van niet-lineaire optimaliseringsproblemen maken gebruik van iteratieve algorithmen die onder andere de volgende, verbaal geformuleerde stappen omvatten

(0) kies een startpunt $\bar{x}^{(0)}$ en zet k := 0;

(i) bepaal met $\bar{x}^{(k)}$ als uitgangspunt een toegelaten punt $x^{(k)}$ (dit is een punt dat voldoet aan de nevenvoorwaarden);

(ii) evalueer de objectfunctie in het punt $x^{(k)}$;

(iii) test of het punt $x^{(k)}$ optimaal is; zo ja, dan klaar; zo nee, dan:

(iv) genereer een (waarschijnlijk) beter punt $\bar{x}^{(k+1)}$, zet k := k + 1 en ga terug naar stap (i);

De hiergenoemde vijf stappen representeren de vijf belangrijkste facetten van minimaliseringsalgorithmen die het succes van deze in de praktijk in grote mate bepalen. Deze facetten zijn

(0) de keuze van een startpunt

(i) de generatie van een toegelaten punt

(ii) de functie-evaluatie

(iii) het optimaliseringscriterium

(iv) de verbetermethode.

Van deze facetten worden de eerste en de derde voornamelijk bepaald door het betreffende praktijkprobleem. In algemene zin kan hier over weinig worden vermeld. Met betrekking tot het vierde facet, de optimaliteits-test kan worden opgemerkt dat hiervoor in de meeste gevallen gebruik wordt gemaakt van combinaties van theoretische en practische noodzakelijke voorwaarden voor optimaliteit. Welke combinatie bij een practisch probleem wordt gehanteerd hangt zowel af van de aard van het probleem als van de instelling van de onderzoeker. De twee resterende facetten, de generatie van een toegelaten punt en de verbeter methode, worden weliswaar eveneens mede bepaald door de aard van het probleem, hierover valt echter ook in algemene termen veel meer te zeggen. Het grootste deel van deze syllabus is dan ook in het bijzonder aan deze twee onderling sterk samenhangende facetten van de methoden gewijd.

1.1.7. De iteratieve algorithmen voor niet-lineaire optimaliseringsproblemen die in deze syllabus aan de orde komen kunnen worden opgevat als voorschriften voor het genereren van een rij van benaderingsoplossingen. Drie aspecten van deze algorithmen treden daarbij op de voorgrond:

a) het concept waarop de algorithme is gebaseerd
b) de convergentie van de gegenereerde rij benaderingsoplossingen, en
c) de snelheid van de convergentie.

In deze syllabus zal de meeste aandacht worden besteed aan het eerstgenoemde aspect van de optimaliseringsalgorithmen. Waar mogelijk zullen echter ook de beide andere aspecten in de beschouwingen worden betrokken.

Notatie.

1.1.8. Een lijst van de belangrijkste in deze syllabus gebruikte notatie werd gegeven vóór deze Inleiding. Deze notatie komt vrijwel geheel overeen met de notatie die werd gepropageerd op de in 1971 en in 1974 in Teddington gehouden IMA/MPL conferenties. Een van de belangrijkste afwijkingen van deze (in de boeken van Murray [1.1.3] en van Gill en Murray [1.1.4] weergegeven) notatie is het gebruik van de letter d voor de zoekrichting inplaats van de letter p. De overige afwijkingen betreffen minder belangrijke verschillen zoals gebruik van een kleine letter f voor de objectfunctie inplaats van een hoofdletter F.

Referenties.

1.1.9. In deze syllabus zal regelmatig worden aangegeven waar besproken resultaten in de literatuur terug te vinden, respectievelijk uitgebreider behandeld zijn. Deze verwijzingen worden genoteerd met behulp van een paar getallen in vierkante haken (bijvoorbeeld [1.1.3]) welk paar correspondeert met de betreffende literatuur ingang (bijvoorbeeld 1.1.3) in de lijst van referenties die wordt gegeven aan het einde van iedere sectie.

1.1.10. De voornaamste referenties voor de in deze syllabus behandelde onderwerpen zijn de eerste 5 boeken in de volgende lijst van algemene referenties. De twee laatste boeken zijn relatief goedkope uitgaven die interessante en up-to-date introducties in het vakgebied representeren.

Algemene referenties.

[1.1.1]: Luenberger, D.G.: "Introduction to linear and nonlinear programming" Addison-Wesley Publ. Co., Reading, Mass. (1973).

[1.1.2]: Jacoby, S.L.S., Kowalik, J.S. and Pizzo, J.T.: "Iterative methods for nonlinear optimization problems", Prentice Hall Inc., Englewood Cliffs, N.J.(1972).

[1.1.3]: Murray, W. (Ed.): "Numerical methods for unconstrained optimization", Academic Press, London. (1972).

[1.1.4]: Gill, P.E. and Murray, W. (Eds.): "Numerical methods for constrained optimization", Academic Press, London (1974).

[1.1.5]: Bryson, Jr. A.E. and Ho, Y.C.: "Applied optimal control", Blaisdell Publ.Cy., Waltham, Mass. (1969).

[1.1.6]: Walsh, G.R.: "Methods of optimization", Wiley, London, (1975). (Prijs (paperback): ± fl. 27,-).

[1.1.7]: Adby, P.R. and Dempster, M.A.H.: "Introduction to optimization methods". Chapman & Hall, London (1974) (Prijs ± fl. 14,-).

2. METHODEN VOOR MINIMALISERING ZONDER NEVENWOORWAARDEN

§ 2.1. Algemeen: Noodzakelijke en voldoende voorwaarden, convergentietheorie.

2.1.1. In dit hoofdstuk komen aan de orde methoden voor de bepaling van de oplossing van problemen van het type UMP (zie paragraaf 1.1., verg (1.1.1))

$$\min\{f(x) \mid x \in \mathbb{R}^n\} \tag{2.1.1}$$

in welke probleem formulering f een reëelwaardige functie voorstelt ($f : \mathbb{R}^n \to \mathbb{R}^1$) Van belang voor een aantal van de methoden zijn de eerste en de tweede afgeleide (vector-, resp. matrix-)functies van f, t.w. de gradiënt van f in x

$$g := \nabla f(x) := \nabla_x f(x) := \frac{df}{dx}^T := \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n}\right)^T \tag{2.1.2}$$

en de Hessiaan van f in x

$$G(x) := \nabla^2 f(x) := \nabla^2_{xx} f(x) := \frac{d^2 f}{dx^2} := \left[\left(\frac{\partial^2 f}{\partial x_i \partial x_j}\right)\right] \tag{2.1.3}$$

T.a.v. de gradiënt wordt in deze syllabus de notatie afspraak gehanteerd, volgens welke de gradiënt $\nabla f(x)$, als gedefinieerd in (2.1.2), als een kolomvector wordt opgevat en de afgeleide vector $f_x = \frac{df}{dx}$ als een rijvector (of matrix). Dit heeft tot gevolg dat bijvoorbeeld voor de differentiaal van f geschreven kan worden als ($f \in C^2$)

$$df = \frac{df}{dx}\,dx = f_x dx = \nabla^T f(x) dx = \langle \nabla f, dx \rangle \tag{2.1.4}$$

Bij de ontwikkeling van de methoden voor de oplossing van probleem UMP (2.1.1) wordt meestal verondersteld dat f twee maal continu differentieerbaar is en een geïsoleerd minimum heeft in een punt x^* in de directe omgeving waarvan de functie kan worden benaderd door een positief definiete kwadratische vorm.

Noodzakelijke en voldoende voorwaarden voor een minimum (als $f \in C^2$).

2.1.2. Opdat in een punt x^* een minimum gevonden wordt van de functie f is het noodzakelijk dat er geen richting $d \in \mathbb{R}^n$ bestaat waarlangs de functie

afneemt d.i.

$$\neg\exists d \in \mathbf{R}^n: \nabla^T f(x^*)d < 0 \qquad (2.1.5)$$

Omdat alle $d \in R^n$ toegelaten zijn volgt hieruit direct als <u>eerste (orde) noodzakelijke voorwaarde</u> voor een minimum dat

$$\nabla f(x^*) = 0 \qquad (2.1.6)$$

Tweede orde benadering van de functie f in een punt x^* dat aan deze eerste orde noodzakelijk voorwaarde voldoet (Een dergelijk punt heet een <u>stationair</u> punt) levert

$$f(x^* + d) = f(x^*) + \frac{1}{2}d^T\nabla^2 f(x^*)d + o(\|d\|^2) \qquad (\|d\| \to 0) \qquad (2.1.7)$$

Hieruit volgt dat indien er een minimum is van f in x^* dat voldaan zal moeten zijn aan de <u>tweede (orde) noodzakelijke voorwaarde</u> voor een minimum t.w.

$$\forall d \in \mathbf{R}^n: d^T\nabla^2 f(x^*)d \geq 0 \qquad (2.1.8.a)$$

of equivalent, als voorwaarde voor de Hessiaan

$$G(x^*) = \nabla^2 f(x^*) \text{ niet negatief definiet} \qquad (2.1.8.b)$$

2.1.3. Voldoende voor het optreden van een minimum van $f \in C^2$ in een punt x^* is de combinatie van de volgende <u>tweede orde voldoende voorwaarden</u> voor een minimum

i) $$\nabla f(x^*) = 0, \qquad (2.1.9)$$

en

iia) $$\forall d \in \mathbf{R}^n: d^T\nabla^2 f(x^*)d > 0 \qquad (2.1.10.a.)$$

of, equivalent, als voorwaarde voor de Hessiaan

iib) $$G(x^*) = \nabla^2 f(x^*) \text{ positief definiet} \qquad (2.1.10.b)$$

Het bewijs dat deze voorwaarden inderdaad voldoende zijn is gebaseerd op de overweging dat het positief definiet zijn van de Hessiaan impliceert dat er een kleinste eigenwaarde $\lambda_0 > 0$ bestaat zodat voor alle d

$$d^T \nabla^2 f(x^*) d \geq \lambda_0 \|d\|^2$$

Substitutie hiervan in (2.1.7) geeft onmiddellijk

$$f(x^* + d) - f(x^*) = \frac{1}{2} d^T \nabla^2 f(x^*) d + o(\|d\|^2)$$

$$\geq \frac{1}{2} \lambda_0 \|d\|^2 + o(\|d\|^2) > 0 \qquad (\|d\| \to 0)$$

Methoden.

2.1.4. Voor het bepalen van de oplossing van minimaliseringsproblemen van het type UMP (2.1.1) bestaan twee duidelijk te onderscheiden klassen van directe methoden (vergelijk paragraaf 1.1, pt. 1.1.5) te weten

a) direct search methoden, en
b) descent methoden.

Bij de eerste klasse van methoden, de direct search methoden, worden bij opvolgende punten uitsluitend de functiewaarden vergeleken en op grond van deze vergelijkingen nieuwe punten geselecteerd. Bij de tweede klasse van methoden, de descent methoden, wordt het originele n-dimensionale probleem (2.1.1) vervangen door een reeks eendimensionale problemen van het type

$$\min\{f(x^{(k)} + \alpha d^{(k)}) \mid \alpha \in \mathbb{R}^1\} \qquad (2.1.11)$$

Bij deze laatste klasse van methoden worden behalve van de functiewaarden, mogelijk ook gebruik gemaakt van de gradiënt en de Hessiaan van de functie in de opvolgende punten.

2.1.5. Descent methoden worden gekenmerkt door een simpele standaard algorithme, die de volgende stappen omvat:

(0) kies een startpunt $x^{(0)}$, zet k := 0

(i) evalueer de functiewaarde $f(x^{(k)})$ en eventueel de gradiënt $\nabla f(x^{(k)})$ in het punt $x^{(k)}$

(ii) ga na of $x^{(k)}$ optimaal is; zo ja, dan klaar; zo nee, dan:

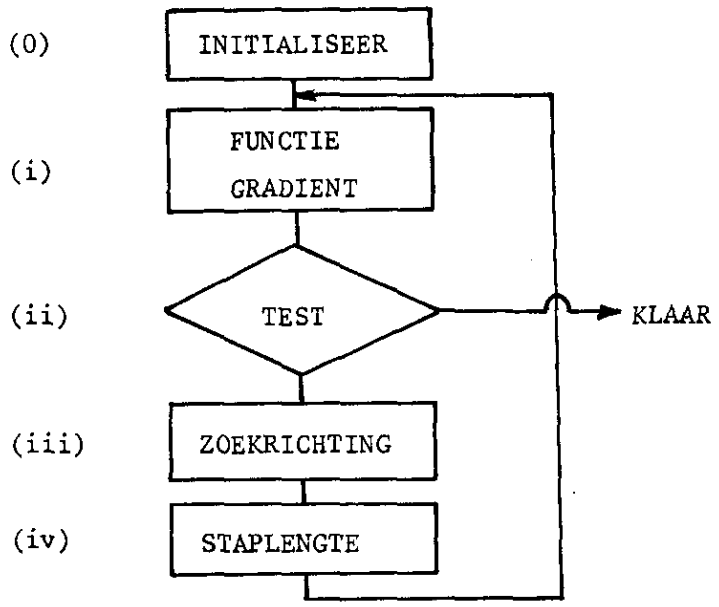
(iii) bepaal een nieuwe zoekrichting $d^{(k)}$

(iv) bepaal een staplengte (factor) $\alpha^{(k)}$ zo dat

$$f(x^{(k)} + \alpha^{(k)} d^{(k)}) < f(x^{(k)}) \qquad (2.1.12)$$

(v) zet $x^{(k+1)} := x^{(k)} + \alpha^{(k)} d^{(k)}$, k := k + 1 en ga terug naar stap (i).

Een flowdiagram van deze algorithme geeft het volgende schematische beeld



2.1.6. Met betrekking tot de individuele stappen van deze algorithme kunnen (in aanvulling op wat reeds gezegd is in pt. 1.1.6 van paragraaf 1.1) de volgende algemene opmerkingen worden gemaakt:

ad (ii) : Als convergentie- of stop criterium kunnen bij problemen van het type UMP de volgende drie practische criteria worden gebruikt

a) $\| \nabla f(x^{(k+1)}) \| < \varepsilon$ (2.1.13.a)

b) $\| f(x^{(k+1)}) - f(x^{(k)}) \| < \varepsilon$ (2.1.13.b)

c) $\| x^{(k+1)} - x^{(k)} \| < \varepsilon$ (2.1.13.c)

In de praktijk is een combinatie van deze criteria veelal het beste.

ad (iii): Voor het genereren van zoekrichtingen bestaan een groot aantal mogelijkheden. Bekende zoekrichtingen welke hierna in meer detail zullen worden besproken zijn onder andere:

a) coördinaat richting (zie paragraaf 2.3)
b) (negatieve) gradiënt richting (zie paragraaf 2.4)
c) geconjugeerde gradiënt richting (zie paragraaf 2.6)
d) Newton-richting (zie paragraaf 2.5).
e) quasi-Newton richting

ad (iv): De meest gebruikte methode voor het bepalen van de stapgrootte (factor) $\alpha^{(k)}$ is lijnminimalisering, d.i. de bepaling van $\alpha^{(k)}$ als oplossing van probleem (2.1.11)

$$f(x^{(k)} + \alpha^{(k)} d^{(k)}) = \min \{ f(x^{(k)} + \alpha d^{(k)}) \mid \alpha \in \mathbb{R}^1 \} \qquad (2.1.14)$$

Bij de meeste descent methoden wordt lijnminimalisering in een verschillende mate van perfectie toegepast. Een aantal van de bekendste methoden voor lijnminimalisering wordt besproken in de volgende paragraaf van dat hoofdstuk (paragraaf 2.2).

Globale convergentie theorie voor algorithmen.

2.1.7. Tot voor enkele jaren bestond de voornaamste ontwikkeling op het gebied van de numerieke methoden voor de oplossing van optimaliseringsproblemen van het type UMP uit het genereren van telkens andere algorithmen. Eerst later is men aandacht gaan besteden aan het onder een noemer brengen van de grote verscheidenheid aan algorithmen. Een opmerkelijke ontwikkeling in dat verband is de globale convergentie theorie van algorithmen van Zangwill [2.1.3], waarvan hieronder enige aspecten zullen worden toegelicht. Doel van deze theorie is te komen tot voorwaarden waaraan algorithmen

moeten voldoen opdat convergentie kan worden gegarandeerd. Centraal in deze theorie staan de volgende definities:

Definitie 2.1.7.a: Onder een algorithme A wordt verstaan een voorschrift (of afbeelding) volgens welk met ieder element x van een verzameling X een deelverzameling van die verzameling X wordt geassocieerd, d.i.

$$\forall x \in X: \quad A(x) \subset X \tag{2.1.15}$$

N.B. Een algorithme A is dus een afbeelding van een verzameling X op de machtsverzameling P(X) van X. Een dergelijke afbeelding wordt in de Engelse literatuur (en bij gebrek aan een beter woord ook hierna in deze syllabus) aangeduid als een "point-to-set-mapping".

Definitie 2.1.7.b: Een door een algorithme A gegenereerde rij $\{x^{(k)}\}_{k=0}^{\infty}$ is een rij waarvan de elementen $x^{(k+1)} \in X$ de eigenschap hebben dat zij behoren tot de deelverzameling $A(x^{(k)})$ gegenereerd door toepassing van de algorithme A op het voorgaande element van de rij, $x^{(k)}$, d.i.

$$x^{(k+1)} \in A(x^{(k)}) \tag{2.1.16}$$

Het eerste element $x^{(0)}$ (dat deze eigenschap niet heeft) wordt aangeduid als startpunt.

N.B. De reden om de door een algorithme A gegenereerde rij op deze niet-eenduidige manier te definiëren is om de mogelijkheid te creëren om over een dergelijke rij uitspraken te kunnen doen zonder dat alle details van het generatieproces volledig bekend zijn.

Voorbeeld 2.1.7.

Zij $X = \mathbf{R}^1$ en A gedefinieerd door

$$A(x) = [-|x|/2, +|x|/2]$$

dan zijn voorbeelden van door A gegenereerde rijen (met $x^{(0)} = 100$)

$$\{100, 50, 25, 12, -6, -2, 1, \tfrac{1}{2}, \ldots\}$$

$$\{100, -40, 20, -5, 2, 1, \tfrac{1}{4}, \ldots\}$$

$$\{100, 10, 1, 0.1, 0.01, 0.001, \ldots\}$$

Definitie 2.1.7.c: Zij Γ een gegeven deelverzameling van X, in het vervolg aangeduid als oplossingsverzameling, en A een algorithme dan heet een reëelwaardige functie Z op X een daalfunctie voor Γ en A indien geldt:

i) als $x \notin \Gamma$ en $y \in A(x)$ dan $Z(y) < Z(x)$

ii) als $x \in \Gamma$ en $y \in A(x)$ dan $Z(y) \leq Z(x)$

N.B. In de gebruikelijke situatie voor het probleem UMP kan de objectfunctie f(x) veelal als daalfunctie fungeren voor een groot aantal algorithmen en voor oplossingsverzamelingen bestaande uit bijvoorbeeld alle stationaire punten.

2.1.8. Een eigenschap van point-to-set-mappings welke van essentieel belang is voor de hierna te bespreken globale convergentiestelling is de generalisatie van het begrip van de continuiteit van afbeeldingen. Deze eigenschap wordt aangeduid als het gesloten zijn van de point-to-set-mapping en wordt als volgt gedefinieerd.

Definitie 2.1.8. Een point-to-set-mapping A van X naar de machtsverzameling van Y heet gesloten in het punt $x \in X$ indien de veronderstellingen

1) $x^{(k)} \to x \quad x^{(k)} \in X$

2) $y^{(k)} \to y \quad y^{(k)} \in A(x^{(k)})$

impliceren dat

3) $y \in A(x)$

Een point-to-set-mapping A heet gesloten op X indien A gesloten is op ieder punt van X

N.B. Het is eenvoudig in te zien dat continue functies op X kunnen worden opgevat als gesloten point-to-set-mappings. Immers als geldt

$$x^{(k)} \to x \quad x^{(k)} \in X$$

en

$$y^{(k)} \to y \quad \text{waar } y^{(k)} = f(x^{(k)})$$

dan volgt uit de continuïteit dat

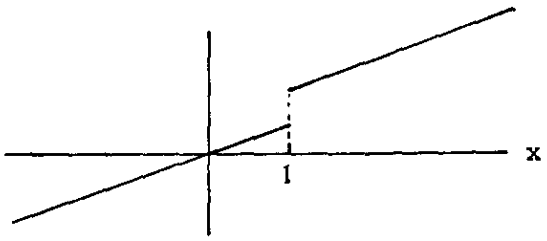
$$y = \lim_{x^{(k)} \to x} f(x^{(k)}) = f(x)$$

Voorbeeld 2.1.8.

Een voorbeeld van een niet-gesloten point-to-set-mapping gedefinieerd op $X = \mathbb{R}^1$ is de afbeelding gedefinieerd door

$$A(x) = \frac{1}{2}x + \frac{1}{2} \qquad x > 1$$

$$= \frac{1}{2}x \qquad x \leq 1$$



2.1.9. De meeste algorithmen voor de oplossing van niet-lineaire optimaliseringsproblemen bestaan niet uit één enkele algorithme in de zin van Definitie 2.1.7.a doch in plaats daarvan uit een samenstelling van twee of meer van dergelijke algorithmen. Van belang is het gesloten zijn van de samengestelde algorithme.

Definitie 2.1.9.a: Als $A : X \to P(Y)$ en $B : Y \to P(Z)$ point-to-set-mappings zijn dan wordt een samengestelde point-to-set-mapping $C = BA : X \to P(Z)$ gedefinieerd door de relatie

$$C(x) = \bigcup_{y \in A(x)} B(y) \qquad (2.1.17)$$

Voor deze samengestelde point-to-set-mappings geldt de volgende uitspraak

Propositie 2.1.9.b: Als $A : X \to P(Y)$ en $B : Y \to P(Z)$ point-to-set-mappings zijn met A gesloten in x , en B gesloten in A(x) dan zal de samengestelde point-to-set-mapping $C = BA$ eveneens gesloten zijn indien voor iedere convergente rij $\{x^{(k)}\}$ met $x^{(k)} \to x$ en iedere rij $\{y^{(k)}\}$ met de eigenschap dat $y^{(k)} \in A(x^{(k)})$ er een $y \in Y$ en een deelrij $\{y^{(k_i)}\}$ te vinden is zodat $y^{(k_i)} \to y$.

Bewijs. Zij $\{x^{(k)}\}$ een rij met $x^{(k)} \to x$ en $\{z^{(k)}\}$ een rij met $z^{(k)} \to z$ en $z^{(k)} \in C(x^{(k)})$. Kies dan een rij $\{y^{(k)}\}$ met $y^{(k)} \in A(x^{(k)})$ en $z^{(k)} \in B(y^{(k)})$ en laat $\{y^{(k_i)}\}$ de deelrij zijn van $\{y^{(k)}\}$ en y het element waarnaar $\{y^{(k_i)}\}$ convergeert (volgens de extra veronderstelling.) Omdat A gesloten is in x volgt dan dat $y \in A(x)$.
Corresponderend met $\{y^{(k_i)}\}$ met $y^{(k_i)} \to y$ bestaat er nu een $\{z^{(k_i)}\}$ met $z^{(k_i)} \to z$ en omdat B gesloten is in y volgt dan dat $z \in B(y) \subset BA(x) = C(x)$ □

2.1.10. Uit deze propositie volgen direct een tweetal practische gevolgen.

Gevolg 2.1.10.a: Als $A : X \to P(Y)$ en $B : Y \to P(Z)$ point-to-set-mappings zijn met A gesloten in x en B gesloten in A(x) en Y is compact, dan geldt dat ook de samengestelde point-to-set-mapping C = BA gesloten is in x.

Gevolg 2.1.10.b: Als $A : X \to Y$ een gewone (point-to-point) afbeelding en $B : Y \to P(Z)$ een point-to-set-mapping met A continu in x en B gesloten in A(x) dan geldt dat ook de samengestelde point-to-set-mapping C = BA gesloten is in x.

2.1.11. Met behulp van de in het voorgaande geïntroduceerde begrippen kan de volgende *globale convergentiestelling voor algorithmen* worden geformuleerd (vergelijk [2.1.1]).

Stelling 2.1.11. (Globale convergentiestelling van Zangwill).
Zij A een algorithme gedefinieerd op X, zij $\{x^{(k)}\}$ een door de algorithme A gegenereerde rij met startpunt $x^{(0)}$ en zij Γ een (vooraf gegeven) oplossingsverzameling. Indien geldt dat:

i) alle elementen $x^{(k)}$ elementen zijn van een compacte verzameling $S \subset X$

ii) er een daalfunctie Z gedefinieerd op X bestaat zodanig dat:
   a) als $x \notin \Gamma$ en $y \in A(x)$ dan $Z(y) < Z(x)$
   b) als $x \in \Gamma$ en $y \in A(x)$ dan $Z(y) \leq Z(x)$

iii) de algorithme A gesloten is in alle punten $x \in X$ buiten Γ

dan behoort de limiet van iedere convergente deelrij van $\{x^{(k)}\}$ tot de oplossingsverzameling.

Bewijs. Stel dat de rij $\{x^{(k)}\}_{k\in K}$ convergeert naar een limiet x. Omdat Z continu is geldt dan dat ook $\{Z(x^{(k)})\}_{k\in K} \to Z(x)$, d.w.z. voor iedere $\varepsilon > 0$ bestaat er een $K \in K$ zodanig dat $Z(x^{(k)}) - Z(x) < \varepsilon$ voor alle $k \in K$ waarvoor $k > K$. Op grond van het monotone gedrag van de functiewaarden van Z in de opvolgende elementen van de door A gegenereerde rij geldt voor alle k en k' met $k' > k$ dat $Z(x^{(k')}) \leq Z(x^{(k)})$. In het bijzonder geldt dan voor alle $k' > k > K$ dat $Z(x^{(k')}) - Z(x) \leq Z(x^{(k)}) - Z(x) < \varepsilon$. Dit impliceert dat $Z(x^{(k)}) \to Z(x)$ voor alle k. Rest te bewijzen dat $x \in \Gamma$. Veronderstel het tegengestelde: $x \notin \Gamma$ en beschouw de rij $\{x^{(k+1)}\}_{k\in K}$ (d.i de rij met als elementen de opvolgers in de originele rij van de elementen van de convergente deelrij). Omdat alle $x^{(k+1)}$ behoren tot een compacte verzameling bestaat er een deelrij $\{x^{(k+1)}\}_{k\in\bar{K}}$ met $\bar{K} \subset K$ waarvoor $x^{(k+1)} \to \bar{x}$ met uiteraard $Z(x^{(k+1)}) \to Z(\bar{x}) = Z(x)$. Er resulteren dan twee convergente deelrijen $\{x^{(k)}\}_{k\in\bar{K}}$ met $x^{(k)} \to x$ en $\{x^{(k+1)}\}_{k\in\bar{K}}$ met $x^{(k+1)} \in A(x^{(k)})$ en $x^{(k+1)} \to \bar{x}$. Omdat volgens de veronderstelling A gesloten is in $x \notin \Gamma$ geldt dat $\bar{x} \in A(x)$ en derhalve dat $Z(\bar{x}) < Z(x)$. Dit laatste is in tegenspraak met de bovenbewezen convergentie van $\{Z(x^{(k)})\}$ naar de limiet Z(x). De veronderstelling $x \notin \Gamma$ is dus onjuist en de stelling bewezen. □

2.1.12. Voorbeeld 2.1.12.a.

Zij A een algorithme gedefinieerd op $\mathbb{R}^1$ door het voorschrift (vergelijk Voorbeeld 2.1.8)

$$A(x) = \tfrac{1}{2}x + \tfrac{1}{2} \qquad x > 1$$

$$= \tfrac{1}{2}x \qquad x \leq 1$$

en laat

$$x^{(0)} = 2,\ \Gamma = \{0\} \quad \text{en} \quad Z = |x|$$

dan volgt voor de elementen van de door A gegenereerde rij $\{x^{(k)}\}$ dat $x^{(k)} \to 1$ maar $1 \notin \Gamma$. Convergentie naar een punt buiten de oplossingsverzameling is hier mogelijk omdat A niet gesloten is in $x = 1$.

Voorbeeld 2.1.12.b.

Zij A een algorithme gedefinieerd op $\mathbb{R}^1$ volgens het voorschrift

$$A(x) = x + 1$$

en laat

$$x^{(0)} = 0,\ \Gamma = \emptyset \text{ en } Z = e^{-x}$$

dan volgt voor de door A gegenereerde rij $\{x^{(k)}\}$ dat $x^{(k)} \to \infty$ en $Z(x^{(k)}) \to 0$. Er kan in dit geval geen convergente deelrij worden afgesplitst omdat de $x^{(k)}$ niet liggen in een compacte verzameling.

Voorbeeld 2.1.12.c.

Zij A een algorithme gedefinieerd op $\mathbb{R}^1_+$ volgens het voorschrift

$$A(x) = [0,x) \qquad 0 < x \leq 1$$

$$= 0 \qquad x = 0$$

en laat

$$x^{(0)} = 1,\ \Gamma = \{0\} \text{ en } Z = x.$$

dan volgt dat de door A gegenereerde rij $\{x^{(k)}\}$ met

$$x^{(k+1)} = x^{(k)} - \frac{1}{2^{k+2}}$$

convergeert naar de limiet $x = \frac{1}{2} \notin \Gamma$. Dit is mogelijk omdat de algorithme A niet gesloten is buiten de oplossingsverzameling.

Convergentiesnelheid.

2.1.13. Van groot practisch belang voor de beoordeling van de bruikbaarheid van algorithmen is de convergentiesnelheid van de door algorithmen gegenereerde rijen. Om hierover uitspraken te kunnen doen is het nodig precies te formuleren wat men met snelle, respectievelijk langzame convergentie bedoelt. De volgende definities die afkomstig zijn van Ortega en Rheinboldt [2.1.4] vormen een basis voor dergelijke uitspraken.

Definitie 2.1.13.

Als $\{x^{(k)}\}$ een rij is die convergeert naar een limiet x^* , (d.i. $x^{(k)} \to x^*$) dan verstaat men onder de orde van convergentie van de rij $\{x^{(k)}\}$ het supremum van de niet-negatieve getallen p waarin geldt

$$0 \leq \lim_{k\to\infty} \sup \frac{\|x^{(k+1)} - x^*\|}{\|x^{(k)} - x^*\|^p} < \infty \qquad (\text{als } x^{(k)} \neq x^*) \qquad (2.1.1$$

N.B. Deze definitie laat zien dat de orde van convergentie te maken heeft met het asymptotisch convergentiegedrag.

Voorbeeld 2.1.13.

a) De rij $\{a^k\}$ met $0 < a < 1$ convergeert naar 0 met een orde van convergentie gelijk aan 1.

b) De rij $\{a^{(2^k)}\}$ met $0 \leq a < 1$ convergeert naar 0 met een orde van convergentie gelijk aan 2.

2.1.14. In het geval van convergentie met orde 1 (en analoog in het geval van convergentie met orde 2, 3 etc.) is een nadere precizering van de convergentiesnelheid mogelijk.

Definitie 2.1.14.a.

Als $\{x^{(k)}\}$ een rij is die convergeert naar een limiet x^* op zodanige wijze dat de limiet

$$\lim_{k\to\infty} \frac{\|x^{(k+1)} - x^*\|}{\|x^{(k)} - x^*\|} = \beta \qquad (x^{(k)} \neq x^*) \qquad (2.1.$$

bestaat, dan spreekt men van

lineaire convergentie als $0 < \beta < 1$

en van

superlineaire convergentie als $\beta = 0$.

Definitie 2.1.14.b.

Als $\{x^{(k)}\}$ een rij is die convergeert naar een limiet x^* op zodanige wijze dat de limiet

$$\lim_{k\to\infty} \frac{\|x^{(k+1)} - x^*\|}{\|x^{(k)} - x^*\|^2} = \beta \qquad (x^{(k)} \neq x^*) \tag{2.1.20}$$

bestaat dan spreekt men van

kwadratische convergentie als $0 < \beta < \infty$

en van

superkwadratische convergentie als $\beta = 0$.

N.B. Iedere convergentie met orde groter dan 1 heet superlineair. De definitie toont aan dat er ook superlineaire convergentie op kan treden met orde 1.

Voorbeeld 2.1.14.

a) De rij $\{1/k\}$ convergeert naar 0 met orde 1. De convergentie is echter niet lineair ($\beta = 1$).

b) De rij $\{(1/k)^k\}$ convergeert naar 0 met orde 1 (immers $x^{(k+1)}/(x^{(k)})^p \to \infty$ voor $p > 1$) en wel superlineair ($\beta = 0$).

Referenties.

2.1.15. Meer over de in deze sectie behandelde onderwerpen is onder andere te vinden in de volgende referenties

[2.1.1]: Zie [1.1.1] Luenberger (1973).

[2.1.2]: Zie [1.1.3] Murray (1972).

[2.1.3]: Zangwill, W.I.: "Nonlinear programming, a unified approach" Prentice Hall Inc., Englewood Cliffs, M.J., (1969).

[2.1.4]: Ortega, A.J. and Rheinboldt, W.: "Iterative solution of nonlinear equations", Academic Press, New York (1970).

[2.1.5]: Polak, E.:"Computational methods in optimization, a unified approach", Academic Press, New York, (1971).

§ 2.2. Eendimensionale minimaliseringsalgorithmen

2.2.1. Restrictie van een functie f(x) van n variabelen, $f : \mathbb{R}^n \to \mathbb{R}^1$, tot een lijn in $\mathbb{R}^n$ gegeven door de parametervoorstelling $x = x^{(k)} + \alpha d^{(k)}$, $\alpha \in \mathbb{R}^1$, genereert de definitie van een functie h(α) van één variabele, $h : \mathbb{R}^1 \to \mathbb{R}^1$, door middel van de uitdrukking

$$h(\alpha) := f(x^{(k)} + \alpha d^{(k)}) \quad . \tag{2.2.1}$$

Is de functie f(x) tweemaal differentieerbaar met als eerste afgeleide de gradiënt ∇f(x) en als tweede afgeleide de Hessiaan

$$G(x) := [(\frac{\partial^2 f}{\partial x_i \partial x_j}(x))]$$

dan geldt voor de eerste en tweede afgeleiden van h(α) naar α respectievelijk

$$h'(\alpha) = \frac{dh(\alpha)}{d\alpha} = \nabla^T f(x^{(k)} + \alpha d^{(k)}) d^{(k)} \tag{2.2.2}$$

$$h''(\alpha) = \frac{d^2 h(\alpha)}{d\alpha^2} = (d^{(k)})^T G(x^{(k)} + \alpha d^{(k)}) d^{(k)} \quad . \tag{2.2.3}$$

Uit deze uitdrukkingen volgen o.a. direct de volgende interessante consequenties:

A: als de functie f(x) in het punt $x^{(k)} + \alpha^{(k)} d^{(k)}$ een <u>lijnminimum</u> heeft op de lijn $x = x^{(k)} + \alpha d^{(k)}$ dan geldt

$$h'(\alpha^{(k)}) = \nabla^T f(x^{(k)} + \alpha^{(k)} d^{(k)}) d^{(k)} = 0 \tag{2.2.4}$$

d.i. in het lijnminimum zijn de gradiënt en de lijn onderling <u>orthogonaal</u>

$$\nabla f(x^{(k)} + \alpha^{(k)} d^{(k)}) \perp d^{(k)}$$

B: als de functie f(x) convex is over een convex gebied $S \subset R^n$ dan is ook h(α) <u>convex</u> over het (convexe) interval van α-waarden waarvoor $x^{(k)} + \alpha d^{(k)} \in S$.

C: de staplengte (-factor) $\alpha^{(k)}$ naar het lijnminimum wordt gegeven door

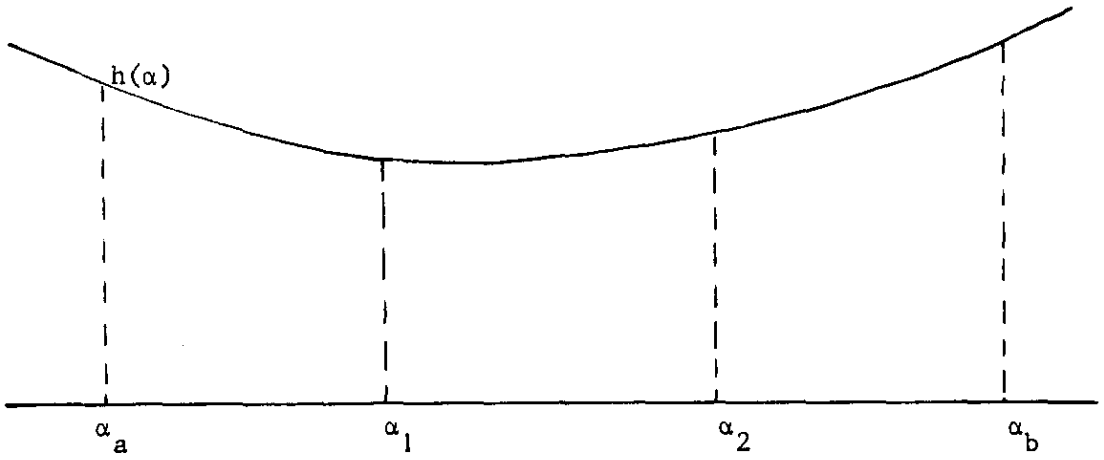
$$\alpha^{(k)} = \frac{\nabla^T f(x^{(k)}) d^{(k)}}{d^{(k)T} G(x^{(k)} + \bar{\alpha} d^{(k)}) d^{(k)}} \qquad 0 < \bar{\alpha} < \alpha^{(k)} \tag{2.2.5}$$

Deze uitdrukking is van belang voor een eerste schatting van de optimale staplengte factor.

2.2.2. Voor het numeriek bepalen van lijnminima, (d.i de minima van de corresponderende functies $h(\alpha)$) worden twee soorten schattingen gebruikt, namelijk, intervalschattingen en puntschattingen. In het eerste geval bepaalt men een interval waarbinnen het minimum ligt, in het tweede geval bepaalt men een enkel punt als benadering voor het lijnminimum. Uitgangspunt voor alle numerieke lijnminimaliseringsmethoden is de veronderstelling dat de functie $h(\alpha)$ unimodaal is over het te beschouwen interval, d.w.z. dat de functie in het interval slechts een stationair punt ($h'(\alpha) = 0$) heeft.

Intervalschattingen

2.2.3. De methoden voor het bepalen van intervalschattingen zijn gebaseerd op de observatie dat om het interval te kunnen verkleinen waarbinnen het minimum van de unimodale functie ligt, de functiewaarden behalve in de eindpunten (α_a en α_b) van het interval ook moet worden berekend in twee inwendige punten (α_1 en α_2) van het interval. Is dat geschied dan kan het interval worden verkleind met de keuzemogelijkheden:



A: als $h(\alpha_1) \leq h(\alpha_2)$, dan kies een nieuw inwendig punt $\alpha_3 \neq \alpha_1$ in het interval $[\alpha_a,\alpha_2]$, evalueer $h(\alpha_3)$, en zet $\alpha_a := \alpha_a, \alpha_b := \alpha_2, \alpha_i := \min(\alpha_1,\alpha_3)$ en $\alpha_2 := \max(\alpha_1,\alpha_3)$

B: als $h(\alpha_1) > h(\alpha_2)$ dan kies een nieuw inwendig punt $\alpha_3 \neq \alpha_2$ in het interval $[\alpha_1,\alpha_b]$, evalueer $h(\alpha_3)$ en zet $\alpha_a := \alpha_1, \alpha_b := \alpha_b, \alpha_1 := \min(\alpha_2,\alpha_3)$ en $\alpha_2 := \max(\alpha_2,\alpha_3)$.

2.2.4. Een efficiënt algorithme ontstaat indien de onderlinge verhoudingen in alle opvolgende intervalverdelingen gehandhaafd blijven. Dit is het geval bij de "gulden-snede"-procedure [2.2.7] waaraan ten grondslag ligt de "gulden-snede"-relatie

$$\frac{\alpha_2 - \alpha_a}{\alpha_b - \alpha_a} = \frac{\alpha_b - \alpha_1}{\alpha_b - \alpha_a} = \frac{\alpha_1 - \alpha_a}{\alpha_2 - \alpha_a} = \frac{\alpha_b - \alpha_2}{\alpha_b - \alpha_1} = \frac{\sqrt{5} - 1}{2} = \tau = 0.618034... \qquad (2.2.6)$$

De algorithme dat hierop is gebaseerd heeft de volgende vorm:

Algorithme voor de "gulden-snede"-procedure:

(0) evalueer de functie $h(\alpha)$ in de eindpunten $\alpha_a^{(0)}$ en $\alpha_b^{(0)}$ van het gegeven interval en bepaal de punten

$$\alpha_1^{(0)} := \alpha_a^{(0)} + (1 - \tau)(\alpha_b^{(0)} - \alpha_a^{(0)})$$

$$\alpha_2^{(0)} := \alpha_b^{(0)} - (1 - \tau)(\alpha_b^{(0)} - \alpha_a^{(0)})$$

bereken $h(\alpha_1^{(0)})$, zet $k := 0$ en ga naar (ii)

(i) bereken $h(\alpha_1^{(k)})$ en ga naar (iii)

(ii) bereken $h(\alpha_2^{(k)})$

(iii) als $h(\alpha_1^{(k)}) \leq h(\alpha_2^{(k)})$ dan zet $\alpha_a^{(k+1)} := \alpha_a^{(k)}$, $\alpha_b^{(k+1)} := \alpha_2^{(k)}$

$$\alpha_2^{(k+1)} := \alpha_1^{(k)}, \quad \alpha_1^{(k+1)} := \alpha_b^{(k+1)} - \tau(\alpha_b^{(k+1)} - \alpha_a^{(k+1)})$$

als $|\alpha_b^{(k+1)} - \alpha_a^{(k+1)}| < \varepsilon$, dan klaar, zo niet, zet $k := k + 1$ en ga naar (i)

(iv) als $h(\alpha_1^{(k)}) > h(\alpha_2^{(k)})$ dan zet $\alpha_a^{(k+1)} := \alpha_1^{(k)}$, $\alpha_b^{(k+1)} := \alpha_b^{(k)}$

$$\alpha_1^{(k+1)} := \alpha_2^{(k)}, \quad \alpha_2^{(k+1)} := \alpha_a^{(k+1)} + \tau(\alpha_b^{(k+1)} - \alpha_a^{(k+1)})$$

als $|\alpha_b^{(k+1)} - \alpha_a^{(k+1)}| < \varepsilon$, dan klaar, zo niet zet $k := k + 1$ en ga naar (ii).

Na n + 1 (+2) functie-evaluaties is de lengte van het interval waarbinnen (onder voorwaarde van unimodaliteit van de functie) het minimum kan worden gegarandeerd verkleind tot τ^n maal de lengte van het begininterval, dus

$$\alpha_b^{(n)} - \alpha_a^{(n)} = \tau^n(\alpha_b^{(0)} - \alpha_a^{(0)}). \qquad (2.2.7)$$

2.2.5. De optimale interval verkleining bij een gegeven aantal stappen wordt gerealiseerd bij de Fibonacci-zoekprocedure [2.2.4]. Deze procedure is gebaseerd op de observatie dat het grootste interval L_n dat met n functie-evaluaties gereduceerd kan worden tot de lengte 1 (volgens de algorithme aangegeven in pt. 2.2.3) voldoet aan de homogene lineaire differentie vergelijking

$$L_n = L_{n-1} + L_{n-2} \qquad (2.2.8.a)$$

met beginwaarden

$$L_0 = L_1 = 1 . \qquad (2.2.8.b)$$

De oplossing van deze differentievergelijking wordt gegeven door de uitdrukking

$$L_i = Ar_1^i + Br_2^i \qquad (2.2.9)$$

waar r_1 en r_2 de oplossingen zijn van de karakteristieke vergelijking

$$r^2 - r - 1 = 0 \qquad (2.2.10)$$

en waar de constanten A en B kunnen worden bepaald door substitutie van de oplossing voor de beginwaarden. De uiteindelijke oplossing krijgt daarmee de vorm

$$L_i = \frac{1}{\sqrt{5}} \{(\frac{1 + \sqrt{5}}{2})^{i+1} - (\frac{1 - \sqrt{5}}{2})^{i+1}\} . \qquad (2.2.11)$$

Uitwerking leert dat de aldus bepaalde maximum-intervallengte L_n voor alle n = 0,1,... juist gelijk is aan het n-de Fibonacci-getal

$$L_n = F_n \qquad (2.2.12)$$

d.i. het (n + 1)-de getal in de rij 1,1,2,3,5,8,13,... . Toepassing van deze overwegingen leidt tot de volgende algorithme:

Algorithme voor de Fibonacci-zoekprocedure

(0) evalueer de functie $h(\alpha)$ in de eindpunten $\alpha_a^{(0)}$ en $\alpha_b^{(0)}$ van het gegeven interval en bepaal de punten

$$\alpha_1^{(0)} := \alpha_a^{(0)} + (1 - \frac{F_{N-1}}{F_N})(\alpha_b^{(0)} - \alpha_a^{(0)})$$

$$\alpha_2^{(0)} := \alpha_b^{(0)} - (1 - \frac{F_{N-1}}{F_N})(\alpha_b^{(0)} - \alpha_a^{(0)})$$

bereken $h(\alpha_1^{(0)})$, zet $k := 0$ en ga naar (ii)

(i) bereken $h(\alpha_1^{(k)})$ en ga naar (iii)

(ii) bereken $h(\alpha_2^{(k)})$

(iii) als $h(\alpha_1^{(k)}) \leq h(\alpha_2^{(k)})$ dan zet $\alpha_a^{(k+1)} := \alpha_a^{(k)}$, $\alpha_b^{(k+1)} := \alpha_2^{(k)}$

$$\alpha_2^{(k+1)} := \alpha_1^{(k)} \quad , \quad \alpha_1^{(k+1)} := \alpha_b^{(k+1)} - (\frac{F_{N-k-2}}{F_{N-k-1}})(\alpha_b^{(k+1)} - \alpha_a^{(k+1)})$$

als $N - k = 2$, dan klaar, zo niet, zet $k := k + 1$ en ga naar (i)

(iv) als $h(\alpha_1^{(k)}) > h(\alpha_2^{(k)})$ dan zet $\alpha_a^{(k+1)} := \alpha_1^{(k)}$, $\alpha_b^{(k+1)} := \alpha_b^{(k)}$

$$\alpha_1^{(k+1)} := \alpha_2^{(k)} \quad , \quad \alpha_2^{(k+1)} := \alpha_a^{(k+1)} + (\frac{F_{N-k-2}}{F_{N-k-1}})(\alpha_b^{(k+1)} - \alpha_a^{(k+1)})$$

als $N - k = 2$, dan klaar, zo niet, zet $k := k + 1$ en ga naar (ii).

Na N (+2) functie-evaluaties is de lengte van het interval waarbinnen (onder voorwaarde van unimodaliteit van de functie) het minimum kan worden gegarandeerd, verkleind tot $\frac{1}{F_N}$ maal de lengte van het begininterval dus,

$$\alpha_b^{(N-1)} - \alpha_a^{(N-1)} = \frac{1}{F_N}(\alpha_b^{(0)} - \alpha_a^{(0)}) \quad . \qquad (2.2.13)$$

2.2.6. Een vergelijking van de "gulden-snede"-procedure en de Fibonacci-zoekprocedure levert o.m. de volgende observaties:

A: bij grote N geldt dat de verhouding van de opvolgende Fibonacci getallen gelijk wordt aan

$$\lim_{N\to\infty} \frac{F_{N-1}}{F_N} = \frac{1}{r_1} = \tau \qquad (2.2.14)$$

waar r_1 de grootste wortel is van de karakteristieke vergelijking die correspondeert met de differentie vergelijking in pt. 2.2.5.

B: de getallen $G_i := 1/\tau^{i-1} = r_1^{i-1}$ vormen de oplossing van de differentievergelijking

$$G_N = G_{N-1} + G_{N-2} \qquad (2.2.15.a)$$

met als beginwaarden

$$G_0 = \tau \ , \quad G_1 = 1 \ . \qquad (2.2.15.b)$$

Eenvoudig valt in te zien dat

$$G_N < F_N < G_{N+1} \qquad (2.2.16.)$$

waaruit volgt dat

$$\tau^N = \frac{1}{G_{N+1}} < \frac{1}{F_N} < \frac{1}{G_N} = \tau^{N-1} \ . \qquad (2.2.17)$$

"Dit impliceert dat het interval dat bij een gelijk aantal functie-evaluaties gegeneerd wordt door de Fibonacci-procedure kleiner is dan het interval dat gegenereerd wordt door de gulden-snede procedure (want $G_{n+1} < F_{n+1}$)."

Op grond van deze feiten, die aantonen dat de "gulden-snede"-procedure voor grotere N nauwelijks minder efficiënt is dan de "optimale" Fibonacci-zoekprocedure, en de omstandigheid dat de "gulden-snede"-procedure veel eenvoudiger te programmeren valt, wordt de voorkeur in de practijk steeds gegeven aan de "gulden-snede"-procedure boven de Fibonacci-zoekprocedure.

Puntschattingen

2.2.7. In de practijk van de meeste numerieke minimaliseringsmethoden moet in iedere iteratieslag een (of meer) keer een lijnminimalisering worden uitgevoerd. Dit impliceert dat beperking van het aantal functie-evaluaties per lijnmini-

mum van bijzonder groot belang kan zijn voor het totaal aantal benodigde functie-evaluaties, dat een belangrijke maat is voor de kwaliteit van de betreffende minimaliseringsprocedure. Een van de middelen om het aantal functie-evaluaties te beperken is beter gebruik te maken van de verkregen informatie door ook de berekende functiewaarden zelf in plaats van alleen hun onderlinge relatie in de lijnminimaliseringsprocedures te betrekken. Een mogelijkheid hiertoe is het gebruik van interpolatieformules die op grond van een gering aantal berekende functiewaarden reeds een benadering van de functie en daarmee tegelijk een benadering van het minimum geven.

Kwadratische interpolatie

2.2.8. Zodra in drie verschillende punten α_1, α_2 en α_3 op de lijn de functiewaarden $h(\alpha_1)$, $h(\alpha_2)$, $h(\alpha_3)$ berekend zijn kan met behulp van de Lagrange-interpolatieformule een parabool $P(\alpha)$ door die drie punten worden gelegd. Deze eerste benadering voor de functie $h(\alpha)$ heeft de vorm:

$$P(\alpha) = \frac{h(\alpha_1)(\alpha - \alpha_2)(\alpha - \alpha_3)}{(\alpha_1 - \alpha_2)(\alpha_1 - \alpha_3)} + \frac{h(\alpha_2)(\alpha - \alpha_3)(\alpha - \alpha_1)}{(\alpha_2 - \alpha_3)(\alpha_2 - \alpha_1)} + \frac{h(\alpha_3)(\alpha - \alpha_1)(\alpha - \alpha_2)}{(\alpha_3 - \alpha_1)(\alpha_3 - \alpha_2)} \,. \qquad (2.2.18)$$

Het minimum van deze parabool wordt gegeven door de equivalente uitdrukkingen

$$\hat{\alpha} = \tfrac{1}{2} \frac{h(\alpha_1)(\alpha_2^2 - \alpha_3^2) + h(\alpha_2)(\alpha_3^2 - \alpha_1^2) + h(\alpha_3)(\alpha_1^2 - \alpha_2^2)}{h(\alpha_1)(\alpha_2 - \alpha_3) + h(\alpha_2)(\alpha_3 - \alpha_1) + h(\alpha_3)(\alpha_1 - \alpha_2)} \,, \qquad (2.2.19.a)$$

$$\hat{\alpha} = \tfrac{1}{2} \frac{\alpha_1^2(h(\alpha_2) - h(\alpha_3)) + \alpha_2^2(h(\alpha_3) - h(\alpha_1)) + \alpha_3^2(h(\alpha_1) - h(\alpha_2))}{\alpha_1(h(\alpha_2) - h(\alpha_3)) + \alpha_2(h(\alpha_3) - h(\alpha_1)) + \alpha_3(h(\alpha_1) - h(\alpha_2))} \,. \qquad (2.2.19.b)$$

en

$$\hat{\alpha} = \tfrac{1}{2} \frac{(\alpha_1 + \alpha_2)[(\alpha_1 - \alpha_2)(h(\alpha_3) - h(\alpha_2))] + (\alpha_2 + \alpha_3)[(\alpha_2 - \alpha_3)(h(\alpha_1) - h(\alpha_2))]}{[(\alpha_1 - \alpha_2)(h(\alpha_3) - h(\alpha_2))] + [(\alpha_2 - \alpha_3)(h(\alpha_1) - h(\alpha_2))]} \qquad (2.2.19.c)$$

De laatste uitdrukking is aantrekkelijk in de praktijk omdat dezelfde uitdrukkingen (tussen de vierkante haken) voorkomen in teller en noemer. Dit betekent een besparing van het rekenwerk. In het geval dat de punten α_1, α_2, α_3 op gelijke afstand $\Delta\alpha$ van elkaar liggen, vereenvoudigen deze uitdrukkingen tot

$$\hat{\alpha} = \alpha_2 - \frac{(h(\alpha_3) - h(\alpha_1))\ \Delta\alpha}{2(h(\alpha_1) - 2h(\alpha_2) + h(\alpha_3))} \qquad (2.2.20)$$

De gebruikelijke procedure bij lijnminimalisering met behulp van de kwadratische interpolatieformule bestaat daaruit dat men eerst een interval bepaalt dat het minimum insluit en dan daarna een nieuw punt bepaalt met behulp van de gegeven uitdrukkingen. In veel gevallen neemt men met deze eerste schatting genoegen. Nauwkeuriger schattingen zijn mogelijk door de functiewaarde te bepalen in het nieuwe punt en afhankelijk van de onderlinge relatie van de 3 + 1 bekende punten daarvan drie punten te kiezen waarvan de twee uitersten het minimum insluiten. Met behulp van de gegeven uitdrukkingen kan opnieuw een nieuw punt worden bepaald en de procedure telkens herhaald. Als stopcriterium van het aldus gedefinieerde iteratieve lijnminimaliseringsproces wordt vaak genomen de voorwaarde dat de nieuwe schatting binnen een zekere nauwkeurigheid overeenkomt met de voorgaande. (zie pt. 2.2.12)

Kubische interpolatie

2.2.9. In plaats van een parabool kan men ook een derdegraads polynoom

$$C(\alpha) = a\alpha^3 + b\alpha^2 + c\alpha + d \qquad (2.2.21)$$

als interpolatie formule gebruiken. Deze procedure past men vaak toe in het geval dat men in twee punten α_1 en α_2 niet alleen de functiewaarden doch tevens de afgeleiden kent. Uitwerking van deze schatting in het geval dat $h'(\alpha_1) < 0$ en $h'(\alpha_2) > 0$ is geeft als schatting voor het minimum de uitdrukking

$$\hat{\alpha} = \alpha_1 + \left(\frac{-b + \sqrt{b^2 - 3ac}}{3a}\right)(\alpha_2 - \alpha_1) \qquad (2.2.22)$$

waarin

$$a = h'(\alpha_2) + h'(\alpha_1) - 2(h(\alpha_2) - h(\alpha_1))$$

$$b = -h'(\alpha_2) - 2h'(\alpha_1) + 3(h(\alpha_2) - h(\alpha_1)) \qquad (2.2.23)$$

$$c = h'(\alpha_1) (< 0)$$

Substitutie hiervan in (2.2.22) geeft

$$\hat{\alpha} = \alpha_1 + \left(\frac{h'(\alpha_1) + z + w}{h'(\alpha_2) + h'(\alpha_1) + 2z}\right)(\alpha_2 - \alpha_1) \qquad (2.2.24.a)$$

waarin

$$z = h'(\alpha_2) + h'(\alpha_1) - 3(h(\alpha_2) - h(\alpha_1)) \; (= -b - h'(\alpha_1))$$

$$w = \sqrt{z^2 - h'(\alpha_1)h'(\alpha_2)} \; . \qquad (= \sqrt{b^2 - 3ac}) \qquad (2.2.24.b)$$

Eenvoudig kan worden aangetoond dat deze uitdrukking equivalent is aan de in de literatuur zeer bekende kubische interpolatieformule van Davidon [2.2.9].

$$\hat{\alpha} = \alpha_1 + \left(1 - \frac{h'(\alpha_2) + w - z}{h'(\alpha_2) - h'(\alpha_1) + 2w}\right)(\alpha_2 - \alpha_1) \qquad (2.2.25.a)$$

waarin als boven

$$z = h'(\alpha_2) + h'(\alpha_1) - 3(h(\alpha_2) - h(\alpha_1))$$

$$w = \sqrt{z^2 - h'(\alpha_1)h'(\alpha_2)} \; . \qquad (2.2.25.b)$$

Ook bij het gebruik van deze kubische interpolatie wordt eerst een interval bepaald waarbinnen het minimum ligt. Anders dan bij de kwadratische interpolatie is dit noodzakelijk vanwege de gemaakte veronderstelling $h'(\alpha_1) < 0$ en $h'(\alpha_2) > 0$. Op dezelfde wijze als bij kwadratische interpolatie kan men met de eerste schatting van het minimum genoegen nemen dan wel door het herhalen van de interpolatie na weglating van een van de punten de schatting van het minimum nauwkeuriger maken.

Onnauwkeurige lijnminimalisering

2.2.10. Toepassing van lijnminimalisering in een numerieke optimaliseringsprocedure impliceert in het algemeen in theorie een oneindig iteratieproces in iedere stap van een op zichzelf eveneens reeds oneindige iteratieprocedure. In de praktijk is exacte lijnminimalisering bij niet-lineaire optimaliseringsproblemen dan ook niet mogelijk en moet in iedere stap gebruik worden gemaakt van een benadering van het lijnminimum. Des te nauwkeuriger deze benadering, des te kleiner is het over het algemeen het totaal aantal benodigde iteraties. Afhankelijk van de minimaliseringsmethode maakt men in de praktijk gebruik van met zo weinig mogelijk functie-evaluaties te bepalen benaderingen van het lijnminimum, die slechts een beperkte verhoging van het totaal aantal iteraties nodig maken. De vraag waar in dit verband het optimale compromis ligt is op dit moment nog volop in onderzoek.

2.2.11. Ook bij theoretische beschouwingen over niet-lineaire optimaliseringsalgorithmen wordt met nauwkeurige lijnminimalisering rekening gehouden. (zie [2.2.6],[2.2.10]). In het bijzonder worden bijvoorbeeld convergentie uitspraken gedaan over algorithmen waarin in plaats van lijnminimalisering gebruik wordt gemaakt van procedures die alleen maar een punt langs de lijn bepalen waar de functiewaarde lager is. Een bekende procedure in dit verband is die van Goldstein en Price [2.2.11] die een punt $x^{(k)} + \hat{\alpha}d^{(k)}$ langs de lijn $d^{(k)}$ bepalen dat voldoet aan de voorwaarde

$$\varepsilon \leq \frac{h(0) - h(\hat{\alpha})}{-h'(0)\hat{\alpha}} \leq 1 - \varepsilon \qquad (2.2.26.a)$$

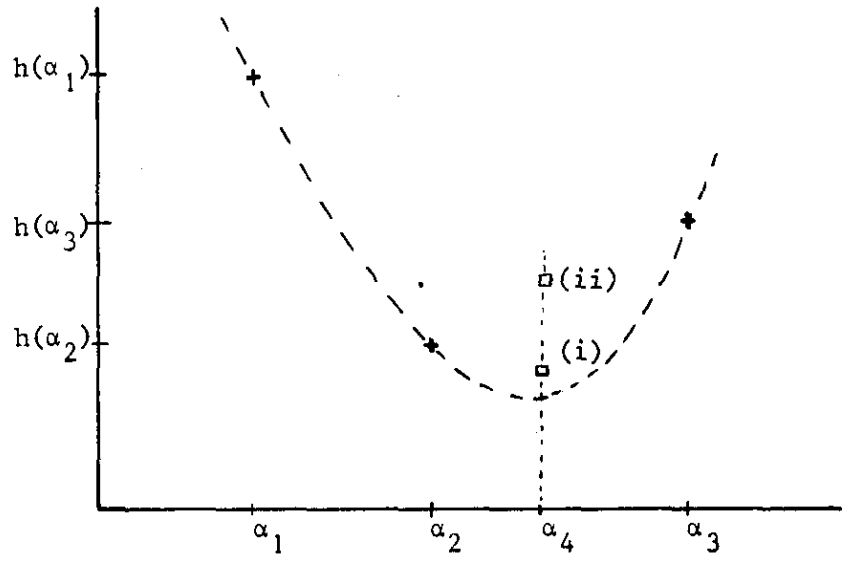
of equivalent

$$\varepsilon\hat{\alpha}\nabla^T f(x^{(k)})d^{(k)} \geq f(x^{(k)} + \hat{\alpha}d^{(k)}) - f(x^{(k)}) \geq (1 - \varepsilon)\hat{\alpha}\nabla^T f(x^{(k)})d^{(k)} \qquad (2.2.26.b)$$

waar $0 < \varepsilon < 0{,}5$. Bij deze voorwaarde die er van uit gaat dat voldaan is aan de voorwaarde $h'(0) = \nabla^T f(x^{(k)})d^{(k)} < 0$ zorgt de rechter ongelijkheid er voor dat $\hat{\alpha}$ van nul verschilt, terwijl de linker ongelijkheid voorkomt dat $\hat{\alpha}$ te groot wordt.

Globale convergentie van lijnminimaliseringsprocedures

2.2.12. Interessant is de vraag of de Globale Convergentiestelling van Zangwill (Stelling 2.1.11) ook van toepassing is op de lijnminimaliseringsprocedures. Het antwoord op deze vraag luidt bevestigend mits een aantal kleine en voor de praktijk niet belangrijke modificaties worden aangebracht. Als voorbeeld diene de lijnminimaliseringsprocedure met behulp van kwadratische interpolatie uitgaande van de situatie waar het lijnminimum is ingesloten (zie figuur en pt. 2.2.8).



Met drie punten α_1, α_2 en α_3 als uitgangspunten wordt een vierde punt α_4 bepaald als het punt waar de benaderingsparabool zijn minimale waarde aanneemt. Afhankelijk van de functiewaarde $h(\alpha_4)$ in α_4 wordt het volgende drietal punten dat het lijnminimum insluit (in het geval van de in de figuur geschetste situatie waar $h(\alpha_3) < h(\alpha_1)$)

(i) $(\alpha_2,\alpha_4,\alpha_3)$ als $h(\alpha_4) \leq h(\alpha_2)$

(ii) $(\alpha_1,\alpha_2,\alpha_4)$ als $h(\alpha_4) > h(\alpha_2)$

Deze overgang van het ene drietal punten naar het volgende drietal kan worden opgevat als een afbeelding van $\mathbf{R}^3$ naar $\mathbf{R}^3$, d.i.

$$A\begin{pmatrix}\alpha_1\\ \alpha_2\\ \alpha_3\end{pmatrix} = \begin{pmatrix}\alpha_2\\ \alpha_4\\ \alpha_3\end{pmatrix} \quad \text{als } h(\alpha_4) \leq h(\alpha_2) \qquad (2.2.27.a)$$

en

$$A\begin{pmatrix}\alpha_1\\ \alpha_2\\ \alpha_3\end{pmatrix} = \begin{pmatrix}\alpha_1\\ \alpha_2\\ \alpha_4\end{pmatrix} \quad \text{als } h(\alpha_4) > h(\alpha_2) \tag{2.2.27.b}$$

Om deze afbeelding A volledig te definiëren worden de volgende aanvullende afspraken gemaakt: Als twee α's aan elkaar gelijk zijn dan wordt het nieuwe punt bepaald als het minimale punt van de interpolatie parabool gebaseerd op de functiewaarden in de twee verschillende punten en de afgeleide in het dubbele punt. Als drie α's aan elkaar gelijk worden dan wordt het nieuwe punt bepaald als het minimum van de interpolatie parabool gebaseerd op de functiewaarde, de eerste en de tweede afgeleide in het drie dubbele punt. De aldus gedefinieerde algorithme A is een continue afbeelding van $\mathbf{R}^3 \rightarrow \mathbf{R}^3$ en voldoet als zodanig aan de in de Globale Convergentiestelling van Zangwill gestelde eisen. In de veronderstelling dat de functie $h(\alpha)$ unimodaal is komt als oplossingsverzameling Γ in aanmerking de verzameling

$$\Gamma := \{(\alpha^*,\alpha^*,\alpha^*)\} \tag{2.2.28}$$

d.i. de verzameling bestaande uit het ene element $(\alpha^*,\alpha^*,\alpha^*)$ waar α^* het punt is waar de functie $h(\alpha)$ zijn minimale waarde aanneemt. Voor daalfunctie van A en Γ kan de functie

$$Z(\alpha) = h(\alpha_1) + h(\alpha_2) + h(\alpha_3) \tag{2.2.29}$$

worden genomen. Onder de voorwaarde dat $h(\alpha)$ unimodaal is, is deze functie monotoon dalend voor opvolgende elementen van de door de hierboven gedefinieerde algorithme A gegenereerde rij. Omdat alle volgende punten α_4 liggen in het insluitingsinterval wordt ook aan de laatste eis van de Globale Convergentiestelling, namelijk dat alle elementen van de door de algorithme gegenereerde rij liggen in een compacte verzameling, voldaan. In het geval dat de functie $h(\alpha)$ unimodaal is binnen een initiëel insluitingsinterval garandeert de Globale Convergentiestelling van Zangwill dus globale convergentie van de lijnminimaliseringsprocedure gebaseerd op kwadratische interpolatie.

Geslotenheid van lijnzoekalgorithmen

2.2.13. In de meeste descent algorithmen worden lijnzoekprocedures gebruikt in iedere stap van het iteratieproces. Het volledige optimaliseringsalgorithme kan daarom in veel gevallen worden opgebouwd gedacht uit de samenstelling van een algorithme $G : \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^n$ voor het genereren van de zoekrichting

$$G(x) := (x,d) \tag{2.2.30}$$

en een lijnzoek algorithme $S : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ gegeven door, bijvoorbeeld,

$$S(x,d) := \{y \mid y = x + \hat{\alpha}d,\ f(y) = \min_{\alpha} f(x + \alpha d)\} \tag{2.2.31}$$

Van belang voor het toepassen van de globale convergentie theorie is de vraag of de aldus gedefinieerde lijnzoekalgorithme gesloten is. Dit blijkt in de veronderstelling dat $f(x)$ een continue functie op $\mathbb{R}^n$ is die een minimum heeft op de lijn $x + \alpha d$ voor $\alpha \geq 0$ relatief eenvoudig te kunnen worden aangetoond indien voldaan wordt aan de voorwaarde dat $d \neq 0$. Voor het bewijs beschouwt men dan de rijen $\{x^{(k)}\}$, $\{d^{(k)}\}$ en $\{y^{(k)}\}$ met $x^{(k)} \to x$, $d^{(k)} \to d$, $y^{(k)} \in S(x^{(k)}, d^{(k)})$ en $y^{(k)} \to y$. Voor iedere k geldt $y^{(k)} = x^{(k)} + \alpha^{(k)} d^{(k)}$ voor zekere $\alpha^{(k)}$ waar als $d^{(k)} \neq 0$ $\alpha^{(k)} = \|y^{(k)} - x^{(k)}\| / \|d^{(k)}\|$. Omdat de rijen $\{y^{(k)}\}$, $\{x^{(k)}\}$ en $\{d^{(k)}\}$ convergeren zal ook de rij $\{\alpha^{(k)}\}$ convergeren en zal gelden $\lim \alpha^{(k)} = \bar{\alpha}$ en $y = x + \bar{\alpha}d$. Voor iedere k geldt dat $f(y^{(k)}) \leq f(x^{(k)} + \alpha d^{(k)})$ voor alle $\alpha \geq 0$ en derhalve zal ook in de limiet $f(y) \leq f(x + \alpha d)$ voor alle $\alpha \geq 0$ en in het bijzonder dus $f(y) \leq \min_{\alpha} f(x + \alpha d)$. Dit laatste impliceert dat $y \in S(x,d)$ en dat de lijnzoek algorithme (2.2.31) dus gesloten is als $\|d\| \neq 0$.

N.B. Met een analoge argumentatie kan ook worden aangetoond dat de lijnzoekprocedure van Goldstein en Price (vgl. pt. 2.2.11) eveneens gesloten is onder de voorwaarde dat $d \neq 0$.

Voorbeeld 2.2.13.

Een voorbeeld, waaruit blijkt dat de voorwaarde $d \neq 0$ essentieel is voor het gesloten zijn van de lijnzoekprocedure (2.2.31) wordt geleverd door de functie $f(x) = (x^2 - 1)^2$. Hiervoor geldt immers $\min\{f(0 + \alpha d) \mid \alpha \geq 0\} = f(1)$ als $d > 0$ en $\min\{f(0 + \alpha 0) \mid \alpha \geq 0\} = f(0)$ voor $d = 0$ waarmee $S(0,d) = 1$ als $d > 0$ en $S(0,d) = 0$ als $d = 0$.

Referenties

2.2.14. Meer informatie over de in deze paragraaf besproken onderwerpen is te vinden in de volgende referenties:

[2.2.1]: Zie [1.1.1] Luenberger (1973).

[2.2.2]: Zie [1.1.2] Jacoby, Kowalik & Pizzo (1972).

[2.2.3]: Zie [1.1.3] Murray (1972).

[2.2.4]: Zie [1.1.6] Walsh (1975).

[2.2.5]: Zie [2.1.3] Zangwill (1969).

[2.2.6]: Zie [2.1.5] Polak (1971).

[2.2.7]: Kowalik, J. and Osborne, M.R.: "Methods for unconstrained optimization problems",American Elsevier Publ. Co., New York (1968).

[2.2.8]: Box, M.J. Davies, D., and Swann, W.H.: "Non-linear optimization techniques", ICI Monograph nr. 5, Oliver & Boyd, Edinburgh (1969).

[2.2.9]: Davidon, W.C. "Solution to Problem 74.3: "Davidon's cubic interpolation",SIAM Review, 17 (1975). p. 170-171.

[2.2.10]: Powell, M.J.D.: "Recent advances in unconstrained optimization, Math. Progr., 1 (1971), p.p. 26-57.

[2.2.11]: Goldstein, A.A. and Price, J.F.: "An effective algorithm for minimization",Numer. Math. 10 (1967), p.p. 184-189.

§ 2.3. "Direct search" methoden

2.3.1. Met de naam "direct search" methoden worden die methoden aangeduid waarvan de strategie uitsluitend gebaseerd is op de kennis van de waarden van de functie in opvolgend geëvalueerde punten. Nieuwe zoekrichtingen worden óf vooraf vastgelegd of worden bepaald op grond van de ervaring met de functiewaarden in eerder geëvalueerde punten.

2.3.2. Een van de meest voor de hand liggende "direct search" methode is de "alternating variable methode". Bij deze methode worden de coördinaatrichtingen telkens opnieuw als zoekrichtingen gebruikt en wordt de staplengte bepaald door lijnminimalisering. In de practijk blijkt deze methode zeer inefficiënt uit het oogpunt van het totaal aantal benodigde functie-evaluaties voor het bepalen van het optimum met een gegeven nauwkeurigheid.

2.3.3. Een efficiëntere "direct search" methode is de "pattern search" methode van Hooke en Jeeves [2.3.5]. Het idee achter deze methode is dat nieuwe punten worden gegenereerd op twee verschillende manieren. Ten eerste door middel van een zgn. "exploratory move", d.i. een serie kleine stappen langs de coördinaatrichtingen die tot doel hebben de richting te bepalen waarlangs de functie het meest afneemt en ten tweede via een zgn. "pattern move" d.i. een "grote" stap in de bij de "exploratory move" bepaalde richting. Het punt van waaruit de stap in de "pattern move" wordt gezet wordt een basispunt genoemd. De methode heeft als voordeel dat zij bijzonder eenvoudig te programmeren is en daarbij weinig geheugenruimte eist. In detail kan de algorithme van de "pattern search" methode als volgt worden weergegeven:

Algorithme voor de "pattern search" methode van Hooke en Jeeves

a) Exploratory move

(0) bepaal de functiewaarde $f(x^{(1)})$ in het startpunt $x^{(1)}$, van de "exploratory move" en zet $j := 1$

(i) zet $x^{(j+1)} := x^{(j)} + e_j\Delta$ (waar e_j de j-de eenheidsvector voorstelt) en evalueer $f(x^{(j+1)})$

(ii) als $f(x^{(j+1)}) \geq f(x^{(j)})$ dan zet $x^{(j+1)} := x^{(j)} - e_j\Delta$ en evalueer $f(x^{(j+1)})$

(iii) als $f(x^{(j+1)}) < f(x^{(j)})$ en $j < n$ dan zet $j := j + 1$ en ga terug naar (i)

(iv) als $f(x^{(j+1)}) \geq f(x^{(j)})$ en $j < n$ dan zet $x^{(j+1)} := x^{(j)}$, $j := j + 1$ en ga terug naar (i)

(v) als $j = n$ dan zet $x_B^{(k+1)} := x^{(j+1)}$ en de "exploratory move" is klaar.

b) Pattern move

(i) zet $\bar{x}_p := x_B^{(k+1)} + (x_B^{(k+1)} - x_B^{(k)})$.

c) Totaal algorithme

(0) zet $x_B^{(0)} := x^{(0)}$ en zet $k := 0$

(i) voer een "exploratory move" uit vanuit het basispunt $x_B^{(k)}$ en vind een nieuw basispunt $x_B^{(k+1)}$

(ii) als $x_B^{(k+1)} = x_B^{(k)}$ verklein Δ en als $\Delta < \varepsilon$ dan klaar, zo niet dan ga terug naar (i)

(iii) voer een "pattern move" uit vanuit $x_B^{(k+1)}$ en vindt $\bar{x}_p$

(iv) voer een "exploratory move" uit vanuit $\bar{x}_p$ en vind $x_B^{(k+2)}$

(v) evalueer $f(x_B^{(k+2)})$

(vi) als $f(x_B^{(k+2)}) > f(x_B^{(k+1)})$ dan zet $k := k + 1$ en ga terug naar (i)

(vii) als $f(x_B^{(k+2)}) \leq f(x_B^{(k+1)})$ dan zet $k := k + 2$ en ga terug naar (i).

2.3.4. Een tweede belangrijke "direct search" methode is de methode van Rosenbrock [2.3.6]. Deze methode kan worden opgevat als een verdere ontwikkeling van de methode van Hooke en Jeeves en verschilt van deze daarin dat in plaats van een "pattern move" na de "exploratory move" een heroriëntatie plaats vindt van het orthogonale assenstelsel waarlangs de stappen in de "exploratory move" worden gezet. Deze heroriëntatie vindt plaats door toepassing van een Gram-Schmidt orthogonalisatie-procedure op een stelsel lineair onafhankelijke vectoren, waarvan de eerste gericht is in de richting van de totale stap gezet in de voorgaande iteratieslag. De op dit idee gebaseerde algorithme luidt in detail als volgt:

Algorithme voor de methode van Rosenbrock

(0) ga uit van een startpunt $x_B^{(k)}$, een rij staplengten $\{s_1^{(k)},\ldots,s_n^{(k)}\}$ en een orthonormaal stelsel vectoren $\{d_1^{(k)},d_2^{(k)},\ldots,d_n^{(k)}\}$, zet $j := 1$, $x^{(1)} := x_B^{(k)}$ en evalueer $f(x^{(1)})$

(i) zet $x^{(j+1)} := x^{(j)} + s_j^{(k)} d_j^{(k)}$ en evalueer $f(x^{(j+1)})$

(ii) als $f(x^{(j+1)}) \leq f(x^{(j)})$, zet $s_j^{(k)} := \alpha s_j^{(k)}$ noteer een "succes" en ga naar (iv)

(iii) als $f(x^{(j+1)}) > f(x^{(j)})$, zet $s_j^{(k)} := -\beta s_j^{(k)}$, $x^{(j+1)} := x^{(j)}$, noteer een "failure" en ga naar (iv)

(iv) als $j < n$, zet $j := j + 1$ en ga terug naar (i)

(v) als $j = n$, check of in ieder van de n coördinaatrichtingen een "succes" gevolgd door een "failure" heeft plaatsgevonden, zo ja ga naar (vi), zo nee zet $j := 1$, $x^{(1)} := x^{(n)}$ en ga terug naar (i)

(vi) einde van de cycle: zet $x_B^{(k+1)} := x^{(n)}$ en bepaal de coëfficiënten $\lambda_i^{(k)}$ in de vergelijking

$$x_B^{(k+1)} - x_B^{(k)} = \sum_{i=1}^{n} \lambda_i^{(k)} d_i^{(k)}$$

(vii) bepaal de vectoren

$$v_j^{(k)} := \sum_{i=j}^{n} \lambda_i^{(k)} d_i^{(k)}$$

en pas een Gram-Schmidt orthonormalisatie procedure toe op de verzameling van lineair onafhankelijke vectoren $\{v_1^{(k)}, v_2^{(k)}, \ldots, v_n^{(k)}\}$ voor het genereren van een nieuwe orthonormale set $\{d_1^{(k+1)}, d_2^{(k+1)}, \ldots, d_n^{(k+1)}\}$:

$$d_i'^{(k+1)} := v_i^{(k)} - \sum_{j=1}^{i-1} (v_i^{(k)T} d_j^{(k+1)})\, d_j^{(k+1)}$$

$$d_i^{(k+1)} := d_i'^{(k+1)} / \| d_i^{(k+1)} \|$$

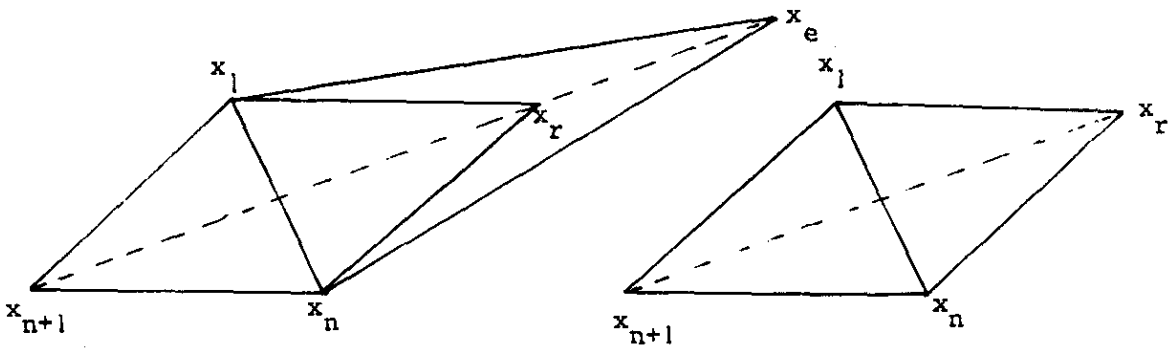
(viii) zet $k := k + 1$, kies een nieuwe rij staplengten $\{s_1^{(k)}, s_2^{(k)}, \ldots, s_n^{(k)}\}$, en ga terug naar (i).

Rosenbrock suggereert voor de constanten α en β de waarden $\alpha = 3$ en $\beta = 0.5$. Het effect van deze waarden is de aanpassing van de staplengte aan het optreden van "successen" en "failures" in voorgaande stappen. In de originele algorithme werd door Rosenbrock niet voorzien in een stopcriterium.

2.3.5. Een bekende variant van de methode van Rosenbrock is de Davies-Swann-Campey-methode ofwel kort D.S.C. methode [2.3.2]. Deze methode verschilt van die van Rosenbrock doordat van een eenvoudige lijnminimalisering gebruik wordt gemaakt bij de "exploratory move", in plaats van van een vector van staplengten. Ook geldt voor deze methode dat slechts eenmaal langs ieder van de orthogonale zoekrichtingen een lijnminimum bepaald wordt, waarna direct een heroriëntatie plaats vindt. Die zoekrichtingen waarlangs een stap van lengte nul gezet werd bij de "exploratory move" worden bij deze heroriëntatie ongemoeid gelaten.

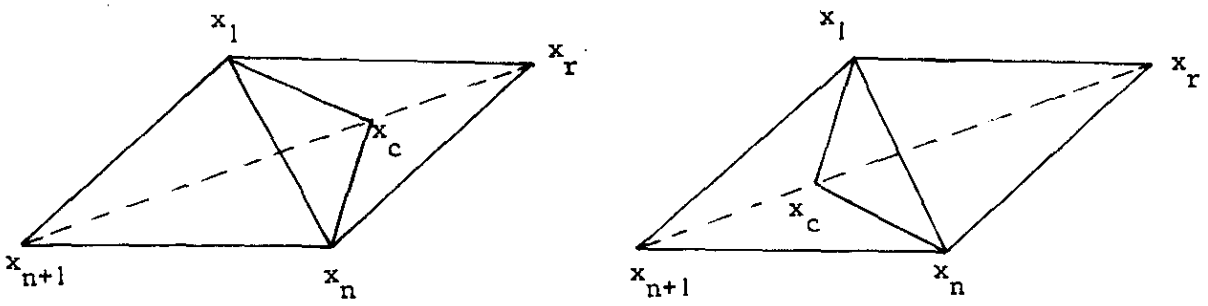
2.3.6. Een ander type "direct search" methode is de simplexmethode van Nelder en Mead [2.3.7]. Bij deze methode, die een verdere ontwikkeling betekent van de (reguliere) simplex methode van Hext, Himsworth en Spendley (zie [2.3.2], [2.3.4]), wordt gebruik gemaakt van het geometrische concept van de simplex. Hieronder wordt verstaan een verzameling van n + 1 punten in $\mathbb{R}^n$ welke verzameling niet ligt in een echte deelruimte van $\mathbb{R}^n$. Dus in $\mathbb{R}^2$ een driehoek, in $\mathbb{R}^3$ een tetraheder, etc. (Men spreekt van een reguliere simplex indien de afstand tussen alle punten gelijk is.) Begonnen wordt met de evaluatie van de objectfunctie in alle hoekpunten van de simplex. Daarna wordt in principe een nieuw punt gegenereerd door het punt met de hoogste objectfunctiewaarde te spiegelen t.o.v. het hypervlak opgespannen door de andere hoekpunten. In dit geval spreekt men van een "reflectie". Is de functiewaarde in het nieuwe punt lager dan de functiewaarden in alle originele hoekpunten van de originele simplex dan wordt de functie nogmaals geëvalueerd voor een tweede punt op de verbindingslijn tussen het originele en het gespiegelde punt en wel in een punt dat voorbij dat laatste ligt. Deze operatie wordt aan geduid als "expansie". Is de functiewaarde in het gereflecteerde punt groter dan de op een na hoogste functiewaarde gevonden voor de originele simplex dan wordt de functie eveneens nogmaals geëvalueerd voor een punt op de eerder genoemde verbindingslijn, en wel in een punt tussen het gereflecteerde punt en het punt van spiegeling als de functiewaarde in het gereflecteerde punt kleiner is dan in het originele te spiegelen punt en in een punt tussen het originele te spiegelen punt en het punt van spiegeling als de functiewaarde in het gereflecteerde punt hoger is dan in het originele te spiegelen punt. (Zie figuur).

Is de functiewaarde in dat punt ook hoger dan de op een na hoogste functiewaarde gevonden voor de originele simplex dan wordt overgegaan op een simplex van kleinere afmetingen door de afstanden van alle punten tot het punt van de originele simplex met de kleinste functiewaarden te verkleinen. In dit geval spreekt men van een "contractie". De procedure wordt gestopt indien de standaarddeviatie van de functiewaarden in de hoekpunten van de simplex kleiner wordt dan de gewenste nauwkeurigheid. Uitwerking van deze ideeën voor twee variabelen is gegeven in de volgende figuur



$f(x_r) < f(x_l)$: expansie

$f(x_l) < f(x_r) < f(x_n)$: reflectie



$f(x_n) < f(x_r) < f(x_{n+1})$: contractie

$f(x_{n+1}) < f(x_r)$: contractie

De uiteindelijke algorithme luidt als volgt:

Algorithme voor de simplexmethode van Nelder en Mead

(0) bepaal de functiewaarden in de hoekpunten $x_1^{(0)}, x_2^{(0)}, \ldots, x_{n+1}^{(0)}$ van een startsimplex en zet $k := 1$

(i) hernummer de verzameling hoekpunten $\{x_1^{(k-1)}, \ldots, x_{n+1}^{(k-1)}\}$ zodat de verzameling $\{x_1^{(k)}, \ldots, x_{n+1}^{(k)}\}$ ontstaat met de eigenschap dat $f(x_1^{(k)}) \leq f(x_2^{(k)}) \leq \ldots \leq f(x_{n+1}^{(k)})$ en bepaal het zwaartepunt

$$c^{(k)} := \frac{1}{n} \sum_{j=1}^{n} x_j^{(k)}$$

in het (hyper) zijvlak van de simplex tegenover het hoekpunt $x_{n+1}^{(k)}$

(ii) bepaal een nieuw punt $x_r^{(k)}$ door reflectie van $x_{n+1}^{(k)}$ t.o.v. het punt $c^{(k)}$:

$$x_r^{(k)} := c^{(k)} + \alpha(c^{(k)} - x_{n+1}^{(k)}), \qquad \alpha > 0$$

(iii) als $f(x_1^{(k)}) \leq f(x_r^{(k)}) \leq f(x_n^{(k)})$ dan zet $x_{n+1}^{(k)} := x_r^{(k)}$ en ga naar (i)

(iv) als $f(x_r^{(k)}) < f(x_1^{(k)})$ dan genereer een tweede nieuw punt door expansie t.o.v. het punt $c^{(k)}$:

$$x_e^{(k)} := c^{(k)} + \gamma(x_r^{(k)} - c^{(k)}), \quad \gamma > 1$$

als daarna $f(x_e^{(k)}) < f(x_r^{(k)})$ dan zet $x_{n+1}^{(k)} := x_e^{(k)}$ en ga naar (i), zo niet, dan zet $x_{n+1}^{(k)} := x_r^{(k)}$ en ga terug naar (i)

(v) als $f(x_r^{(k)}) > f(x_{n+1}^{(k)})$ dan genereer een tweede nieuw punt door contractie van het punt $x_{n+1}^{(k)}$ t.o.v. het punt $c^{(k)}$

$$x_c^{(k)} := c^{(k)} + \beta(x_{n+1}^{(k)} - c^{(k)}), \quad 0 < \beta < 1$$

als daarna $f(x_c^{(k)}) < f(x_{n+1}^{(k)})$ dan zet $x_{n+1}^{(k)} := x_c^{(k)}$ en ga naar (i), zo niet, dan ga naar (vii)

(vi) als $f(x_n^{(k)}) < f(x_r^{(k)}) \leq f(x_{n+1}^{(k)})$ dan genereer een tweede nieuw punt door contractie van het punt $x_r^{(k)}$ t.o.v. het punt $c^{(k)}$

$$x_c^{(k)} := c^{(k)} + \beta(x_r^{(k)} - c^{(k)}), \quad 0 < \beta < 1$$

als daarna $f(x_c^{(k)}) < f(x_r^{(k)})$ dan zet $x_{n+1}^{(k)} := x_c^{(k)}$ en ga naar (i), zo niet, dan zet $x_{n+1}^{(k)} := x_r^{(k)}$ en ga naar (vii)

(vii) krimp de simplex in door de hoekpunten $x_j^{(k)}$, $j = 1,\ldots,n+1$ te vervangen door de hoekpunten

$$x_j^{(k+1)} := \tfrac{1}{2}(x_j^{(k)} + x_1^{(k)}), \qquad j = 1,\ldots,n+1$$

(viii) evalueer de functie in de nieuwe hoekpunten en bereken de "steekproefvariantie"

$$(s^{(k+1)})^2 = \frac{1}{n}\sum_{i=1}^{n+1}[f(x_i^{(k+1)}) - \frac{1}{n+1}\sum_{i=1}^{n+1} f(x_i^{(k+1)})]^2$$

als $s < \varepsilon$ dan klaar, zo niet dan zet $k := k + 1$ en ga terug naar (i).

Als waarden voor de reflectiecoëfficiënt α, de contractiecoëfficiënt β en de expansiecoëfficiënt γ suggereren Nelder en Mead zelf de waarden

$$\alpha = 1 \ , \quad \beta = \tfrac{1}{2} \quad \text{en} \quad \gamma = 2$$

Met deze waarden blijkt de algorithme een van de meest efficiënte "direct search" algorithmen.

2.3.7. Een bijzondere categorie "direct search" methoden wordt gevormd door de zgn. "random search" methoden [2.3.8]. Hieronder worden verstaan die "direct search" methoden waarbij de volgende punten waar de functie wordt geëvalueerd tijdens het zoekproces mede worden bepaald door het toeval. De methoden waarbij de nieuwe punten uitsluitend of nagenoeg uitsluitend door het toeval worden bepaald blijken bijzonder inefficiënt. Beter ligt die situatie met betrekking tot die "random search" methoden waarbij de nieuwe punten voor een groot gedeelte worden bepaald met behulp van de ervaring opgedaan op een vroeger tijdstip in het zoekproces en slechts voor een klein gedeelte door het toeval. Als voorbeeld kan worden genoemd de algorithme van Matyas waarbij een nieuw punt wordt bepaald met behulp van de uitdrukking

$$x^{(i+1)} := x^{(i)} + d^{(i)} + T^{(i)}z^{(i)}$$

waarin

$$d^{(i)} := c_0 d^{(i-1)} + c_1(x^{(i)} - x^{(i-1)})$$

$z^{(i)}$:= realisatie van een n-dimensionale stochastische vector variabele waarvan de componenten een verdeling N(0,1) bezitten

$T^{(i)}$:= $n \times n$ matrix voor de introductie van correlaties tussen de componenten

en waarin de coëfficiënten c_0 en c_1 afhangen van de omstandigheid of in de laatste stap een "succes" ($f(x^{(i)}) < f(x^{(i-1)})$) of een "failure" werd ondervonden, t.w.:

a) bij een "succes" in de laatste stap:

$$0 < c_0 < 1 \qquad c_1 > 0 \qquad c_0 + c_1 > 1$$

b) bij een "failure" in de laatste stap

$$0 < c_0 < 1 \qquad c_1 < 0 \qquad |c_0 + c_1| < 1 \ .$$

Uiteraard zijn er vele strategieën op deze manier mogelijk. Uit de literatuur blijkt dat voor sommige toepassingen, zoals bijvoorbeeld in de gevallen dat

a) de functiewaarden slechts bij benadering bepaald kunnen worden als gevolg van stochastische fouten
b) men te maken heeft met veel variabelen en eenvoudige functies
c) het gaat om het bepalen van een eerste niet noodzakelijk zeer nauwkeurige schatting van het optimum.

"random search" methoden zelfs relatief zeer efficiënt kunnen zijn.

2.3.8. Resumerend kunnen als voordelen van "direct search" methoden worden genoemd

1) de algemeenheid van de methoden: alleen functiewaarden worden gebruikt
2) de toepasbaarheid op niet-differentieerbare functies en de relatieve ongevoeligheid voor kleine stochastische meet- of rekenfouten bij de evaluatie van de functies
3) de eenvoudige programmeerbaarheid en de veelal kleine benodigde geheugenruimte.

Als nadelen kunnen worden genoemd:

1) het relatief groot aantal functie-evaluaties door een inefficiënt gebruik van de verkregen informatie
2) de veelal langzame en niet gegarandeerde convergentie.

Referenties

2.3.9. Meer informatie over de in deze paragraaf behandelde onderwerpen is te vinden in de volgende referenties:

[2.3.1]: Zie [1.1.2] Jacoby, Kowalik & Pizzo (1972).

[2.3.2]: Zie [1.1.3] Murray (1972).

[2.3.3]: Zie [2.2.7] Kowalik & Osborne (1968).

[2.3.4]: Zie [2.2.8] Box, Davies, Swann (1969).

[2.3.5]: Hooke, R. and Jeeves, T.A.: "Direct search solutions of numerical and statistical problems". J. ACM. 8 (1961) pp. 221-229.

[2.3.6]: Rosenbrock, H.H.: "An automatic method for finding the greatest or least value of a function." Computer J, 3 (1960), pp. 175-184.

[2.3.7]: Nelder, J.A. and Mead, R.: "A simplex method for function minimization". Computer J, V (1965), pp. 308-313.

[2.3.8]: White, R.C.: "A survey of random methods for parameter optimization". THE-Report 70-E-16,(1970).

[2.3.9]: Gaviano, M.: "On the convergence of random search algorithms for minimization problems" in "Towards global optimization", (L.C.W. Dixon & G.P. Szegö, Eds.) North-Holland Publ. Co., Amsterdam (1975).

§ 2.4. Methode van de steilste helling en gradiënt-methoden

2.4.1. Een van de oudste klassen van methoden voor het minimaliseren van functies van meer variabelen is de klasse van methoden die gebruik maken van de locaal te bepalen gradiënt van de objectfunctie als (tegengestelde van de) zoekrichting. In termen van de standaardalgorithme (pt.2.1.5) betekent dit dat in het punt $x^{(k)}$ als zoekrichting wordt gekozen

$$d^{(k)} := -\nabla f(x^{(k)}) \tag{2.4.1}$$

en dat het volgende punt $x^{(k+1)}$ bepaald wordt uit

$$x^{(k+1)} := x^{(k)} - \alpha^{(k)} \nabla f(x^{(k)}) \ . \tag{2.4.2}$$

Wordt de staplengtefactor $\alpha^{(k)}$ bepaald door lijnminimalisering dan spreekt men van de methode van de steilste helling ofwel de "steepest-descent"-methode. Wordt deze staplengte factor op andere wijze bepaald dan spreekt men van een gradiënt-methode. (De methode van de steilste helling wordt soms ook aangeduid als de "optimale gradiënt-methode".) De eerste beschrijving van de methode van de steilste helling werd gegeven door Cauchy in 1847 (vandaar de soms gebruikte naam "Cauchy-methode"). In de meer recente tijd verkreeg de methode opnieuw bekendheid na een artikel van Curry [2.4.4].

2.4.2. De keuze van de negatieve gradiënt als zoekrichting wordt intuïtief ingegeven door de overweging dat indien men in de mist vanaf een berg naar het dal moet dat men het snelst dat dal zal bereiken indien men de richting neemt van de steilste helling (d.i. de gradiënt-richting). Wiskundig volgt de gradiënt-richting als oplossing van het probleem van de bepaling van het minimum van de locaal gelineariseerde functie (d.i. de lineaire benadering van de objectfunctie in het laatst gevonden punt) binnen een cirkel of (hyper)-bol met een gegeven (kleine) straal δ. Toepassing van de Kuhn-Tucker-voorwaarden op het minimaliseringsprobleem

$$\min\{f(x^{(k)}) + \nabla^T f(x^{(k)})(x - x^{(k)}) \mid \|x - x^{(k)}\|^2 \leq \delta^2\} \tag{2.4.3}$$

geeft onmiddellijk de oplossing

$$x^* - x^{(k)} := - \frac{\nabla f(x^{(k)})}{\|\nabla f(x^{(k)})\|} \delta. \tag{2.4.4}$$

2.4.3. Uitgeschreven levert de methode van de steilste helling de volgende algorithme (Andere gradiënt methoden verschillen slechts in stap (iv)).

Algorithme voor de methode van de steilste helling

(0) kies een startpunt $x^{(0)}$, zet k := 0

(i) bepaal de functiewaarde $f(x^{(k)})$ en de gradiënt $\nabla f(x^{(k)})$ in $x^{(k)}$

(ii) ga na of $x^{(k)}$ optimaal is, zo ja, dan klaar, zo nee, dan:

(iii) kies als zoekrichting

$$d^{(k)} := - \nabla f(x^{(k)})$$

(iv) bepaal een staplengte (-factor) uit

$$f(x^{(k)} + \alpha^{(k)} d^{(k)}) = \min\{f(x^{(k)} + \alpha d^{(k)}) \mid \alpha \in \mathbf{R}^1_+\}$$

(v) zet $x^{(k+1)} := x^{(k)} + \alpha^{(k)} d^{(k)}$ en k := k + 1 en ga terug naar stap (i).

2.4.4. Diverse auteurs geven stellingen voor de convergentie van de methode van de steilste helling en andere gradiënt methoden. Een goed voorbeeld daarvan is de volgende stelling van Goldstein [2.4.6].

Stelling 2.4.4. (Goldstein): Zij gegeven dat $f : \mathbb{R}^n \to \mathbb{R}^1$, $x^{(0)} \in \mathbf{R}^n$ $f(x^{(0)})$ gedefinieerd en $f \in C^2$ voor alle $x \in S$ waar

$$S := \{x \in \mathbf{R}^n \mid f(x) \leq f(x^{(0)})\} \qquad (2.4.5)$$

Als verder voldaan is aan de voorwaarden:

(i) f is naar beneden begrensd op S

(ii) $\| G(x) \| \leq M < \infty$ voor alle $x \in S$

(iii) $x^{(k+1)} := x^{(k)} - \alpha^{(k)} \nabla f(x^{(k)})$ waar $\alpha^{(k)}$ voldoet aan

ôf:

(α) $f(x^{(k)} - \alpha^{(k)} \nabla f(x^{(k)})) \leq f(x^{(k)} - \alpha \nabla f(x^{(k)}))$ voor $0 \leq \alpha < \infty$

ôf:

(β) $\delta \leq \alpha^{(k)} \leq 2/M - \delta$ waar $0 < \delta \leq 1/M$

dan geldt

a) de opvolgende punten $x^{(k)}$ van de rij $\{x^{(k)}\}$ behoren tot S, de rij van gradiënten $\{\nabla f(x^{(k)})\}$ convergeert naar 0 en de rij van functiewaarden $\{f(x^{(k)})\}$ convergeert naar een limiet.

b) in het geval dat S begrensd is geldt voor ieder verdichtingspunt $\bar{x}$ van $\{x^{(k)}\}$ dat $\nabla f(\bar{x}) = 0$. Is x^* het enige punt in S waar $\nabla f(x^*) = 0$ dan convergeert $\{x^{(k)}\}$ naar x^* en geldt $f(x^*) = \min\{f(x) \mid x \in S\}$.

c) als voor alle $y \in \mathbb{R}^n$ en alle $x \in S$ geldt $y^T G(x) y \geq m\|y\|^2$ met $m > 0$ en S convex is dan bestaat er een uniek minimum $f(x^*)$ van f in $x^* \in S$ waarnaar de rij $\{x^{(k)}\}$ convergeert.

Bewijs: ad a): Veronderstel $\nabla f(x^{(k)}) \neq 0$ en $x^{(k)} \in S$.
Definieer $x^{(k)}(\alpha) := x^{(k)} - \alpha \nabla f(x^{(k)})$ en $h^{(k)}(\alpha) := f(x^{(k)}(\alpha))$ en beschouw de functie

$$\Delta(x^{(k)};\alpha) = h^{(k)}(\alpha) - h^{(k)}(0) \qquad (2.4.6)$$

Omdat $\nabla f(x^{(k)}) \neq 0$ en $h^{(k)}(\alpha)$ differentieerbaar in $\alpha = 0$ met $h'^{(k)}(0) = -\|\nabla f(x^{(k)})\|^2$ geldt voor kleine $\alpha > 0$ dat $\Delta(x^{(k)};\alpha) < 0$ en daarmee dat $x^{(k)}(\alpha) \in S$. Met de middelwaardestelling volgt voor dergelijke α met $0 < \bar{\alpha} < \alpha$ en $y := \nabla f(x^{(k)}) / \|\nabla f(x^{(k)})\|$ dat

$$\Delta(x^{(k)};\alpha) = \|\nabla f(x^{(k)})\|^2 [-\alpha + \frac{\alpha^2}{2} y^T G(x^{(k)}(\bar{\alpha}))y]$$

$$\leq \|\nabla f(x^{(k)})\|^2 [-\alpha + \frac{\alpha^2 M}{2}]$$

waaruit volgt dat $\Delta(x^{(k)};\alpha) < 0$ voor $0 < \alpha < 2/M$. Wordt $\alpha^{(k)}$ gekozen volgens een van de voorschriften (iii α) of (iii β) dan geldt dat

$$\Delta(x^{(k)};\alpha^{(k)}) \leq \|\nabla f(x^{(k)})\|^2 [-\delta + \frac{\delta^2 M}{2}]$$

$$< 0 \quad \text{als } \nabla f(x^{(k)}) \neq 0.$$

Dit impliceert dat zolang $\nabla f(x^{(k)}) \neq 0$ geldt dat $f(x^{(k+1)}) < f(x^{(k)})$ en

derhalve dat $x^{(k+1)} \in S$. Zodra voor zekere $x^{(k)}$ geldt $\nabla f(x^{(k)}) = 0$ dan volgt dat $x^{(k+\ell)} = x^{(k)}$ en daarmee $x^{(k+\ell)} \in S$ voor $\ell \geq 0$. De rij $\{f(x^{(k)})\}$ is (zolang $\nabla f(x^{(k)}) \neq 0$) monotoon dalend en van onder begrensd en convergeert daarom naar een limiet. De absolute waarden van de functie $\Delta(x^{(k)};\alpha^{(k)})$ en de rij $\{\|\nabla f(x^{(k)})\|\}$ convergeren dienovereenkomstig naar 0.

Bewijs: ad b): Zij $\{x^{(k_i)}\}$ een deelrij die naar het verdichtingspunt $\bar{x}$ convergeert dan geldt op grond van de continuïteit van ∇f dat de rij $\{\nabla f(x^{(k_i)})\}$ convergeert naar $\nabla f(\bar{x})$ waar op grond van het voorgaande geldt $\nabla f(\bar{x}) = 0$. Als x^* het enige punt in S waar $\nabla f = 0$ dan convergeert de totale rij $\{x^{(k)}\}$ naar x^*. Immers zo niet dan zou er een tweede verdichtingspunt $z \neq x^*$ moeten zijn, waar, in tegenspraak tot de veronderstelling eveneens gold $\nabla f(z) = 0$. Omdat S compact en f continu is er een punt $\hat{x}$ waar het minimum van f wordt bereikt. In dat punt moet gelden $\nabla f(\hat{x}) = 0$ waaruit dus volgt dat $\hat{x} = x^*$.

Bewijs: ad c): De voorwaarde dat voor alle $x \in S$ en alle $y \in \mathbb{R}^n$ geldt dat $y^T G(x) y \geq m\|y\|^2$ met $m > 0$ impliceert dat f(x) strikt convex is en derhalve dat S begrensd is en dat er in S een uniek punt x^* is waar f(x) zijn minimum bereikt (zie bijvoorbeeld [2.4.11]). Toepassing van uitspraak b) completeert het bewijs. □

2.4.5. Convergentie van de methode van de steilste helling (en van de andere gradiënt methoden) kan in de veronderstelling dat de verzameling S (2.4.5) begrensd is ook worden aangetoond met de Globale Convergentiestelling (Stelling 2.1.11) van Zangwill. De door de algorithme van de steilste helling methode gegenereerde rij $\{x^{(k)}\}$ wordt daartoe opgevat als te zijn gegenereerd door de samengestelde algorithme (in de zin van Definitie 2.1.9.a)

$$A = SG \tag{2.4.7}$$

waar G de zoekrichtinggeneratiealgorithme is, die een continue afbeelding is van $\mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^n$ overeenkomstig de relatie

$$G(x) := (x, \nabla f(x)) \tag{2.4.8}$$

en S een lijnminimaliseringsalgorithme is die een (mogelijke) point-to-set mapping is van $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ overeenkomstig de relatie

$$S(x,d) := \{y \mid y = x + \hat{\alpha}d,\ f(y) = \min\{f(x + \alpha d) \mid 0 \leq \alpha < \infty\}\} \qquad (2.4.9)$$

Omdat S gesloten is als $d \neq 0$ en G continu volgt dat ook $A = SG$ gesloten is als $\nabla f(x) \neq 0$. Wordt dan als oplossingsverzameling gekozen de verzameling

$$\Gamma := \{x \in S \mid \nabla f(x) = 0\} \qquad (2.4.10)$$

dan voldoet algorithme A aan de in de Globale Convergentiestelling gestelde voorwaarden. Als daalfunctie voor Γ en A komt in aanmerking de objectfunctie f(x) zelf. Zolang $\nabla f(x^{(k)}) \neq 0$ geldt immers dat

$$f(x^{(k+1)}) = \min\{f(x^{(k)} - \alpha \nabla f(x^{(k)})) \mid 0 \leq \alpha \leq \infty\} < f(x^{(k)}) \qquad (2.4.11)$$

Met deze ongelijkheid en de veronderstelling dat S begrensd en daarmee compact is volgt dat de elementen van de rij $\{x^{(k)}\}$ in de verzameling S blijven, waarmee dan voldaan is aan alle voorwaarden voor convergentie op grond van de Globale Convergentiestelling.

N.B. Opgemerkt kan worden dat het in het voorgaande punt 2.4.4 gegeven bewijs van de convergentie van de methode van de steilste helling sterke overeenkomsten vertoont met de argumentatie om aan te tonen dat aan de voorwaarden van de Globale Convergentiestelling wordt voldaan en de eerder gegeven argumentatie (zie pt. 2.1.11) dat deze voorwaarden inderdaad voldoende zijn voor convergentie. Verondersteld mag worden dat het bewijs van de stelling van de convergentie van de methode van de steilste helling model heeft gestaan voor de Globale Convergentiestelling van Zangwill.

N.B. De hierboven gegeven convergentiestelling voor de methode van de steilste helling en andere gradiëntmethoden kan relatief eenvoudig worden gegeneraliseerd voor toepassingen van dezelfde methoden voor de minimalisering van functionalen gedefinieerd op algemenere ruimten zoals Hilbert en Banachruimten (zie bijvoorbeeld [2.4.6], [2.4.7] [2.4.8] en [2.4.9]).

2.4.6. Met betrekking tot de convergentiesnelheid van de methode van de steilste helling en de andere gradiëntmethoden worden door diverse auteurs min of meer gelijkluidende uitspraken gedaan. Een vereiste daarbij is dat extra eisen, zoals bijvoorbeeld convexiteit, aan de te minimaliseren functies worden opgelegd. Representanten van deze uitspraken zijn de volgende twee stellingen van Goldstein en Polak, waarvan de eerste een aanvulling is van de hierboven weergegeven stelling 2.4.4 en de tweede betrekking heeft op een wat algemenere klasse van descentalgorithmen.

Stelling 2.4.6. (Goldstein): Onder de voorwaarden i) ii) en iii β) en de extra veronderstelling in punt c) van Stelling 2.4.4 geldt voor de convergentiesnelheid van de daar gedefinieerde rij $\{x^{(k)}\}$ de afschatting

$$\|x^{(k+1)} - x^*\| \leq (1 - \delta m)\|x^{(k)} - x^*\| \qquad (2.4.12)$$

dat wil zeggen $\{x^{(k)}\}$ convergeert lineair met een convergentiefactor kleiner dan $1 - \delta m$.

Bewijs. (vergelijk [2.4.6]) De generatie van het punt $x^{(k+1)}$ uit $x^{(k)}$ kan worden opgevat als het resultaat van een afbeelding: $P^{(k)} : \mathbf{R}^n \to \mathbf{R}^n$ gedefinieerd door

$$P^{(k)}(y) = y - \alpha^{(k)}\nabla f(y) \qquad (2.4.13)$$

Onder de door iii β) gegarandeerde voorwaarde dat $\alpha^{(k)} > 0$ geldt in het minimale punt x^* dat $P^{(k)}(x^*) = x^*$, dat wil zeggen x^* is een vast punt van de afbeelding $P^{(k)}$. Er geldt daarom

$$\begin{aligned}
\|x^{(k+1)} - x^*\| &= \|P^{(k)}(x^{(k)}) - P^{(k)}(x^*)\| \\
&\leq \sup_{0\leq\sigma\leq 1}\{\|\nabla P^{(k)}(x^{(k)} - \sigma(x^{(k)} - x^*))\|\|x^{(k)} - x^*\|\} \\
&\leq \sup_{0\leq\sigma\leq 1}\{\|I - \alpha^{(k)}G(x^{(k)} - \sigma(x^{(k)} - x^*))\|\|x^{(k)} - x^*\|\} \\
&\leq (1 - \alpha^{(k)}m)\|x^{(k)} - x^*\| \\
&\leq (1 - \delta m)\|x^{(k)} - x^*\| \qquad (2.4.14.a)
\end{aligned}$$

in welke afleiding gebruik wordt gemaakt van de gegeneraliseerde middelwaardestelling voor differentieerbare afbeeldingen en van de eigenschap dat voor een symmetrische matrix A geldt

$$\|A\| = \sup\{|y^T Ay| \mid \|y\| = 1\} \qquad \square$$

N.B. De grootste waarde voor δ waarvoor stelling 2.4.4 in het geval van de gradiëntmethoden nog convergentie garandeert is $\delta = 1/M$.
De afschatting (2.4.14.a) gaat daarmee over in

$$\|x^{(k+1)} - x^*\| \leq (1 - \frac{m}{M})\|x^{(k)} - x^*\| \qquad (2.4.14.b)$$

voor de methode van steilste helling die het snelst convergeert van alle gradiëntmethoden zal deze afschatting steeds gelden.

2.4.7. Een uitspraak over de convergentiesnelheid van een wat algemenere klasse van descent methoden werd gegeven door Polak [2.4.3] in de volgende stelling.

Stelling 2.4.7.(Polak):Zij gegeven dat $f : \mathbb{R}^n \to \mathbb{R}^1$, $x^{(0)} \in \mathbb{R}^n$, $f(x^{(0)})$ gedefinieerd en $f \in C^2$ voor alle

$$x \in S := \{x \in \mathbb{R}^n \mid f(x) \leq f(x^{(0)})\}.$$

Als verder verondersteld wordt dat:

(i) $f(x)$ is strikt convex en voor alle $x \in S$ en $y \in \mathbb{R}^n$ geldt

$$m\|y\|^2 \leq \langle y, G(x)y\rangle^{\overline{\dagger}} \leq M\|y\|^2 \qquad 0 < m < M < \infty \qquad (2.4.15)$$

waar $G(x)$ de Hessiaan is $[(\frac{\partial^2 f}{\partial x_i \partial x_j})(x)]$ van $f(x)$ in x

(ii) $x^{(k+1)} := x^{(k)} + \alpha^{(k)} d^{(k)}$ waar $d^{(k)}$ voldoet aan

$$\langle \nabla f(x^{(k)}), d^{(k)}\rangle \leq -\rho\|\nabla f(x^{(k)})\|\|d^{(k)}\| \qquad 0 < \rho < 1 \qquad (2.4.16)$$

$\overline{\dagger}$ Voor een inwendig product van twee vectoren a en b in $\mathbb{R}^n$ wordt naast de notatie $a^T b$ ook gebruik gemaakt van de notatie $\langle a,b\rangle$.

en $\alpha^{(k)}$ aan

$$f(x^{(k)} + \alpha^{(k)}d^{(k)}) := \min\{f(x^{(k)} + \alpha d^{(k)}) \mid \alpha \in \mathbb{R}_+^1\}$$

dan geldt dat

a) de rij $\{x^{(k)}\}$ convergeert naar het unieke minimale punt x^* van $f(x)$ in S

b) voor de convergentiesnelheid van de rij $x^{(k)}$ geldt

$$\|x^{(k)} - x^*\| \leq \theta^k \sqrt{\frac{2}{m}[f(x^{(0)}) - f(x^*)]} \qquad (2.4.17.a)$$

waar

$$\theta = \sqrt{(1 - \frac{\rho^2 m^2}{M^2})} < 1. \qquad 2.4.17.b)$$

Bewijs: ad a): Dit resultaat volgt met wat kleine aanpassingen direct uit het bewijs van Stelling 2.4.4. ad b): Met behulp van de voor alle $x \in S$ en $t \in [0,1]$ te definiëren functies

$$h(t) := f(x^* + t(x - x^*)) - f(x^*) = \int_0^t h'(\tau)d\tau \qquad (2.4.18.a)$$

met afgeleiden

$$h'(t) = \nabla^T f(x^* + t(x - x^*))(x - x^*) = \int_0^t h''(\tau)d\tau \qquad (2.4.18.b)$$

$$h''(t) = (x - x^*)^T G(x^* + t\,(x - x^*))(x - x^*) \qquad (2.4.18.c)$$

en de ongelijkheid (2.4.15) kunnen de volgende algemene schattingen worden afgeleid

$$m\|x - x^*\|^2 t \leq h'(t) \leq M\|x - x^*\|^2 t \qquad (2.4.19)$$

$$\frac{1}{2}m\|x - x^*\|^2 \leq h(1) = f(x) - f(x^*) \leq \frac{1}{2}M\|x - x^*\|^2 \qquad (2.4.20)$$

Voor het minimum van de lijnfunctie

$$\Delta(x^{(k)};\alpha) = f(x^{(k)} + \alpha d^{(k)}) - f(x^{(k)})$$

waarvoor geldt (met $0 < \bar{\alpha} < \alpha$)

$$\Delta(x^{(k)};\alpha) = \alpha\nabla^T f(x^{(k)})d^{(k)} + \frac{1}{2}\alpha^2 d^{(k)T} G(x^{(k)} + \bar{\alpha}d^{(k)})d^{(k)}$$

$$\leq -\alpha\rho\|\nabla f(x^{(k)})\|\,\|d^{(k)}\| + \frac{1}{2}\alpha^2 M\|d^{(k)}\|^2$$

$$\leq \|d^{(k)}\|\left[-\alpha\rho\|\nabla f(x^{(k)})\| + \frac{1}{2}\alpha^2 M\|d^{(k)}\|\right]$$

volgt dat

$$\min\{\Delta(x^{(k)};\alpha)\,|\,\alpha \in \mathbb{R}^1\} \leq \|d\|\left[-\left(\frac{\rho\|\nabla f\|}{M\|d\|}\right)\rho\|\nabla f\| + \frac{1}{2}\left(\frac{\rho\|\nabla f\|}{M\|d\|}\right)^2 M\|d\|\right]$$

$$\leq -\frac{1}{2}\frac{\rho^2}{M}\|\nabla f(x^{(k)})\|^2$$

zodat

$$f(x^{(k+1)}) - f(x^{(k)}) \leq -\frac{1}{2}\frac{\rho^2}{M}\|\nabla f(x^{(k)})\|^2 \leq -\frac{1}{2}\frac{\rho^2}{M}m^2\|x^{(k)} - x^*\|^2$$

$$\leq -\frac{\rho^2 m^2}{M^2}\left(\frac{M}{2}\|x^{(k)} - x^*\|^2\right) \leq -\frac{\rho^2 m^2}{M^2}[f(x^{(k)}) - f(x^*)]$$

waaruit weer kan worden afgeleid

$$f(x^{(k+1)}) - f(x^*) \leq \left(1 - \frac{\rho^2 m^2}{M^2}\right)[f(x^{(k)}) - f(x^*)]$$

en, na herhaalde toepassing,

$$f(x^{(k+1)}) - f(x^*) \leq \left(1 - \frac{\rho^2 m^2}{M^2}\right)^{k+1}[f(x^{(0)}) - f(x^*)]$$

Met

$$\frac{m}{2}\|x^{(k+1)} - x^*\|^2 \leq f(x^{(k+1)}) - f(x^*)$$

volgt tenslotte de gevraagde ongelijkheid

$$\|x^{(k+1)} - x^*\|^2 \leq \frac{2}{m}\left(1 - \frac{\rho^2 m^2}{M^2}\right)^{k+1}[f(x^{(0)}) - f(x^*)]$$ □

2.4.8. Voor het geval van het minimaliseren van een positief definiete kwadratische vorm $f(x) = \frac{1}{2}x^T Ax + b^T x + c$ kunnen afschattingen voor de convergentiesnelheid van de methode van de steilste helling worden gegeven die scherper zijn dan de afschatting in Stelling 2.4.6 (met $\delta = 1/M$). De reden hiervoor is dat in dit geval analytische uitdrukkingen bestaan voor de opvolgende punten $x^{(k)}$. De scherpst mogelijke afschatting is die gegeven door Kantorovich [2.4.8], [2.4.9]

$$\| x^{(k)} - x^* \| \leq \frac{1}{m} \left(\frac{M - m}{M + m}\right)^k \| \nabla f(x^{(0)}) \| \qquad (2.4.21)$$

waarin M en m resp. de grootste en de kleinste eigenwaarden van de positief definiete matrix A voorstellen. Deze afschatting kan worden afgeleid met behulp van de zg. Kantorovich (-Bergstrom) ongelijkheid

$$\frac{\langle y,y\rangle^2}{\langle y,Ay\rangle \langle y,A^{-1}y\rangle} \geq \frac{4Mm}{(M+m)^2} . \qquad (2.4.22)$$

Een simpel bewijs hiervan is o.a. te vinden in [2.4.1] en [2.4.10].

2.4.9. Als voor- en nadelen van de methode van de steilste helling en de gradiëntmethode kunnen respectievelijk worden genoemd:

a) Voordelen:

1) altijd verlaging van de functiewaarde in opvolgende punten waardoor onder ruime voorwaarden theoretische convergentie kan worden gegarandeerd;
2) intuïtief eenvoudig en tussenoplossingen interpreteerbaar;
3) geschikt voor het vinden van een eerste (grove) benadering van het minimum.

b) Nadelen:

1) zeer langzame convergentie in de nabijheid van het minimum;
2) convergentiesnelheid sterk afhankelijk van de conditionering en de schaling van het probleem (verhouding van de grootste tot de kleinste eigenwaarde van de Hessiaan in de omgeving van het optimum van belang);
3) relatief nauwkeurige lijnminimalisering vereist;
4) berekeningen in opvolgende stappen onafhankelijk van elkaar.

Referenties

[2.4.1]: Zie [1.1.1] Luenberger (1973).

[2.4.2]: Zie [2.2.7] Kowalik en Osborne (1968).

[2.4.3]: Zie [2.1.5] Polak (1971).

[2.4.4]: Curry, H.B.: The method of steepest descent for nonlinear minimization problems, Qu. Appl. Math. 2 (1944), pp. 258-261.

[2.4.5]: Goldstein, A.A.: Cauchy's method of minimization. Num. Math. 4 (1962), p.p. 146-150.

[2.4.6]: Goldstein, A.A.: Constructive real analysis, Harper & Row, New York, 1967.

[2.4.7]: Luenberger, D.G.: Optimization by vector space methods, J. Wiley & Sons. Inc. , New York (1969).

[2.4.8]: Kantorovich, L.V. and Akilov, G.P.: Functional analysis in normed spaces, MacMillan, New York,(1964).

[2.4.9]: Antosiewicz, H.A. en Rheinboldt, W.C.: Numerical analysis and functional analysis, Ch. 14 in Survey of numerical analysis, J. Todd (Ed.), McGraw-Hill, New York,(1962).

[2.4.10]: Aoki, M.: Introduction to optimization techniques, fundamentals and applications of nonlinear programming, Macmillan Co., New York,(1971).

[2.4.11]: Mangasarian, O.L.: Nonlinear programming, McGraw-Hill Inc., New York (1969).

§ 2.5. Methode van Newton en enige modificaties daarvan

2.5.1. Toepassing van de methode van de steilste helling op de positief definiete kwadratische vorm $f(x) = \frac{1}{2}x^TAx + b^Tx + c$ resulteert in de iteratieformule

$$x^{(k+1)} := x^{(k)} - \alpha^{(k)}\nabla f(x^{(k)}) = x^{(k)} - \alpha^{(k)}(Ax^{(k)} + b) \qquad (2.5.1.a)$$

of met $x^* = -A^{-1}b$

$$x^{(k+1)} := x^{(k)} - \alpha^{(k)}A(x^{(k)} - x^*) \ . \qquad (2.5.1.b)$$

Realiseert men eerst een coördinatentransformatie $x =: Qy$, waar Q de eigenschap heeft dat $Q^TAQ = I$ met als resultaat de kwadratische vorm in y

$$g(y) := f(Qy) = \tfrac{1}{2}y^TQ^TAQy + b^TQy + c \ ,$$

en past men vervolgens de methode van de steilste helling toe, dan resulteert convergentie in een stap (met $\alpha^{(0)} := 1$)

$$y^{(1)} := y^{(0)} - \alpha^{(0)}\nabla g(y^{(0)}) = y^{(0)} - Q^T(AQy^{(0)} + b) = -Q^Tb \ .$$

Teruggetransformeerd naar de originele coördinaten luidt deze laatste iteratieformule

$$x^{(1)} := x^{(0)} - QQ^T\nabla f(x^{(0)}) \qquad (2.5.2)$$

welke uitdrukking met $QQ^T = A^{-1}$ overgaat in

$$x^{(1)} := x^{(0)} - A^{-1}\nabla f(x^{(0)}) = x^{(0)} - A^{-1}A(x^{(0)} - x^*) = x^* \ .$$

Deze laatste formule vormt het uitgangspunt van de <u>methode van Newton</u> (of <u>Newton-Raphson</u>) voor het minimaliseren van algemene functies $f(x)$ met gradiënt $\nabla f(x)$ en (positief definiete)Hessiaan $G(x) := [(\frac{\partial^2 f}{\partial x_i \partial x_j})(x)]$. Centraal in die methode staat de iteratieformule

$$x^{(k+1)} := x^{(k)} - [G(x^{(k)})]^{-1}\nabla f(x^{(k)}) \ . \qquad (2.5.3)$$

2.5.2. De methode van Newton kan ook worden beschouwd als een toepassing van de bekende methode van Newton-Raphson voor het bepalen van een oplossing van een stelsel niet-lineaire vergelijkingen op de vectorvergelijking

$$\nabla f(x) = 0. \tag{2.5.4}$$

Linearisering van het linkerlid van deze vergelijking rond de laatst bepaalde benadering $x^{(k)}$ voor de oplossing leidt tot de lineaire vergelijking

$$\nabla f(x^{(k)}) + G(x^{(k)})\ (x - x^{(k)}) = 0 \tag{2.5.5}$$

waarvan de oplossing (als $[G(x^{(k)})]^{-1}$ bestaat) gegeven wordt door

$$x = x^{(k)} - [G(x^{(k)})]^{-1}\nabla f(x^{(k)})\ . \tag{2.5.6}$$

2.5.3. Een derde manier om de iteratieformule van de methode van Newton af te leiden gaat uit van de lokale tweede-orde of kwadratische benadering van de functie $f(x)$ rond het laatst bepaalde punt $x^{(k)}$, d.i. uit (2.5.7)

$$f(x) = f(x^{(k)}) + \nabla^T f(x^{(k)})(x-x^{(k)}) + \tfrac{1}{2}(x-x^{(k)})^T G(x^{(k)})(x-x^{(k)}) + o(\|x-x^{(k)}\|^2)\ .$$

Wordt de kleine orde term verwaarloosd dan resteert een kwadratische vorm in $(x-x^{(k)})$ die zijn minimum bereikt (als $G(x^{(k)})$ positief definiet) in het punt

$$(x-x^{(k)}) = -[G(x^{(k)})]^{-1}\nabla f(x^{(k)})\ . \tag{2.5.6}$$

2.5.4. Bij de uitwerking van de methode van Newton in een algorithme in de terminologie van de standaardalgorithme (pt. 2.1.5) kan de iteratieformule worden opgevat als het resultaat van de combinatie van de stap voor de bepaling van de nieuwe zoekrichting

$$d^{(k)} := -[G(x^{(k)})]^{-1}\nabla f(x^{(k)}) \tag{2.5.8}$$

en de stap voor de bepaling van de stapgrootte

$$\alpha^{(k)} := 1\ .$$

In de meeste gerealiseerde algorithmen worden deze stappen gescheiden en wordt in de tweede stap de $\alpha^{(k)}$ niet bepaald met het voorschrift $\alpha^{(k)} := 1$ doch in plaats daarvan met lijnminimalisering. De algorithme van de methode van Newton krijgt met deze modificaties de vorm

Algorithme voor de methode van Newton met lijnminimalisering

(0) kies een startpunt $x^{(0)}$, zet k := 0;

(i) bepaal de functiewaarde $f(x^{(k)})$ en de gradiënt $\nabla f(x^{(k)})$

(ii) ga na of $x^{(k)}$ optimaal is; zo ja, dan klaar; zo nee, dan

(iii) bepaal de Hessiaan $G(x^{(k)})$ in $x^{(k)}$ en de zoekrichting $d^{(k)}$ uit

$$G(x^{(k)})d^{(k)} = -\nabla f(x^{(k)}) \tag{2.5.9}$$

(iv) bepaal een staplengte (factor) $\alpha^{(k)}$ uit

$$f(x^{(k)} + \alpha^{(k)}d^{(k)}) = \min\{f(x^{(k)} + \alpha d^{(k)}) \mid \alpha \in \mathbf{R}_+^1\}$$

(v) zet $x^{(k+1)} := x^{(k)} + \alpha^{(k)}d^{(k)}$ en k := k + 1 en ga terug naar stap (i).

Convergentie van de methode van Newton

2.5.5. Voor de convergentie van de methode van Newton werd door L.V. Kantorovich in 1948 een nu klassieke stelling gepubliceerd [2.5.3]. Deze stelling geldt de toepassing van de methode van Newton voor de oplossing van de operatorvergelijking

$$T(x) = 0 \tag{2.5.10}$$

waar T een in het algemeen niet-lineaire afbeelding is die een deelverzameling D van een Banach-ruimte X afbeeldt in een Banach-ruimte Y. In deze vorm kunnen nagenoeg alle problemen waarop de methode van Newton kan worden toegepast, worden herformuleerd.

2.5.6. Essentieel voor de formulering van de stelling van Kantorovich is het concept van de eerste en de tweede Frechet-afgeleide (of gewoon afgeleiden [2.5.4], [2.5.5], [2.5.6]) T' en T'' van de operator T. Deze afgeleiden zijn gedefinieerd als de lineaire begrensde operatoren

$$T': D \subset X \to \mathcal{L}(X \to Y), \qquad T'': D \subset X \to \mathcal{L}(X \to \mathcal{L}(X \to Y)) \tag{2.5.11}$$

waarvoor geldt

$$\lim_{h\to 0} \frac{1}{\|h\|} (\|T(x+h) - T(x) - T'(x)h\|) = 0 \tag{2.5.12}$$

en

$$\lim_{h\to 0} \frac{1}{\|h\|} (\|T'(x+h) - T'(x) - T''(x)h\|) = 0 . \qquad (2.5.13)$$

Zij zijn dus juist de generalisatie van het begrip functionaal operator bij vectorfuncties van meer variabelen. Voor Frechet-afgeleiden gelden soortgelijke relaties als voor functionaal operatoren. Van veel belang voor de convergentie van de methode van Newton zijn de volgende, voor tweemaal continu (Frechet-) differentieerbare operatoren geldende afschattingen

$$\|T(x+h) - T(x)\| \leq \|h\| \sup_{0\leq\alpha\leq 1} \{\|T'(x+\alpha h)\|\} \qquad (2.5.14)$$

en

$$\|T(x+h) - T(x) - T'(x)h\| \leq \tfrac{1}{2}\|h\|^2 \sup_{0\leq\alpha\leq 1} \{\|T''(x+\alpha h)\|\} \qquad (2.5.15)$$

welke uiteraard slechts gelden zolang $x + \alpha h \in D$ voor $0 \leq \alpha \leq 1$. Met behulp van de inverse $[T'(x)]^{-1}$ van de Frechet-afgeleide $T'(x)$ kan de k-de iteratiestap van de methode van Newton worden weergegeven door

$$x^{(k+1)} := x^{(k)} - [T'(x^{(k)})]^{-1} T(x^{(k)}) . \qquad (2.5.16)$$

2.5.7. Om enige "feeling" te krijgen voor de convergentievoorwaarden vervat in de stelling van Kantorovich is het nuttig de iteratieformule van de methode van Newton te beschouwen als een stap volgens de methode van de successieve approximatie, d.i.

$$x^{(k+1)} := P(x^{(k)}) = x^{(k)} - [T'(x^{(k)})]^{-1} T(x^{(k)}) \qquad (2.5.17)$$

waar P een operator is die de deelverzameling $D \subset X$ afbeeldt in X. Uit de theorie van de dekpuntsstelling is bekend dat de methode van de successieve approximatie convergeert indien de operator P een zg. contractie is, d.i. indien voor alle x en y in D geldt dat

$$\|P(y) - P(x)\| \leq \alpha\|y-x\| \quad \text{met } 0 \leq \alpha < 1 . \qquad (2.5.18)$$

Aan deze voorwaarde voor P zal worden voldaan indien geldt dat

$$\|P'(x)\| < 1 \qquad (2.5.19)$$

voor alle $x \in D \subset X$. Uitwerking van deze conditie levert voor de Newton-iteratie

$$\| (x-[T'(x)]^{-1}T(x))' \| = \| I+[T'(x)]^{-1}T''(x)[T'(x)]^{-1}T(x)-[T'(x)]^{-1}T'(x) \|$$

$$= \| [T'(x)]^{-1}T''(x)[T'(x)]^{-1}T(x) \| < 1$$

en hieraan is voldaan indien voor alle $x \in D$ geldt

$$\| [T'(x)]^{-1} \| \leq \beta, \ \| T''(x) \| \leq K, \ \| [T'(x)]^{-1}T(x) \| \leq \eta \qquad (2.5.20)$$

en

$$\beta K \eta < 1 \ . \qquad (2.5.21)$$

2.5.8. De convergentievoorwaarden in de stelling van Kantorovich hebben een soortgelijke vorm. Anders dan de hier gegeven voorwaarden hebben zij echter voornamelijk betrekking op de condities in het beginpunt van de iteratie. Door deze voorwaarden iets scherper te stellen dan hierboven kon Kantorovich bewijzen dat in alle opvolgende punten van de iteratie steeds aan dezelfde (iets scherpere) voorwaarden wordt voldaan. Dit is de grondgedachte van het bewijs van zijn stelling (zie [2.5.4], [2.5.5] en [2.5.15]).

Stelling 2.5.8 (Kantorovich). Laten X en Y twee Banach-ruimten zijn en T een operator die een deelverzameling D van X afbeeldt in Y en die Frechet-differentieerbaar is in een omgeving van het punt $x^{(0)} \in D$. Als verder voldaan wordt aan de voorwaarden

(i) er bestaat een bol $S(x^{(0)};r^{(0)})$ waarbinnen T tweemaal Frechet-differentieerbaar is en waarbinnen geldt $\| T''(x) \| \leq K$

(ii) de inverse van de Frechet-afgeleide $[T'(x^{(0)})]^{-1}$ bestaat in $x^{(0)} \in D$ en voldoet aan $\| [T'(x^{(0)})]^{-1} \| \leq \beta^{(0)}$ en $\| [T'(x^{(0)})]^{-1}T(x^{(0)}) \| = \| x^{(1)} - x^{(0)} \| \leq \eta^{(0)}$

(iii) de constante $h^{(0)} := \beta^{(0)}K\eta^{(0)}$ voldoet aan $h^{(0)} \leq \frac{1}{2}$ en de straal $r^{(0)}$ aan

$$r^{(0)} > \frac{1}{(h^{(0)})} (1 - \sqrt{1 - 2h^{(0)}})\eta^{(0)} \qquad (2.5.22)$$

dan geldt dat:

a) de rij $\{x^{(k)}\}$ gegenereerd met iteratieformule van de methode van Newton: $x^{(k+1)} := x^{(k)} - [T'(x^{(k)})]^{-1}T(x^{(k)})$, $k \geq 0$, convergeert naar de oplossing x^* van $T(x) = 0$ in D

b) de convergentiesnelheid wordt gegeven door

$$\|x^{(k)} - x^*\| \leq (\tfrac{1}{2})^{k-1}(2h^{(0)})^{2^{k-1}}\eta^{(0)} \qquad (2.5.23)$$

c) de oplossing x^* is uniek in iedere gesloten bol $\bar{S}(x^{(0)};r) \subset S(x^{(0)};r^{(0)})$

met $r < \left(\frac{1}{h^{(0)}}\right)(1 + \sqrt{1 - 2h^{(0)}})\,\eta^{(0)}$

Bewijs. Zie [2.5.3], [2.5.4], [2.5.5].

Opmerking. In het geval dat aan de schattingen in (i) en (ii) wordt voldaan in een convex gebied $D \subset X$ i.p.v. in het punt $x^{(0)} \in D$ alleen en $h = \mathfrak{K}\eta < \frac{1}{2}$ dan kan worden aangetoond (zie [2.5.15]) dat $\{x^{(k)}\}$ convergeert naar een (niet noodzakelijk uniek) punt $\hat{x} \in S(x^{(0)};r) \subset D$ met $r = \eta/(1-h)$ waar $T(\hat{x}) = 0$.

Convergentiesnelheid van de methode van Newton

2.5.9. De stelling van Kantorovich geeft een afschatting voor de convergentiesnelheid van de methode van Newton. Een tweede uitdrukking daarvoor kan worden afgeleid met de in pt. 2.5.6 gegeven afschattingen voor de Frechet-afgeleiden toegepast op de in pt. 2.5.7 geïntroduceerde operator P

$$P(x) := x - [T'(x)]^{-1}T(x) \,. \qquad (2.5.24)$$

Aangezien (zie pt. 2.5.7)

$$P'(x) = [T'(x)]^{-1}T''(x)[T'(x)]^{-1}T(x) \qquad (2.5.25)$$

geldt als x^* de oplossing is van de vergelijking $T(x) = 0$ dat

$$P(x^*) = x^* \quad \text{en} \quad P'(x^*) = 0 \,. \qquad (2.5.26)$$

Voor het $(k+1)^e$ element van de rij $\{x^{(k)}\}$ gegenereerd door toepassing van de methode van Newton voor de oplossing van de vergelijking $T(x) = 0$ geldt (onder de veronderstelling dat P tenminste tweemaal (Frechet-) differentieerbaar is op de lijn $x := x^* + \alpha(x^{(k)} - x^*)$ voor $0 \leq \alpha \leq 1$) dat

$$\begin{aligned}\|x^{(k+1)} - x^*\| &= \|P(x^{(k)}) - P(x^*)\| \\ &= \|P(x^{(k)}) - P(x^*) - P'(x^*)(x^{(k)} - x^*)\| \\ &\leq \tfrac{1}{2}\|x^{(k)} - x^*\|^2 \sup_{0\leq\alpha\leq 1}\{\|P''(x^* + \alpha(x^{(k)} - x^*))\|\}.\end{aligned}$$

Voor de opvolgende punten gegenereerd met behulp van de methode van Newton geldt nu, als $\sup_{0\le\alpha\le 1} \{\|P''(x^* + \alpha(x^{(k)} - x^*))\|\} \le M$, dat

$$\|x^{(k+1)} - x^*\| \le \frac{M}{2}\|x^{(k)} - x^*\|^2 \qquad (2.5.27)$$

en de methode is dan volgens de definitie in pt. 2.1.14 kwadratisch convergent (en bezit tweede-orde convergentie eigenschappen). Bij practische problemen is niet altijd aan de voorwaarde dat P tweemaal (Frechet-) differentieerbaar is op alle lijnen die de iteratiepunten verbinden met de oplossing voldaan. In dat geval kan slechtere of soms, in het geheel geen convergentie optreden.

2.5.10. In het geval van de toepassing van de methode van Newton voor de minimalisering van een functie f(x) kan op illustratieve wijze een afschatting worden gegeven voor de norm van het verschil tussen twee opvolgende iteratiepunten (welke uiteraard overeenkomt met de afschatting in pt. 2.5.9). Er geldt

$$x^{(k+1)} - x^{(k)} = -[G(x^{(k)})]^{-1}\nabla f(x^{(k)})$$

$$= -[G(x^{(k)})]^{-1}[\nabla f(x^{(k-1)}) + G(x^{(k-1)})(x^{(k)} - x^{(k-1)}) + O(\|x^{(k)} - x^{(k-1)}\|^2)].$$

Volgens de methode van Newton geldt voor het punt $x^{(k)}$ juist dat

$$\nabla f(x^{(k-1)}) + G(x^{(k-1)})\,(x^{(k)} - x^{(k-1)}) = 0$$

waarmee direct volgt dat, indien de norm van $[G(x^{(k)})]^{-1}$ begrensd is, geldt

$$\|x^{(k+1)} - x^{(k)}\| \le M\,\|x^{(k)} - x^{(k-1)}\|^2 . \qquad (2.5.28)$$

Uiteraard geldt deze relatie alleen wanneer $\|x^{(k)} - x^{(k-1)}\|$ klein is d.w.z. indien $x^{(k)}$ de oplossing redelijk goed benadert. In de praktijk betekent deze afschatting dat bij de hypothetische afwezigheid van numerieke onnauwkeurigheden in de omgeving van de oplossing het aantal nullen in de getalwaarde voor de norm van het verschil tussen twee opvolgende iteratiepunten gegenereerd door de methode van Newton per iteratiestap verdubbelt.

Voor- en nadelen van de methode van Newton

2.5.11. De methode van Newton voor het minimaliseren van functies van meer variabelen heeft als belangrijke voordelen:

1) zeer snelle (kwadratische of tweede-orde) convergentie als convergentie optreedt
2) in principe geen lijnminimalisering vereist.

Als nadelen staan hier tegenover:

1) bij iedere iteratiestap moeten naast de n componenten van de gradiënt nog $\frac{1}{2}(n^2+n)$ componenten van de Hessiaan worden berekend en in het computergeheugen worden opgeborgen
2) bij iedere iteratieslag moet (in stap (iii) van de algorithme in pt. 2.5.4) een stelsel van n lineaire vergelijkingen (ofwel een matrix-vector-vergelijking) worden opgelost voor het bepalen van de zoekrichting
3) convergentie kan (zoals uit de stelling van Kantorovich (St. 2.5.8) blijkt) alleen worden gegarandeerd wanneer het beginpunt wordt gekozen in de directe omgeving van de oplossing. Wordt het startpunt te ver weg gekozen dan convergeert de (ongemodificeerde) methode van Newton vaak in het geheel niet (mede om reden van de eigenschappen hieronder genoemd als nadeel 4))
4) het op te lossen stelsel lineaire vergelijkingen genoemd bij het nadeel 2) heeft in de practijk niet altijd een oplossing (de Hessiaan G(x) bezit niet altijd een inverse) en de gevonden oplossing (= zoekrichting) resulteert niet altijd in een vermindering van de functiewaarde: De Hessiaan G(x) is in de practijk niet steeds positief definiet in welk geval het mogelijk wordt dat $\nabla^T f(x^{(k)}) d^{(k)} = -\nabla^T f(x^{(k)})[G(x^{(k)})]^{-1} \nabla f(x^{(k)}) > 0$.

Modificaties van de methode van Newton

2.5.12. Gemotiveerd door de in het vorige punt gesignaleerde voor- en nadelen van de methode van Newton zijn in de literatuur een aantal modificaties gesuggereerd die beogen de nadelen op te heffen onder behoud van zoveel mogelijk van de voordelen. In de eerste plaats zijn er twee modificaties (waarvan de eerste eigenlijk standaard practijk geworden is (vgl. pt. 2.5.4)) die vaak in theoretische beschouwingen worden opgenomen en die beide in de literatuur soms worden aangeduid als de gemodificeerde methode van Newton. Deze twee modificaties betreffen respectievelijk:

1) de vervanging van de theoretische stapgroottefactor $\alpha^{(k)} := 1$ door een stapgroottefactor $\alpha^{(k)}$ bepaald door lijnminimalisering

2) de vervanging van de Hessiaan $G(x^{(k)})$ in het k-de iteratiepunt in de matrix-vector-vergelijking voor de berekening van de k-de zoekrichting $d^{(k)}$ door de Hessiaan $G(x^{(0)})$ in het startpunt. Bij gebruik van deze modificatie kan de berekening van de zoekrichting worden gereduceerd tot een eenmalige matrix-inversie (van $G(x^{(0)})$) in de eerste iteratieslag en een matrix-vector-vermenigvuldiging $d^{(k)} := -[G(x^{(0)})]^{-1}\nabla f(x^{(k)})$ in alle opvolgende iteratieslagen. Voor de aldus gemodificeerde methode kan nog slechts lineaire convergentie worden gegarandeerd (zie [2.5.4], [2.5.5]).

2.5.13. Naast deze twee min of meer klassieke modificaties zijn er in de laatste jaren een aantal practische modificaties voorgesteld welke alle drie betrekking hebben op de in het voorgaande punt genoemde "klassieke" nadeel van de methode van Newton, nl. dat de methode niet convergeert indien het startpunt niet ligt in de directe omgeving van de oplossing. Drie van deze modificaties zullen hieronder worden besproken:

3) de modificatie van Goldfeld, Quandt en Trotter [2.5.8]
4) de modificatie van Greenstadt [2.5.9]
5) de modificatie van Fiacco-McCormick [2.5.10].

Modificatie van Goldfeld, Quandt en Trotter

2.5.14. De voornaamste reden voor het slechte convergentiegedrag van de methode van Newton in het geval het startpunt niet in de directe omgeving van de oplossing ligt is de omstandigheid dat de Hessiaan vaak niet meer positief definiet is. In dat geval bestaan er gradiënt-richtingen die aanleiding geven tot Newton-methode-zoekrichtingen die een scherpe hoek maken met de positieve gradiënt, d.i.

$$\nabla^T f(x^{(k)}) d^{(k)} = -\nabla^T f(x^{(k)}) [G(x^{(k)})]^{-1} \nabla f(x^{(k)}) > 0 \ . \qquad (2.5.29)$$

Een methode om deze situatie waarin het bewegen in de zoekrichting leidt tot verhoging van de functiewaarde te voorkomen werd in 1944 gesuggereerd door Levenberg in verband met niet-lineaire kleinste-kwadratenproblemen [2.5.11] werd uitgewerkt voor de methode van Newton door Goldfeld, Quandt en Trotter [2.5.12]. Het idee daarbij is dat de matrix $G(x^{(k)})$ bij de bepaling van de zoekrichting wordt vervangen door een positief definiete matrix $[G(x^{(k)}) + \mu^{(k)} I]$

waar $\mu^{(k)}$ een positieve constante is, d.w.z. $d^{(k)}$ wordt bepaald als oplossing van de vergelijking

$$[G(x^{(k)}) + \mu^{(k)}I]d^{(k)} = -\nabla f(x^{(k)}) \ . \tag{2.5.30}$$

Door $\mu^{(k)}$ groot genoeg te kiezen kan de matrix $[G(x^{(k)}) + \mu^{(k)}I]$ steeds positief definiet worden gemaakt met als resultaat dat de gevonden zoekrichting een scherpe hoek maakt met de negatieve gradiënt. Is de matrix $G(x^{(k)})$ al positief definiet van zichzelf dan kan voor $\mu^{(k)}$ de waarde $\mu^{(k)} = 0$ worden gekozen en heeft men de originele methode van Newton terug. Laat men de waarde van $\mu^{(k)}$ zeer groot worden, d.i. $\mu^{(k)} \to +\infty$, dan overheerst in de matrix $[G(x^{(k)}) + \mu^{(k)}I]$ de correctietermen en benadert de zoekrichting de negatieve gradiënt

$$d^{(k)} \approx -\frac{1}{\mu^{(k)}} \nabla f(x^{(k)}) \ . \tag{2.5.31}$$

Men heeft dan bij benadering de methode van de steilste helling. De methode kan in de practijk dus worden beschouwd als een tussenvorm tussen de altijd maar langzaam convergerende methode van de steilste helling en de ver van de oplossing *niet, maar dichtbij de oplossing snel convergerende methode van* Newton.

2.5.15. Opgemerkt kan worden dat bij de modificatie van Goldfeld, Quandt en Trotter lijnminimalisering achterwege gelaten wordt en dat de stapgrootte geregeld wordt met de grootte van de constante $\mu^{(k)}$. Aangetoond kan worden dat, indien $\mu^{(k)} > \max\{0,-\lambda_1^{(k)}\}$ waar $\lambda_1^{(k)}$ de kleinste eigenwaarde is van de Hessiaan $G(x^{(k)})$, geldt dat het nieuwe iteratiepunt

$$x^{(k+1)} := x^{(k)} - [G(x^{(k)}) + \mu^{(k)}I]^{-1}\nabla f(x^{(k)}) \tag{2.5.32}$$

de oplossing is van het beperkte minimaliseringsprobleem (vgl. pt.2.5.3)

$$\min\{f(x^{(k)}) + \nabla^T f(x^{(k)})(x-x^{(k)}) + \tfrac{1}{2}(x-x^{(k)})^T G(x^{(k)})(x-x^{(k)})$$

$$\Big| \|x-x^{(k)}\|^2 \le \|[G(x^{(k)}) + \mu^{(k)}I]^{-1}\nabla f(x^{(k)})\|^2 \ , \ x \in \mathbb{R}^n\}. \tag{2.5.33}$$

(als ook van het onbeperkte minimaliseringsprobleem

$$\min\{f(x^{(k)}) + \nabla^T f(x^{(k)})(x - x^{(k)}) + \tfrac{1}{2}(x - x^{(k)})^T G(x^{(k)})(x - x^{(k)})$$

$$+ \tfrac{1}{2}\mu^{(k)}\|x - x^{(k)}\|^2\}) \tag{2.5.34}$$

Tevens geldt dat $r(\mu^{(k)}) = \|[G(x^{(k)}) + \mu^{(k)}I]^{-1}\nabla f(x^{(k)})\|$ voor $\mu^{(k)} > -\lambda_1^{(k)}$ een monotoon dalende functie van $\mu^{(k)}$ is (en dat $\lim_{\mu^{(k)}\to\infty} r(\mu^{(k)}) = 0$). In verband met dit resultaat suggereren Goldfeld, Quandt en Trotter voor $\mu^{(k)}$ de waarde

$$\mu^{(k)} := \max\{0, -\lambda_1^{(k)} + (\frac{1}{R^{(k)}})\|\nabla f(x^{(k)})\|\} \qquad (2.5.35)$$

waarin $\lambda_1^{(k)}$ weer de kleinste eigenwaarde voorstelt van de Hessiaan $G(x^{(k)})$ en $R^{(k)}$ de straal aangeeft van een hyperbol met $x^{(k)}$ als middelpunt waarop of waarbinnen het volgende punt $x^{(k+1)}$ gevonden zal worden. De keuze van deze straal wordt door hen afhankelijk gesteld van de mate waarin de functie $f(x)$ lokaal benaderd kan worden door een kwadratische vorm. Opvolgen van deze suggestie van Goldfeld, Quandt en Trotter impliceert dat in iedere iteratiestap de kleinste eigenwaarde van de Hessiaan $G(x^{(k)})$ moet worden bepaald hetgeen een hoeveelheid extra rekenwerk betekent. In plaats daarvan kan men ook de strategie volgen die lijkt op de strategie gesuggereerd door Marquardt (zie pt.2.10.10) bij een analoge methode voor niet-lineaire kleinste-kwadrate problemen. Deze strategie houdt in dat het streven naar het gebruiken van de kleinst mogelijke $\mu^{(k)} \geq 0$ wordt gerealiseerd door in het begin van het iteratieproces $\mu^{(k)} := 0$, $k = 0,1,2,\ldots$ te kiezen en eerst dan $\mu^{(k)}$ een positieve waarde $\mu^{(k)} := \mu_0 > 0$ te geven indien de functiewaarde in het met deze $\mu^{(k)} := 0$ berekende nieuwe iteratiepunt $x^{(k+1)}$ groter is dan in het voorgaande punt $x^{(k)}$. Hierna wordt in elke nieuwe iteratiestap $\mu^{(k)} := \mu^{(k-1)}/\nu$ (met $\nu > 1$) gekozen en wordt afhankelijk van het al dan niet groter zijn van de oude functiewaarde $f(x^{(k)})$ dan de nieuwe functiewaarde $f(x^{(k+1)})$ beslist om het nieuwe punt $x^{(k+1)}$ te accepteren (als $f(x^{(k+1)}) < f(x^{(k)})$)of om een nieuwe $\mu^{(k)} := \nu\mu^{(k)}$ te proberen (als $f(x^{(k+1)}) \geq f(x^{(k)})$). (Een veel gebruikte praktijkwaarde van ν is $\nu = 10$.) De strategie van Marquardt wijkt af van de hier geschetste omdat daarbij steeds begonnen wordt met $\mu^{(0)} = 0.01 > 0$.

Modificatie van Greenstadt

2.5.16. Een tweede bekende modificatie van de methode van Newton om te voorkomen dat bewegen in de richting van de met de algorithme bepaalde zoekrichting leidt tot hogere in plaats van lagere functiewaarden is de door Greenstadt [2.5.9]

gesuggereerde modificatie. Deze modificatie is gebaseerd op de observatie dat de "pathologische" zoekrichtingen lineaire combinaties bevatten van die eigenvectoren van de (niet positief definiete) Hessiaan die corresponderen met de negatieve eigenwaarden. Dit kan worden verduidelijkt door gebruik te maken van de eigenschap dat iedere reële symmetrische matrix een orthonormale basis van eigenvectoren bezit. Indien de met de (symmetrische) Hessiaan $G(x^{(k)})$ corresponderende orthonormale basis wordt weergegeven als $P^{(k)} := [p_1^{(k)}, p_2^{(k)}, \ldots, p_n^{(k)}]$ waar $p_j^{(k)}$ de j-de eigenvector is

$$G(x^{(k)})p_j^{(k)} = \lambda_j^{(k)} p_j^{(k)} \tag{2.5.36}$$

dan kan de Hessiaan worden herschreven als

$$G(x^{(k)}) = P^{(k)}\Lambda^{(k)}P^{(k)T} = \sum_{j=1}^{n} \lambda_j^{(k)} p_j^{(k)} p_j^{(k)T} . \tag{2.5.37}$$

(Deze uitdrukking staat bekend als de spectrale decompositie van de symmetrische matrix $G(x^{(k)})$.) In het geval dat geen der eigenwaarden gelijk aan nul is, bestaat de inverse van de matrix $G(x^{(k)})$ en is gelijk aan

$$[G(x^{(k)})]^{-1} = P^{(k)}(\Lambda^{(k)})^{-1}P^{(k)T} = \sum_{j=1}^{n} (\lambda_j^{(k)})^{-1} p_j^{(k)} p_j^{(k)T} . \tag{2.5.38}$$

Substitutie van deze uitdrukking in de formule voor de zoekrichting van de methode van Newton geeft

$$\begin{aligned} d^{(k)} &:= -\sum_{j=1}^{n} \frac{1}{\lambda_j^{(k)}} p_j^{(k)} p_j^{(k)T} \nabla f(x^{(k)}) \\ &= -\sum_{j=1}^{n} \frac{1}{\lambda_j^{(k)}} < p_j^{(k)}, \nabla f(x^{(k)}) > p_j^{(k)} \end{aligned} \tag{2.5.39}$$

waar $\lambda_j^{(k)}$ en $p_j^{(k)}$ respectievelijk de j-de eigenwaarde en de j-de eigenvector zijn van de Hessiaan $G(x^{(k)})$ in het k-de iteratiepunt. De gevonden uitdrukking ontleedt als het ware de generatie van de zoekrichting van de methode van Newton uit de negatieve gradiëntrichting: componenten van de negatieve gradiënt in de richting van eigenvectoren die corresponderen met eigenwaarden

$\lambda_j^{(k)} > 1$ worden verkleind, componenten in de richting van eigenvectoren die corresponderen met eigenwaarden $0 < \lambda_j^{(k)} < 1$ worden vergroot en componenten van de negatieve gradiënt in de richting van eigenvectoren die corresponderen met eigenwaarden $\lambda_j^{(k)} < 0$ worden van teken omgedraaid. Een en ander wordt geïllustreerd voor een 2-dimensionale situatie in de volgende schets

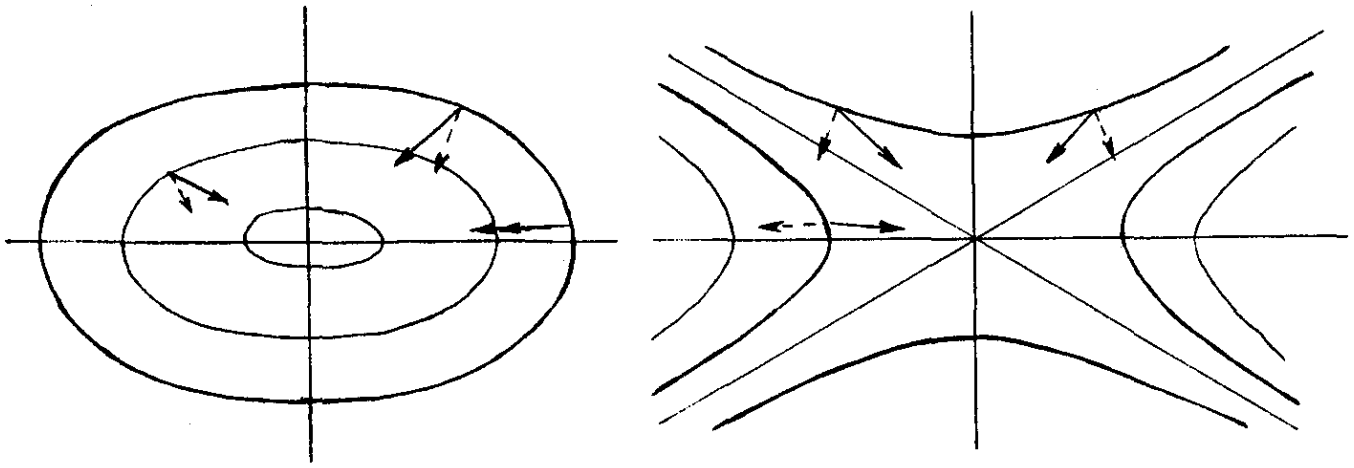


Fig.2.5.16 Generatie van de zoekrichting van de methode van Newton.

2.5.17. Gemotiveerd door het hierboven geschetste gedrag van de methode van Newton bij het genereren van zoekrichtingen en in de overweging dat in iedere iteratiestap de functiewaarde zoveel mogelijk dient te worden verlaagd suggereerde Greenstadt om de zoekrichting in de methode van Newton in plaats van op de gebruikelijke manier te bepalen als oplossing van de vergelijking

$$\bar{G}(x^{(k)})d^{(k)} = -\nabla f(x^{(k)}) \tag{2.5.40}$$

waar

$$\bar{G}(x^{(k)}) = \sum_{j=1}^{n} \beta_j^{(k)} p_j^{(k)} p_j^{(k)T} \tag{2.5.41}$$

en

$$\beta_j^{(k)} = |\lambda_j^{(k)}| \quad \text{als} \quad |\lambda_j^{(k)}| \geq \varepsilon > 0 \tag{2.5.42}$$

$$= \varepsilon \quad \text{als} \quad |\lambda_j^{(k)}| < \varepsilon \ .$$

Voor het getal ε dat werd ingevoerd ter voorkoming van het singulier worden van de matrix $\bar{G}(x^{(k)})$ kan het kleinste getal worden gekozen waarmee de resulterende matrix in de computeralgorithmen juist als niet-singulier wordt aangemerkt.

2.5.18. Vergelijking van de modificatie van Goldfeld, Quandt en Trotter (pt. 2.5.15) met de modificatie van Greenstadt (pt. 2.5.16) leert dat in de terminologie van de spectrale decompositie van de $G(x^{(k)})$ de zoekrichting in het eerste geval gelijk wordt aan

$$d^{(k)} := \sum_{j=1}^{n} \frac{1}{(\lambda_j^{(k)}+\mu^{(k)})} < p_j^{(k)}, \nabla f(x^{(k)}) > p_j^{(k)} \qquad (2.5.43)$$

en in het tweede geval gelijk aan

$$d^{(k)} := \sum_{j=1}^{n} \frac{1}{\max\{|\lambda_j^{(k)}|, \varepsilon\}} < p_j^{(k)}, \nabla f(x^{(k)}) > p_j^{(k)} \; . \qquad (2.5.44)$$

Aangezien $\mu^{(k)}$ zo gekozen dient te worden dat $\mu^{(k)} > -\lambda_1^{(k)}$ waar $\lambda_1^{(k)}$ de kleinste eigenwaarde is van de Hessiaan $G(x^{(k)})$ volgt dat de correctie volgens de eerste methode veel ingrijpender is dan die volgens de tweede die subtieler corrigeert. Anderzijds geldt in het geval van een grote meest negatieve eigenwaarde $\lambda_1^{(k)} << 0$ met $\mu^{(k)} > -\lambda_1^{(k)}$ dat

$$\frac{1}{\lambda_1^{(k)} + \mu^{(k)}} \geq \frac{1}{\lambda_j^{(k)} + \mu^{(k)}} \qquad j = 2,\ldots,n$$

terwijl niet steeds behoeft te gelden dat

$$\frac{1}{|\lambda_1^{(k)}|} \geq \frac{1}{\max\{|\lambda_j^{(k)}|, \varepsilon\}} \qquad j = 2,\ldots,n \; .$$

Dit impliceert dat bij de modificatie van Goldfeld, Quandt en Trotter de richting van de eigenvector die correspondeert met de meest negatieve eigenwaarde sterker benadrukt wordt dan in het geval van de modificatie van Greenstadt. Omdat die richting in het algemeen gunstig is omdat de functie daarlangs steeds sterker afneemt (de tweede afgeleide van de lijnfunctie $h(\alpha)$ (vgl.pt. 2.2.1) is negatief) biedt de eerste methode wellicht meer voordeel dan de tweede (vgl.[2.5.13]).

2.5.19. Een nadeel van de modificatie van Greenstadt is dat in iedere iteratieslag de eigenwaarden en eigenvectoren van de Hessiaan moeten worden berekend.

Dit betekent een aanzienlijke vermeerdering van de hoeveelheid rekenwerk per iteratieslag. Een mogelijkheid om op eenvoudige wijze een gedeelte van dit werk te voorkomen is zo mogelijk gebruik maken van de Cholesky-decompositie-methode voor de oplossing van de vergelijking van de originele Newton methode

$$[G(x^{(k)})]d^{(k)} = -\nabla f(x^{(k)})$$

d.w.z. gebruik te maken van de Cholesky-decompositie van de Hessiaan

$$G(x^{(k)}) = L^{(k)}D^{(k)}L^{(k)T} \qquad (2.5.45)$$

waar $L^{(k)}$ een onderdriehoeksmatrix is met enen op de diagonaal en waar $D^{(k)}$ een diagonaalmatrix is. De vergelijking kan met deze decompositie worden opgelost door achtereenvolgens op te lossen

$$L^{(k)}y^{(k)} = -\nabla f(x^{(k)})$$

en (2.5.46)

$$D^{(k)}L^{(k)T}d^{(k)} = y^{(k)} \ .$$

Door het gebruik van deze methode, die niet altijd mogelijk is wanneer $G(x^{(k)})$ niet positief definiet is, kan zonder extra werk worden geconstateerd of $G(x^{(k)})$ negatieve eigenwaarden heeft aan het feit of diagonaalmatrix $D^{(k)}$ negatieve elementen bevat. Is dit het geval dan kan alsnog een eigenwaarde en eigenvectorberekening worden gestart. In alle gevallen waarin dit niet het geval is kan de tijdrovende eigenwaardeberekening achterwege worden gelaten.

Modificatie van Fiacco-McCormick

2.5.20. Een practische modificatie van de methode van Newton die minder extra rekenwerk vereist dan de modificatie van Greenstadt werd gesuggereerd door Fiacco en McCormick [2.5.10]. Het idee achter deze modificatie is dat, in het geval dat de Hessiaan niet positief definiet is, de richtingen $s^{(k)}$ waarvoor geldt

$$s^{(k)T}G(x^{(k)})s^{(k)} < 0$$

aantrekkelijke zoekrichtingen zijn omdat de functie in die richting steeds sneller afneemt (vgl. pt.2.2.1, $h''(\alpha) < 0$). Ook aantrekkelijk zijn de richtingen $s^{(k)}$ waarvoor geldt

$$s^{(k)T}G(x^{(k)})s^{(k)} = 0 \ .$$

In beide gevallen betekent het bewegen langs deze richtingen meestal het wegbewegen van zadelpunten en (lokale) maxima. Een manier om deze richtingen te genereren is weer (zo mogelijk) gebruik te maken van de Cholesky-decompositiemethode voor het oplossen van de originele Newton-methode-vergelijking (2.5.9)

$$G(x^{(k)})d^{(k)} = -\nabla f(x^{(k)})$$

welke dan overgaat in (2.5.45)

$$L^{(k)}D^{(k)}L^{(k)T}d^{(k)} = -\nabla f(x^{(k)}) \ .$$

Ontdekt men bij de Cholesky-decompositie dat een of meer diagonaalelementen van de matrix $D^{(k)}$ negatief worden dan bepaalt men de richting $\bar{d}^{(k)}$ als de oplossing van de vergelijking

$$L^{(k)T}\bar{d}^{(k)} = a^{(k)} \qquad (2.5.47)$$

waar $a^{(k)}$ een vector is met als componenten

$$a_j^{(k)} := 1 \text{ als } D_{jj}^{(k)} < 0 \qquad (2.5.48)$$

$$:= 0 \text{ als } D_{jj}^{(k)} \geq 0 \ .$$

Het volgt onmiddellijk dat de gevonden richting er een is waarvoor geldt

$$\bar{d}^{(k)T}G(x^{(k)})\bar{d}^{(k)} = a^{(k)T}D^{(k)}a^{(k)} < 0 \ . \qquad (2.5.49)$$

De uiteindelijke zoekrichting wordt gelijk genomen aan

$$d^{(k)} := -\text{sgn}(\langle \nabla f(x^{(k)}), \bar{d}^{(k)} \rangle)\bar{d}^{(k)} \ . \qquad (2.5.50)$$

In het geval dat de matrix $D^{(k)}$ geen negatieve elementen bevat doch wel elementen gelijk aan nul wordt de zoekrichting bepaald uit (2.5.45)

$$L^{(k)}D^{(k)}L^{(k)T}d^{(k)} = -\nabla f(x^{(k)})$$

of, als deze vergelijking geen oplossing heeft, een richting $\bar{d}^{(k)}$ uit

$$L^{(k)}D^{(k)}L^{(k)T}\bar{d}^{(k)} = 0 \qquad (2.5.51)$$

en de uiteindelijke zoekrichting weer uit

$$d^{(k)} := -\text{sgn}(\langle \nabla f(x^{(k)}), \bar{d}^{(k)} \rangle)\bar{d}^{(k)} \ . \qquad (2.5.52)$$

In het geval dat geen Cholesky-decompositie mogelijk is kan de hierboven geschetste procedure niet worden toegepast. In dit uitzonderingsgeval kan als zoekrichting worden genomen

$$d^{(k)} := -\nabla f(x^{(k)}) \ . \qquad (2.5.53)$$

2.5.21. Resumerend omvat de methode van Fiacco-McCormick de volgende modificatie van de algorithme voor de methode van Newton met lijnminimalisering:

Algorithme van de modificatie van Fiacco-McCormick

Vervang in de algorithme in pt. 2.5.4 stap (iii) door:

(iii)(a) bepaal de Cholesky-decompositie $L^{(k)}D^{(k)}L^{(k)T}$ van de matrix $G(x^{(k)})$

(b) is decompositie niet mogelijk, zet dan $d^{(k)} = -\nabla f(x^{(k)})$;
is decompositie wel mogelijk,bepaal dan de zoekrichting als volgt

(c) zijn alle $D_{jj}^{(k)} > 0$ dan bepaal $d^{(k)}$ uit $L^{(k)}D^{(k)}L^{(k)T}d^{(k)} = -\nabla f(x^{(k)})$

(d) zijn er negatieve $D_{jj}^{(k)}$'s dan bepaal de vector $a^{(k)} :=$

$$\max\left(-\frac{D_{jj}^{(k)}}{|D_{jj}^{(k)}|},\ 0\right) \text{ en } d^{(k)} \text{ uit } L^{(k)T}\bar{d}^{(k)} = a^{(k)} \text{ en}$$

$$d^{(k)} := \operatorname{sgn}(<\bar{d}^{(k)}, \nabla f(x^{(k)})>)\bar{d}^{(k)}$$

(e) zijn er geen negatieve $D_{jj}^{(k)}$'s doch wel $D_{jj}^{(k)}$'s gelijk aan nul bepaal dan $d^{(k)}$ uit $L^{(k)}D^{(k)}L^{(k)T}d^{(k)} = -\nabla f(x^{(k)})$ of een richting $\bar{d}^{(k)}$ uit $L^{(k)}D^{(k)}L^{(k)T}\bar{d}^{(k)} = 0$ en $d^{(k)}$ uit $d^{(k)} := -\operatorname{sgn}(<\bar{d}^{(k)}, \nabla f(x^{(k)})>)\bar{d}^{(k)}$.

Opmerking: De meest gebruikte practische algorithmen voor de Cholesky-decompositie komen in moeilijkheden wanneer er een of meer elementen $D_{jj}^{(k)}$ zijn waarvoor $D_{jj}^{(k)} = 0$. Wordt een dergelijk algorithme gebruikt dan kan dit laatste gedeelte (e) van stap (iii) worden weggelaten

2.5.22. Uit recente vergelijkingsonderzoeken (bv. [2.5.14])blijkt dat de methode van Newton met de hierboven in detail geschetste modificatie van Fiacco-McCormick voor veel minimaliseringsproblemen een van de meest efficiënte methoden is die op dit moment bekend zijn.

2.5.23. Referenties

[2.5.1]: Zie [1.1.3] Murray (1972)

[2.5.2]: Zie [2.2.7] Kowalik en Osborne (1968)

[2.5.3]: Kantorovich, L.V.: Functional analysis and applied mathematics, Uspekhi Mat. Nauk. 3 (1948), pp. 89-185. Translation: C.D. Benster, National Bureau of Standards Report 1509, G.E. Forsythe, ed. (1952).

[2.5.4]: Zie [2.4.8] Kantorovich en Akilov (1964).

[2.5.5]: Zie [2.4.9] Antosiewicz en Rheinboldt (1962).

[2.5.6]: Zie [2.4.7] Luenberger (1969).

[2.5.7]: Zie [1.1.1] Luenberger (1973).

[2.5.8]: Goldfeld, S.M. and Quandt, R.E.: Nonlinear methods in econometrics, North Holland Publ.Cy., Amsterdam (1972).

[2.5.9]: Greenstadt, J.E.: On the relative efficiencies of gradient methods, Math. of Comp. 21 (1967) pp. 23-26.

[2.5.10]: Fiacco, A.V. en McCormick, G.P.: Nonlinear programming: sequential unconstrained minimization techniques, Wiley, New York (1968).

[2.5.11]: Levenberg, K.: A method for the solution of certain nonlinear problems in least squares, Qu. Appl. Math. 2 (1944), pp. 164-168.

[2.5.12]: Goldfeld, S.M., Quandt, R.E. and Trotter, H.F.: Maximization by quadratic hill-climbing, Econometrica 34 (1966), pp. 541-551.

[2.5.13]: Zie [2.2.10] Powell (1971).

[2.5.14]: Lootsma, F.A.: Penalty function performance of several unconstrained minimization techniques, Philips Res. Repts. 27 (1972), pp. 358-385.

[2.5.15]: Stoer, J.: Einführung in die Numerische Mathematik I, Heidelberger Taschenbücher Band 105, Springer Verlag, Berlin (1972).

§ 2.6 Methoden gebaseerd op het gebruik van geconjugeerde richtingen

2.6.1. De n lineair onafhankelijke richtingen $\{d_0,d_1,\ldots,d_{n-1}\}$ in $\mathbb{R}^n$ heten A-orthogonaal of onderling A-geconjugeerd (of geconjugeerd t.o.v. een (symmetrisch positief definiete) matrix A) indien geldt

$$\begin{aligned} d_i^T A d_j &= 0 \qquad \text{als } i \neq j \\ d_i^T A d_j &\neq 0 \qquad \text{als } i = j \ . \end{aligned} \tag{2.6.1}$$

2.6.2. A-geconjugeerde richtingen zijn de generalisatie van de zgn. toegevoegde middellijnen van een ellips (zie schets, fig. 2.6.1), die op hun beurt weer de (scheve) projectie zijn van orthogonale assenkruizen van de tot ellips geprojecteerde cirkel.

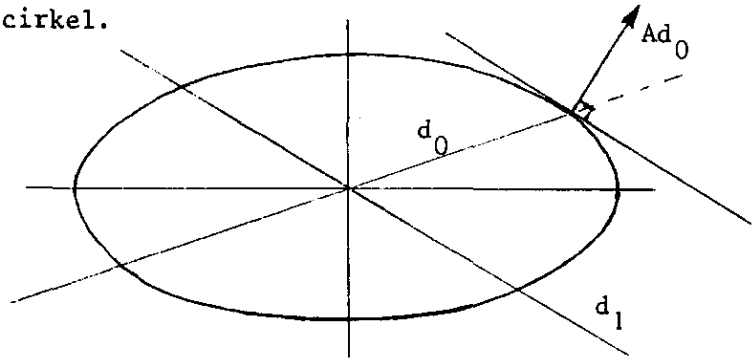


Fig. 2.6.1.

Dezelfde lineaire transformatie x=: Qy met $Q^TAQ = I$ welke (vgl. pt. 2.5.1) niveau oppervlakken van de kwadratische vorm

$$f(x) = \tfrac{1}{2}x^TAx + b^Tx + c$$

in hyperbollen transformeert, transformeert ieder stelsel A-orthogonale richtingen $\{d_0,d_1,\ldots,d_{n-1}\}$ in een orthogonaal assenstelsel. Immers voor de richtingen d_i' en d_j' die uit de A-orthogonale richtingen d_i en d_j worden gegenereerd met behulp van de transformatie $d_i =: Qd_i'$ en $d_j =: Qd_j'$ geldt

$$d_i'^T d_j' = (d_i'^T Q^T)A(Qd_j') = d_i^T A d_j = 0 \quad \text{als } i \neq j$$

en

$$d_i'^T d_j' \neq 0 \qquad \text{als } i = j \ . \tag{2.6.2}$$

2.6.3. Indien een symmetrische matrix A positief definiet is dan bestaat er steeds een volledig stelsel A-geconjugeerde richtingen, d.i. er bestaan n lineair onafhankelijke richtingen die onderling A-geconjugeerd zijn. Immers bij iedere symmetrische positief definiete matrix A behoort een orthonormaal stelsel van eigenvectoren P en de kolommen van de matrix $Q =: P\Lambda^{-\frac{1}{2}}$ vormen dan een A-orthogonaal stelsel. Deze matrix is juist dezelfde als de transformatiematrix Q uit 2.6.2.

2.6.4. Het gebruik van geconjugeerde richtingen bij minimaliseringsalgorithmen is gebaseerd op de volgende twee eigenschappen:

Eigenschap I. Indien de verzameling $\{d_0,\ldots,d_{n-1}\}$ een volledig stelsel A-geconjugeerde richtingen vormt en x_0 een willekeurig punt is in $\mathbb{R}^n$ en $\hat{x}$ het minimale punt is van de kwadratische vorm $f(x) = \frac{1}{2}x^TAx + b^Tx + c$ dan geldt met de notatie $g_0 := \nabla f(x_0)$ voor de gradiënt in x_0

$$\hat{x} = x_0 + \sum_{k=0}^{n-1} \alpha_k d_k \qquad (2.6.3.a)$$

waar

$$\alpha_k = \frac{-d_k^T(Ax_0 + b)}{d_k^TAd_k} = -\frac{d_k^Tg_0}{d_k^TAd_k} \qquad (2.6.3.b)$$

en:

Eigenschap II. Wordt voor de minimalisering van de positief definiete kwadratische vorm $f(x) = \frac{1}{2}x^TAx + b^Tx + c$ gebruik gemaakt van de standaardalgorithme met lijnminimalisering (zie pt. 2.1.5, 2.1.6) en worden de geconjugeerde richtingen $\{d_0,d_1,\ldots,d_{n-1}\}$ gebruikt als successievelijke zoekrichtingen

$$d^{(k)} := d_k \qquad (2.6.4.a)$$

dan geldt voor de staplengte factor $\alpha^{(k)}$ in de (k + 1)-de lijnminimalisering met de notatie $g^{(k)} := \nabla f(x^{(k)})$ voor de gradiënt in punt $x^{(k)}$

$$\alpha^{(k)} = \frac{-d^{(k)T}g^{(k)}}{d^{(k)T}Ad^{(k)}} = \frac{-d^{(k)T}g^{(0)}}{d^{(k)T}Ad^{(k)}} . \qquad (2.6.4.b)$$

2.6.5. Combinatie van de eigenschappen I en II leidt tot de volgende belangrijke uitspraak:

Stelling 2.6.5: Wordt bij de minimalisering van een positief definiete kwadratische vorm $f(x) := \frac{1}{2}x^TAx + b^Tx + c$ gebruik gemaakt van de standaardalgorithme met lijnminimalisering (zie pt. 2.1.5, 2.1.6) en telkens nog niet eerder gebruikte A-geconjugeerde richtingen als zoekrichtingen dan wordt het minimum van $f(x)$ in ten hoogste n iteratiestappen gevonden. De volgorde waarin de geconjugeerde richtingen gebruikt worden speelt hierbij geen rol.

2.6.6. De uitspraak in pt. 2.6.5 kan ook worden afgeleid met behulp van de volgende overwegingen: Veronderstel dat ten behoeve van de minimalisering van de positief definiete kwadratische vorm $f(x) := \frac{1}{2}x^TAx + b^Tx + c$ bij ieder van de eerste n iteratiestappen een stap van willekeurige lengte (d.i. $\alpha^{(k)}$ willekeurig) wordt gezet langs een nog niet eerder gebruikte A-geconjugeerde richting $d^{(k)} := d_k$ en dat gebruik wordt gemaakt van de volgende (in deze theorie gebruikelijke) notatie

$$s^{(k)} := x^{(k+1)} - x^{(k)} \qquad S^{(k)} := [s^{(0)}, s^{(1)}, \ldots, s^{(k-1)}] \tag{2.6.5}$$

$$y^{(k)} := g^{(k+1)} - g^{(k)} \qquad Y^{(k)} := [y^{(0)}, y^{(1)}, \ldots, y^{(k-1)}] \tag{2.6.6}$$

waar dus

$$s^{(k)} := \alpha^{(k)} d^{(k)} \qquad D^{(k)} := [d^{(0)}, d^{(1)}, \ldots, d^{(k-1)}] \,. \tag{2.6.7}$$

Met

$$y^{(k)} = As^{(k)} \qquad Y^{(k)} = AS^{(k)} \tag{2.6.8}$$

$$S^{(k)} = D^{(k)}[\alpha_D^{(k)}] := D^{(k)} \begin{bmatrix} \alpha^{(0)} & & \\ & \alpha^{(1)} & \\ & & \ddots \\ & & & \alpha^{(k-1)} \end{bmatrix} \tag{2.6.9}$$

volgen onmiddellijk de relaties

$$s^{(\ell)T}As^{(k)} = \alpha^{(\ell)}\alpha^{(k)} d^{(\ell)T}Ad^{(k)} = 0 \qquad k \neq \ell \tag{2.6.10}$$

$$Y^{(\ell)T}s^{(k)} = \alpha^{(k)} Y^{(\ell)T}d^{(k)} = 0 \qquad \ell \leq k \tag{2.6.11}$$

en

$$s^{(\ell)T}y^{(k)} = [\alpha_D^{(\ell)}]D^{(\ell)T}y^{(k)} = 0 \qquad \ell \leq k \tag{2.6.12}$$

Met deze terminologie kan de fundamentele eigenschap die ten grondslag ligt aan het gebruik van geconjugeerde richtingen bij minimaliseringsalgorithmen (en die in de Engelse literatuur bekend staat als de Expanding Subspaces Theorem) eenvoudig als volgt worden geformuleerd (en bewezen).

Stelling 2.6.6: Wordt bij minimalisering van een positief definiete kwadratische vorm gebruik gemaakt van onderling A-geconjugeerde richtingen $d^{(0)},d^{(1)},\ldots,d^{(k)}$ als zoekrichtingen en van lijnminimalisering voor de stapgroottebepaling, d.w.z. wordt in iedere stap geëist dat

$$g^{(\ell+1)T}d^{(\ell)} = 0 \qquad 0 \leq \ell \leq k \leq n-1 \tag{2.6.13}$$

dan geldt voor iedere k, $k = 0,1,\ldots,n-1$

$$g^{(k+1)T}D^{(k+1)} = 0 \tag{2.6.14}$$

en dus ook in het bijzonder

$$g^{(n)T}D^{(n)} = 0 \tag{2.6.15}$$

waaruit volgt, omdat $D^{(n)}$ de gehele $\mathbb{R}^n$ opspant dat

$$g^{(n)} = 0 \tag{2.6.16}$$

2.6.7. De in de voorgaande punten 2.6.5, 2.6.6 gevonden convergentie in (hoogstens) n stappen bij toepassing van de algorithme op de minimalisering van een positief definiete kwadratische vorm wordt in de literatuur soms aangeduid met de naam kwadratische convergentie. Dit is enigszins verwarrend daar dezelfde naam ook wordt gebruikt voor de convergentie van b.v. de algorithme van de methode van Newton (zie pt. 2.1.14, 2.5.9) die bij toepassing op een positief definiete kwadratische vorm (vgl. pt. 2.5.1) convergentie geeft in één stap. In verband met deze mogelijke verwarring zegt men van algorithmen die toegepast op de minimalisering van een positief definiete kwadratische vorm in (hoogstens) n stappen convergentie leveren wel dat zij de Qn-eigenschap dan wel n-stappen-convergentie vertonen.

2.6.8. De overweging dat de objectfuncties van veel practische minimaliseringsproblemen in de omgeving van het minimum benaderd kunnen worden door positief definiete kwadratische vormen, gecombineerd met de wetenschap dat toepassing op kwadratische vormen van de standaardalgorithme met lijnminimalisering bij gebruik van geconjugeerde richtingen als zoekrichtingen convergentie geeft

in (hoogstens) n stappen, heeft ertoe geleid dat een groot aantal minimaliseringsalgorithmen zijn ontwikkeld die gebaseerd zijn op het gebruik van geconjugeerde richtingen. Enkele van de meest bekende daarvan worden hieronder besproken t.w.:

1) de methode van Powell (1964) zonder afgeleiden [2.6.3], [2.6.4]
2) de Partan-methode [2.6.5]
3) de geconjugeerde gradiënt-methoden van resp.
   a) Hestenes-Stiefel [2.6.6]
   b) Fletcher-Reeves [2.6.7]
4) de geprojecteerde gradiënt-methode van Pearson [2.6.8]

De algorithmen van deze methoden verschillen onderling voornamelijk in de wijze waarop de geconjugeerde richtingen worden gegenereerd.

Methode van Powell (1964)

2.6.9. Gerelateerd aan het idee dat geconjugeerde richtingen de generalisaties zijn van de toegevoegde middellijnen van een ellips is de eenvoudig aan te tonen eigenschap dat voor de verbindingslijn s van twee minima van een positief definiete kwadratische vorm:

$$f(x) := \tfrac{1}{2}x^T Ax + b^T x + c$$

op twee evenwijdige lineaire variëteiten

$$X_1 := \{x \mid x = x_1 + \sum_{j=0}^{i} \lambda_j d_j , \quad i \leq n-2, \ \lambda_j \in \mathbb{R}^1\}$$

en

$$X_2 := \{x \mid x = x_2 + \sum_{j=0}^{i} \lambda_j d_j , \quad i \leq n-2, \ \lambda_j \in \mathbb{R}^1\}$$

met

$$X_1 \cap X_2 = \emptyset$$

geldt dat deze A-geconjugeerd is met alle richtingsvectoren van de lineaire variëteiten, d.i.

$$s^T A d_j = 0 , \qquad j = 0,\ldots,i .$$

Deze eigenschap impliceert dat geconjugeerde richtingen kunnen worden gegenereerd door de verbindingslijnen te bepalen van minima op evenwijdige lineaire variëteiten. Dit idee vormt de basis van de methode van Powell [2.6.3].

2.6.10. De algorithme van de methode van Powell bestaat in zijn basisvorm daaruit dat telkens n lijnminimaliseringen worden uitgevoerd langs n lineair onafhankelijke richtingen. Hierna wordt de verbindingslijn getrokken tussen het startpunt en het laatste bepaalde lijnminimum. Lijnminimalisering langs deze verbindingslijn geeft het volgende startpunt. Voordat hierna met de volgende serie lijnminimaliseringen begonnen wordt, wordt de eerste van de originele richtingsvectoren weggelaten en worden de andere richtingen hernummerd door de indices met 1 te verlagen. Als n-de richting wordt ingevoerd de richting $y := x_n - x_0 / \| x_n - x_0 \|$. Schematisch

iteratie (0) : $d_0, d_1, \ldots, d_{n-2}, d_{n-1}$

iteratie (1) : $d_1, d_2, \ldots, d_{n-1}, y_0$

iteratie (2) : $d_2, d_3, \ldots, d_{n-1}, y_0, y_1$

iteratie (3) : $d_3, d_4, \ldots, y_0, y_1, y_2$

. . . . . etc .

Na k iteratieslagen zijn op deze wijze bij minimalisering van een positief definiete kwadratische vorm de laatste k-1 richtingen onderling geconjugeerd.

2.6.11. Bij toepassing van de in pt. 2.6.10 geschetste basisalgorithme van de methode van Powell is het geenszins gegarandeerd dat de lineaire onafhankelijkheid van de zoekrichtingen gehandhaafd blijft (zie [2.6.4]). Teneinde deze voor een correcte minimalisering noodzakelijke eigenschap te handhaven modificeerde Powell zijn basisalgorithme zodanig dat niet telkens de eerste van de oude zoekrichtingen wordt vervangen doch in plaats daarvan een zoekrichting door wiens vervanging de lineaire onafhankelijkheid niet vermindert.

Voor de uitwerking van zijn modificatie ging Powell [2.6.3] als volgt te werk: Als maat voor de lineaire onafhankelijkheid koos hij in principe de absolute waarde van de determinant van de matrix $D := [d_0, d_1, \ldots, d_{n-1}]$ met als kolommen de zoekrichtingen nadat deze zo zijn geschaald dat voor alle j, $j = 0, \ldots, n-1$ geldt

$$d_j^T A d_j = \|d_j\|_A^2 = 1 \qquad (2.6.17)$$

waar A de symmetrische positief definiete matrix is van de te minimaliseren kwadratische vorm. Aangetoond kan worden dat deze determinant zijn maximum absolute waarde dan en slechts dan bereikt wanneer de geschaalde richtingsvectoren (kolommen van de matrix) onderling A-geconjugeerd zijn. Teneinde dit determinantcriterium toe te kunnen passen zonder de matrix A expliciet te kennen maakte Powell gebruik van een andere eigenschap van de geschaalde richtingsvectoren, nl. dat voor de stapgroottefactor α_j die correspondeert met het <u>minimum</u> x_{j+1} van de positief definiete kwadratische vorm

$$f(x) := \tfrac{1}{2}x^T A x + b^T x + c$$

langs de lijn $x := x_j + \alpha d_j$ geldt dat

$$x_{j+1} = x_j + \alpha_j d_j$$

en

$$\alpha_j = \sqrt{2f(x_j) - 2f(x_{j+1})} \cdot \operatorname{sgn}[(x_{j+1} - x_j)^T d_j] \qquad (2.6.18)$$

Als y de overeenkomstig geschaalde richtingsvector is die correspondeert met de verbindingslijn $x_n - x_0$ tussen het startpunt en het laatste lijnminimum in dezelfde iteratie, d.i. $y := (x_n - x_0)/\|x_n - x_0\|_A$ en geldt dat

$$x_n - x_0 =: \mu y = \sum_{j=0}^{n-1} \alpha_j d_j = \|x_n - x_0\|_A \left(\frac{x_n - x_0}{\|x_n - x_0\|_A}\right) \qquad (2.6.19)$$

dan zal vervanging van een van de oude richtingsvectoren d_j door de nieuwe richting y resulteren in een verandering van de absolute waarde van de determinant van de matrix van richtingsvectoren met een factor $|\alpha_j/\mu|$. Powell maakte van deze eigenschappen gebruik door na n lijnminimaliseringen voor vervanging uit te kiezen die richtingsvector d_ℓ met index ℓ waarvoor geldt

$$|\alpha_\ell| = \max_j |\alpha_j| = \max_j \sqrt{2(f(x_j) - f(x_{j+1}))} =: \sqrt{2\Delta} \qquad (2.6.20)$$

en vervanging alleen te doen plaatsvinden in het geval dat

$$\sqrt{2\Delta} > \mu \qquad (2.6.21)$$

Aangezien

$$\sqrt{2\Delta} \leq \sqrt{2(f(x_0) - f(x_{n+1}))} \tag{2.6.22}$$

en

$$\mu = \sqrt{2(f(x_0) - f(x_{n+1}))} \pm \sqrt{2(f(x_n) - f(x_{n+1}))} \tag{2.6.23}$$

zal alleen aan (2.6.21) kunnen worden voldaan indien in deze uitdrukking voor μ het minteken van toepassing is, d.w.z. indien

$$\alpha_n = \sqrt{2(f(x_n) - f(x_{n+1}))}\,\mathrm{sgn}\,((x_{n+1} - x_n)^T(x_n - x_0)) > 0 \tag{2.6.24}$$

Wordt niet aan (2.6.21) en (2.6.24) voldaan dan vindt geen vervanging plaats en worden de oude zoekrichtingen opnieuw gebruikt in de volgende iteratieslag. Teneinde te voorkomen dat in dit laatste geval onnodig een nieuw lijnminimum zou worden bepaald, verving Powell in het vervangingscriterium de echte functiewaarde in het lijnminimum x_{n+1} door de parabolische benadering f_s voor deze waarde gebaseerd op de functiewaarden $f_1 := f(x_0)$, $f_2 := f(x_n)$ en $f_3 := f(x_0 + 2(x_n - x_0))$, d.i.:

$$f_s := f_2 - \frac{1}{8}\frac{(f_1 - f_3)^2}{(f_1 - 2f_2 + f_3)} . \tag{2.6.25}$$

Hiermee wordt de conditie dat <u>geen</u> vervanging van de oude zoekrichtingen plaats mag vinden indien aan een of beide van de volgende voorwaarden voldaan wordt:

a) $$f_3 \geq f_1 \tag{2.6.26}$$

b) $$(f_1 - 2f_2 + f_3)(f_1 - f_2 - \Delta)^2 \geq \tfrac{1}{2}\Delta(f_1 - f_3)^2 . \tag{2.6.27}$$

In de practijk wordt deze laatste "truc" van Powell niet altijd gebruikt.

2.6.12. Als eerste zoekrichtingen worden bij de methode van Powell veelal de coördinaatrichtingen gebruikt. Uitgaande hiervan geeft uitwerking van het bovenstaande het volgende resultaat:

<u>Algorithme van de methode van Powell</u> (1964)

(0) zet $x_0^{(0)}$:= gegeven startpunt, en voor j := 0,1,...,n-1 zet
$d_j^{(0)} := e_j$ (j-de eenheidsvector); zet k := 0

(i) voor $j := 0,1,\ldots,n-1$ bepaal $x_{j+1}^{(k)}$ en $\alpha_j^{(k)}$ zodat

$$f(x_{j+1}^{(k)}) = f(x_j^{(k)} + \alpha_j^{(k)} d_j^{(k)}) := \min_{\alpha} f(x_j^{(k)} + \alpha d_j^{(k)})$$

en

$$\Delta_j^{(k)} := f(x_j^{(k)}) - f(x_{j+1}^{(k)})$$

(ii) is $x_n^{(k)}$ optimaal dan klaar; zo niet, bepaal

$$\Delta^{(k)} := \Delta_\ell^{(k)} := \max_j \Delta_j^{(k)}$$

(iii) zet $f_1^{(k)} := f(x_0^{(k)})$, $f_2^{(k)} := f(x_n^{(k)})$ en bepaal

$$f_3^{(k)} := f(x_n^{(k)} + (x_n^{(k)} - x_0^{(k)}))$$

(iv) check of: $f_3^{(k)} \geq f_1^{(k)}$ en/of

$$(f_1^{(k)} - 2f_2^{(k)} + f_3^{(k)})(f_1^{(k)} - f_2^{(k)} - \Delta^{(k)})^2 \geq \tfrac{1}{2}\Delta^{(k)}(f_1^{(k)} - f_3^{(k)})^2$$

a) zo ja, voor $j := 0,1,\ldots,n-1$, zet $d_j^{(k+1)} := d_j^{(k)}$, zet $k := k + 1$, $x_0^{(k+1)} := x_n^{(k)}$ en ga terug naar stap (i).

b) zo nee, bepaal $y^{(k)} := (x_n^{(k)} - x_0^{(k)})/\|x_n^{(k)} - x_0^{(k)}\|_A$ en $\alpha_n^{(k)}$ zo dat

$$f(x_n^{(k)} + \alpha_n^{(k)} y^{(k)}) := \min_{\alpha} f(x_n^{(k)} + \alpha y^{(k)})$$

zet

$$x_0^{(k+1)} := x_{n+1}^{(k)} := x_n^{(k)} + \alpha_n^{(k)} y^{(k)}$$

$$d_\ell^{(k+1)} := y^{(k)}$$

en voor $j := 0,1,\ldots,\ell-1,\ell+1,\ldots,n-1$

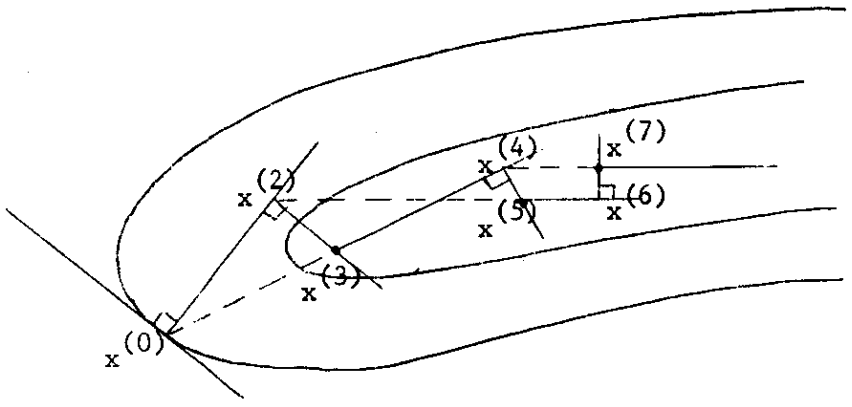
$$d_j^{(k+1)} := d_j^{(k)}$$

zet $k := k + 1$ en ga terug naar stap (i).

.6.13. De algorithme van Powell is een van de bekendste methoden voor het minimaliseren van functies van meer variabelen waarbij geen gebruik wordt gemaakt van analytische uitdrukkingen voor de afgeleiden. Als zodanig behoort de algorithme eigenlijk thuis bij de in § 2.3 besproken "direct-search"-methoden. Als geconjugeerde-richtingenmethode convergeert de methode van Powell echter in het algemeen sneller dan de daar besproken methoden. Sinds de publicatie van de methode zijn nog een aantal andere algorithmen gepubliceerd die eveneens gebruik maken van geconjugeerde richtingen en eveneens geen gebruik van analytische uitdrukkingen voor de afgeleiden [2.6.4], [2.6.9]. Van deze methoden wordt meestal geclaimed dat zij efficiënter zijn dan de methode van Powell. Desondanks blijkt de algorithme nog redelijk vaak toegepast te worden.

Partan-methode

.6.14. Een andere algorithme die juist als de methode van Powell gebaseerd is op de geometrische interpretatie van de geconjugeerde richtingen is de zgn. Partan-methode, welke vooral bekendheid heeft gekregen door het werk van Shah, Buehler en Kempthorne [2.6.5]. Deze methode is gebaseerd op de eigenschap (zie pt. 2.6.9) dat de verbindingslijn van twee evenwijdige raakhypervlakken (eng: parallel tangents) aan de niveauoppervlakken van een positief definiete kwadratische vorm geconjugeerd is met alle rechten in die hypervlakken. De methode bestaat daaruit dat afwisselend als zoekrichtingen worden gebruikt de (negatieve) locale gradiënt en de verbindingslijn tussen het lijnminimum langs die gradiënt en het minimum langs de op een na laatste verbindingslijn (zie schets: bij de nummering wordt de index 1 i.v.m. de systematiek niet gebruikt)



Aangetoond kan worden [2.6.10] dat bij minimalisering van een positief definiete kwadratische vorm de successievelijke verbindingslijnen $x^{(2)} - x^{(0)}$, $x^{(4)} - x^{(2)}$, $x^{(6)} - x^{(4)}$,... onderling geconjugeerd zijn.

2.6.15. Bij de Partan-methode moeten in alle "even" iteratiepunten de gradiënt van de functie worden berekend (anders dan bij de methode van Powell). Dit impliceert dat in een te realiseren computerprogramma een procedure moet worden ingebouwd voor het bepalen van deze gradiënt. In het geval dat een dergelijke faciliteit om analytische dan wel numerieke afgeleiden te bepalen, beschikbaar is, bestaan er echter efficiëntere methoden zoals o.a. de hierna te bespreken geconjugeerde gradiëntmethoden en de later te bespreken quasi-Newtonmethoden. Dit feit heeft ertoe geleid dat de Partan-methode nu niet veel meer wordt toegepast en eigenlijk op dit moment voornamelijk van historisch belang is.

Geconjugeerde-gradiëntmethoden

2.6.16. Een min of meer voor de hand liggende methode voor het genereren van A-geconjugeerde of A-orthogonale richtingsvectoren is de toepassing van het analogon voor A-orthogonalisatie van de Gram-Schmidt-orthogonalisatie algorithme: Uitgaande van een stelsel lineair onafhankelijke vectoren $p_0, p_1, \ldots, p_{n-1}$ in $\mathbb{R}^n$ kan een volledig stelsel A-geconjugeerde richtingen worden gegenereerd met behulp van de algorithme:

$$d_0 := p_0$$

$$d_{k+1} := p_{k+1} - \sum_{j=0}^{k} \frac{p_{k+1}^T A d_j}{d_j^T A d_j} d_j \quad , \quad k = 0,\ldots,n-2 \; . \tag{2.6.28}$$

Stapsgewijze toepassing van deze algorithme op de negatieve gradiënten gevonden in de punten $x^{(k+1)}$ bij toepassing van de standaardalgorithme met lijnminimalisering (pt. 2.1.5, 2.1.6) voor de minimalisering van de positief definiete kwadratische vorm $f(x) := \frac{1}{2}x^T A x + b^T x + c$ resulteert in een volledig stelsel geconjugeerde zoekrichtingen

$$d^{(0)} := -g^{(0)}$$

$$d^{(k+1)} := -g^{(k+1)} + \sum_{j=0}^{k} \frac{g^{(k+1)T} A d^{(j)}}{d^{(j)T} A d^{(j)}} d^{(j)} \quad , \quad k = 0,\ldots,n-2 \; . \tag{2.6.29}$$

2.6.17. Een belangrijke eigenschap van de in pt. 2.6.16 geformuleerde generatie van geconjugeerde zoekrichtingen is het volgende resultaat:

Stelling 2.6.17: Wordt de Gram – Schmidt A-orthogonalisatie procedure toegepast voor het genereren van A-geconjugeerde zoekrichtingen bij het minimaliseren van een positief definiete kwadratische vorm zoals beschreven in het voorgaande punt 2.6.16 dan geldt dat de successievelijk gevonden gradiënten $g^{(k+1)} := \nabla f(x^{(k+1)})$ een orthogonaal stelsel vormen, d.i.

$$g^{(k+1)T} g^{(j)} = 0 , \quad j \leq k,\ k = 0,1,\ldots,k_{max} \leq n-1 . \tag{2.6.30}$$

Bewijs. Met de in pt. 2.6.6 gevonden eigenschap dat $g^{(k+1)T} D^{(k+1)} = 0$ en de observatie dat $g^{(j)} \in \{D^{(k+1)}\}$ voor $0 \leq j \leq k$ volgt het resultaat direct. □

2.6.18. Substitutie van deze orthogonaliteitsrelatie in de formules voor het genereren van geconjugeerde richtingen resulteert in de volgende twee † (voor de minimalisering van positief definiete kwadratische vormen equivalente) algorithmen van respectievelijk Hestenes-Stiefel:

$$d^{(0)} := -g^{(0)} \tag{2.6.31}$$

$$d^{(k+1)} := -g^{(k+1)} + \frac{g^{(k+1)T} y^{(k)}}{d^{(k)T} y^{(k)}} d^{(k)} = -g^{(k+1)} + \frac{g^{(k+1)^T} Ad^{(k)}}{d^{(k)T} Ad^{(k)}} d^{(k)}$$

en van Fletcher-Reeves:

$$d^{(0)} := -g^{(0)} \tag{2.6.32}$$

$$d^{(k+1)} := -g^{(k+1)} + \frac{g^{(k+1)T} g^{(k+1)}}{g^{(k)T} g^{(k)}} d^{(k)} = -g^{(k+1)} + \frac{\| g^{(k+1)} \|^2}{\| g^{(k)} \|^2} d^{(k)}$$

Deze algorithmen vormen de basis voor de volgende twee bekende minimaliseringsmethoden:

Methode van Hestenes-Stiefel[2.6.6]

(0) zet $x^{(0)}$:= gegeven startpunt en k := 0

(i) bepaal de functiewaarde $f(x^{(k)})$ en de gradiënt $g^{(k)} := \nabla f(x^{(k)})$

† Een derde bekende formule is die van Polak-Ribière [2.6.13]:

$$d^{(k+1)} := -g^{(k+1)} + \frac{g^{(k+1)T} y^{(k)}}{g^{(k)T} g^{(k)}} d^{(k)} .$$

(ii) ga na of $x^{(k)}$ optimaal is; zo ja, dan klaar; zo nee, dan

(iii) bepaal (als $k \geq 1$) $y^{(k-1)} := g^{(k)} - g^{(k-1)}$ en als zoekrichting

$$d^{(k)} := - g^{(k)} + \frac{g^{(k)T} y^{(k-1)}}{d^{(k-1)T} y^{(k-1)}} d^{(k-1)} \qquad \text{als } k \geq 1$$

$$:= - g^{(k)} \qquad \text{als } k = 0$$

(iv) bepaal een staplengte (-factor) $\alpha^{(k)}$ uit

$$f(x^{(k)} + \alpha^{(k)} d^{(k)}) = \min\{f(x^{(k)} + \alpha d^{(k)}) \mid \alpha \in \mathbb{R}^1_+\}$$

(v) zet $x^{(k+1)} := x^{(k)} + \alpha^{(k)} d^{(k)}$, $k := k + 1$ en ga terug naar stap (i)

Methode van Fletcher-Reeves [2.6.7]

(0)-(ii) als bij de methode van Hestenes-Stiefel

(iii) bepaal als zoekrichting

$$d^{(k)} := - g^{(k)} + \frac{g^{(k)T} g^{(k)}}{g^{(k-1)T} g^{(k-1)}} d^{(k-1)} \qquad \text{als } k \geq 1$$

$$:= - g^{(k)} \qquad \text{als } k = 0$$

(iv)-(v) als boven bij de methode van Hestenes-Stiefel.

Geprojecteerde-gradiëntmethode van Pearson

2.6.19. Een andere mogelijkheid voor het genereren van een nieuwe A-geconjugeerde richting uit de gradiënt in het nieuwste iteratiepunt (bij minimalisering van een positief definiete kwadratische vorm met de standaardalgorithme met lijnminimalisering) werd bestudeerd door Pearson [2.6.8]. Uit de observatie dat het doel van dit generatieproces een richting $d^{(k+1)}$ is waarvoor geldt (vgl. notatie in pt. 2.6.6)

$$D^{(k+1)T} A d^{(k+1)} := [\alpha_D^{(k+1)}]^{-1} Y^{(k+1)T} d^{(k+1)} = 0 \qquad (2.6.33)$$

volgt direct dat een voor de hand liggende manier hiervoor is de projectie van de (negatieve) gradiënt op het orthogonale complement van de lineaire deelruimte opgespannen door de kolommen van $Y^{(k+1)}$, d.i.

$$\{Y^{(k+1)}\}^{\perp} := \{y \in \mathbb{R}^n \mid y^T y^{(j)} = 0,\ j = 0,\ldots,k\} \ . \qquad (2.6.34)$$

Deze projectie wordt gegeven door de symmetrische idempotente lineaire transformatie

$$d^{(k+1)} := -H^{(k+1)}g^{(k+1)} \tag{2.6.35}$$

met

$$H^{(k+1)} := [I - Y^{(k+1)}[Y^{(k+1)T}Y^{(k+1)}]^{-1}Y^{(k+1)T}]. \tag{2.6.36}$$

De minimaliseringsmethode welke gebruik maakt van zoekrichtingen gegenereerd met deze projectieformule, wordt in de literatuur soms aangeduid als de gradiëntprojectiemethode (van Pearson). (vgl. [2.6.11]).

2.6.20. De projectieformule van de methode van Pearson (pt. 2.6.19) kan ook op recursieve manier worden bepaald. Hiervoor geldt de volgende uitspraak

Stelling 2.6.20. De projectieoperator voor het genereren van nieuwe geconjugeerde richtingen in de algorithme van de gradiëntprojectiemethode

$$d^{(k+1)} := -[I - Y^{(k+1)}[Y^{(k+1)T}Y^{(k+1)}]^{-1}Y^{(k+1)T}]g^{(k+1)} \tag{2.6.37}$$

kan op recursieve manier worden bepaald met

$$d^{(k+1)} := -H^{(k+1)}g^{(k+1)} \tag{2.6.38}$$

met

$$H^{(0)} := I$$

en

$$H^{(k+1)} := H^{(k)} - \frac{H^{(k)}y^{(k)}y^{(k)T}H^{(k)}}{y^{(k)T}H^{(k)}y^{(k)}} . \tag{2.6.39}$$

Bewijs. Het bewijs is een directe toepassing van het volgende lemma:

2.6.21. Lemma 2.6.21: Zij voor iedere k, k = 0,...,n-1, $\{Y^{(k+1)}\}$ de (k+1)-dimensionale deelruimte in $\mathbb{R}^n$ opgespannen door de kolommen $y^{(j)}$, j = 0,...,k, van de n × (k+1)-dimensionale matrix $Y^{(k+1)}$, zij $\{Y^{(k+1)}\}^{\perp}$ het orthogonale complement van $\{Y^{(k+1)}\}$ en zij $H^{(k+1)}$ de orthogonale projectie van $\mathbb{R}^n$ op $\{Y^{(k+1)}\}^{\perp}$ d.w.z.

$$H^{(k+1)}y = 0 \quad \text{als } y \in \{Y^{(k+1)}\}$$
$$H^{(k+1)}y = y \quad \text{als } y \in \{Y^{(k+1)}\}^{\perp} \tag{2.6.40}$$

dan geldt dat $H^{(k+1)}$ recursief kan worden bepaald uit (2.6.39)

$$H^{(0)} := I$$
$$H^{(k+1)} := H^{(k)} - \frac{H^{(k)}y^{(k)}y^{(k)T}H^{(k)}}{y^{(k)T}H^{(k)}y^{(k)}} \quad .$$

2.6.22. Het beoogde doel van de gradiëntprojectieoperator, d.i. het genereren van een richting $d^{(k+1)}$ zodat

$$Y^{(k+1)T}d^{(k+1)} = 0 \tag{2.6.41}$$

kan ook worden bereikt met de algemenere projectieoperator

$$H^{(k+1)} := R[I - Y^{(k+1)}[Y^{(k+1)T}R\,Y^{(k+1)}]^{-1}Y^{(k+1)T}R] \tag{2.6.42}$$

waar R een willekeurige symmetrische positief definiete matrix voorstelt. Deze projectieoperator beeldt het R-orthogonale complement van $\{Y^{(k+1)}\}$, d.i.

$$\{Y^{(k+1)}\}^{\perp R} := \{y \in \mathbb{R}^n \mid y^TRy^{(j)} = 0,\ j = 0,\ldots,k\} \tag{2.6.43}$$

af op het orthogonale complement van $\{Y^{(k+1)}\}$. Immers

$$H^{(k+1)}y = 0 \quad \text{als } y \in \{Y^{(k+1)}\}$$
$$H^{(k+1)}y = Ry \quad \text{als } y \in \{Y^{(k+1)}\}^{\perp R} \tag{2.6.44}$$

en

$$Ry \in \{Y^{(k+1)}\}^{\perp} \text{ als } y \in \{Y^{(k+1)}\}^{\perp R} \quad . \tag{2.6.45}$$

In recursieve vorm kan deze algemene projectieoperator worden bepaald uit

$$H^{(0)} := R$$
$$H^{(k+1)} := H^{(k)} - \frac{H^{(k)}y^{(k)}y^{(k)T}H^{(k)}}{y^{(k)T}H^{(k)}y^{(k)}} \quad . \tag{2.6.46}$$

Opgemerkt kan worden dat men in dit geval te doen heeft met een symmetrische, niet-idempotente lineaire operator.

2.6.23. Uitwerking van de standaardalgorithme met lijnminimalisering met zoekrichtingen bepaald volgens de gradiëntprojectie van Pearson levert het volgende resultaat:

Algorithme van de gradiëntprojectiemethode van Pearson

(0) zet $x^{(0)}$:= gegeven startpunt en zet k := 0

(i) bepaal de functiewaarde $f(x^{(k)})$ en de gradiënt $g^{(k)} := \nabla f(x^{(k)})$

(ii) ga na of $x^{(k)}$ optimaal; zo ja, dan klaar; zo nee, dan

(iii) bepaal (als $k \geq 1$) $y^{(k-1)} := g^{(k)} - g^{(k-1)}$ en

$$H^{(k)} := H^{(k-1)} - \frac{H^{(k-1)}y^{(k-1)}y^{(k-1)T}H^{(k-1)}}{y^{(k-1)T}H^{(k-1)}y^{(k-1)}} \qquad \text{als } k \geq 1$$

$$:= R \qquad \text{als } k = 0$$

en als zoekrichting

$$d^{(k)} = - H^{(k)}g^{(k)}$$

(iv) bepaal een staplengte (-factor) $\alpha^{(k)}$ uit

$$f(x^{(k)} + \alpha^{(k)}d^{(k)}) = \min\{f(x^{(k)} + \alpha d^{(k)}) \mid \alpha \in \mathbf{R}_+^1\}$$

(v) zet $x^{(k+1)} := x^{(k)} + \alpha^{(k)}d^{(k)}$, k := k + 1 ga terug naar stap (i).

Vergelijking Gram-Schmidt A-orthogonalisatie en Pearson's projectie

2.6.24. Ter vergelijking met Pearson's projectieformule (pt. 2.6.19, 2.6.20) kan de iteratieformule van de Gram-Schmidt A-orthogonalisatieprocedure (pt. 2.6.16)

$$d^{(k+1)} := -g^{(k+1)} + \sum_{j=0}^{k} \frac{d^{(j)T}Ag^{(k+1)}}{d^{(j)T}Ad^{(j)}} d^{(j)} \qquad (2.6.47)$$

ook worden geschreven in de vorm

$$d^{(k+1)} := - H^{(k+1)}g^{(k+1)}$$

waarin dan

$$H^{(k+1)} := I - \sum_{j=0}^{k} \frac{d^{(j)}d^{(j)T}A}{d^{(j)T}Ad^{(j)}}$$

en

$$H^{(0)} := I \qquad (2.6.48)$$

ofwel, equivalent

$$H^{(k+1)} := [I - D^{(k+1)}[D^{(k+1)T}AD^{(k+1)}]^{-1}D^{(k+1)T}A] \tag{2.6.49}$$

en

$$H^{(0)} = I .$$

Uit deze formulering blijkt duidelijk dat voor de operator $H^{(k+1)}$ geldt

$$\begin{aligned} H^{(k+1)}d &= 0 \quad \text{als } d \in \{D^{(k+1)}\} \\ H^{(k+1)}d &= d \quad \text{als } d \in \{D^{(k+1)}\}^{\perp A} . \end{aligned} \tag{2.6.50}$$

Een met de recursieve vorm van de gradiëntprojectieoperator van Pearson corresponderende recursieve vorm van de Gram—Schmidt A-orthogonalisatieprocedure wordt gegeven door

$$\begin{aligned} H^{(k+1)} &:= H^{(k)} - \frac{H^{(k)}d^{(k)}d^{(k)T}H^{(k)T}AH^{(k)}}{d^{(k)T}H^{(k)T}AH^{(k)}d^{(k)}} \\ H^{(0)} &:= I \end{aligned} \tag{2.6.51}$$

of equivalent, omdat $d^{(k)} \in \{D^{(k)}\}^{\perp A}$ (anders dan $y^{(k)}$ waarvoor niet noodzakelijk geldt dat $y^{(k)} \in \{Y^{(k)}\}^{\perp}$)

$$\begin{aligned} H^{(k+1)} &:= H^{(k)} - \frac{d^{(k)}d^{(k)T}A}{d^{(k)T}Ad^{(k)}} \\ H^{(0)} &:= I . \end{aligned} \tag{2.6.52}$$

Opgemerkt kan worden dat de Gram—Schmidt A-orthogonalisatieoperator een <u>niet-symmetrische, idempotente</u> lineaire operator is.

2.6.25. Als q_1 de vector is die resulteert bij toepassing van de k-de Gram—Schmidt A-orthogonale projectieoperator op een willekeurige vector $p \in \mathbb{R}^n$, d.i.

$$q_1 := [I - D^{(k)}[D^{(k)T}AD^{(k)}]^{-1}D^{(k)T}A]p \tag{2.6.53}$$

en q_2 de vector is die resulteert bij toepassing van corresponderende k-de projectieoperator van de methode van Pearson, d.i.

$$q_2 := [I - Y^{(k)}[Y^{(k)T}Y^{(k)}]^{-1}Y^{(k)T}]p \qquad (2.6.54)$$

dan zullen de vectoren q_1 en q_2 in het algemeen niet aan elkaar gelijk zijn. Geldt echter dat de lineaire deelruimte opgespannen door p en de kolommen van $D^{(k)}$ dezelfde is als die opgespannen door p en de kolommen van $Y^{(k)}$, d.i.

$$\{p,D^{(k)}\} = \{p,Y^{(k)}\} \qquad (2.6.55)$$

dan volgt dat

$$q_1 = \alpha q_2 \cdot \qquad (2.6.56)$$

Zowel bij de toepassing van de geconjugeerde gradiëntmethoden gebaseerd op de Gram–Schmidt A-orthogonalisatieprocedure (zie pt. 2.6.16), als bij de toepassing van de gradiëntprojectiemethode van Pearson voor de minimalisering van een positief definiete kwadratische vorm, geldt (mits $d^{(0)} := -g^{(0)}$) dat

$$\{g^{(k+1)},D^{(k+1)}\} = \{g^{(k+1)},Y^{(k+1)}\} = \{g^{(k+1)},g^{(k)},\ldots,g^{(0)}\} \cdot (2.6.57)$$

Met deze eigenschap kan worden aangetoond dat de op de Gram–Schmidt A-orthogonalisatieprocedure berustende geconjugeerde gradiëntmethoden en de gradiëntprojectiemethode van Pearson (met R = I) toegepast op de minimalisering van dezelfde positief definiete kwadratische vorm bij hetzelfde startpunt in het geval van exacte lijnminimalisering dezelfde zoekrichtingen en dezelfde iteratiepunten genereren.

Niet kwadratische objectfuncties

2.6.26. Bij niet-kwadratische objectfuncties is de tweede afgeleiden-matrix of Hessiaan niet overal hetzelfde en kan men dan ook op zijn hoogst nog slechts lokaal over onderling geconjugeerde richtingen praten.
Het gebruik van geconjugeerde-richtingen-methoden, zo die al te realiseren zijn, voor minimalisering van niet-kwadratische objectfunkties geeft dan ook niet die gunstige eindige convergentieresultaten als in het geval van kwadratische objectfuncties. De omstandigheid dat de meeste practische objectfuncties van minimaliseringsproblemen zich in de nabijheid van het minimum gedragen als kwadratische functies maakt dat de meeste geconjugeerde-richtingen-methoden desondanks bij toepassing op niet-kwadratische objectfuncties met de voor de hand liggende modificaties toch redelijke, zij het geen eindige, convergentie resultaten opleveren. In het geval van de minimalisering van strikt convexe

functies kan voor een aantal geconjugeerde-richtingen-methoden zogenaamde n-stap-superlineaire convergentie worden aangetoond, dat wil zeggen bewezen kan worden [2.6.12] dat voor de door die algorithmen gegenereerde rij geldt

$$\lim_{c\to\infty} \frac{\| x^{(nc + n)} - x^* \|}{\| x^{(nc)} - x^* \|} = 0 \qquad (2.6.58)$$

Een bijzonder aspect van de toepassing van geconjugeerde-richtingen-methoden op niet-kwadratische minimaliseringsproblemen is de vraag wat te doen na n iteratiestappen. Bij kwadratische objectfuncties geldt volgens het Expanding Subspaces Theorem (Stelling 2.6.6) dat de gradiënt in het nieuw te bepalen punt $x^{(k+1)}$ loodrecht staat op de ruimte opgespannen door alle (k+1) doorlopen zoekrichtingen en derhalve in het bijzonder dat na n stappen de gradiënt in het punt $x^{(n)}$ loodrecht staat op de gehele $\mathbb{R}^n$, dat wil zeggen gelijk is aan nul. Bij de geconjugeerde-gradiënt-methoden komt dit zowel in het geval van de Gram-Schmidt A-orthogonalisatie operator (2.6.48) als in het geval van de projectie operator van Pearson (2.6.36) (of (2.6.39)) duidelijk tot uiting doordat hiervoor geldt

$$H^{(n)} = 0 \qquad (2.6.59)$$

Bij niet kwadratische functies is de analoog te formeren operator $H^{(n)} \neq 0$ en rijst de vraag wat de betekenis van deze $H^{(n)}$ en wat het resultaat is wanneer doorgegaan wordt met het genereren van nieuwe $H^{(n+k)}$'s. Een voor de hand liggende en in de praktijk gebruikelijke oplossing uit dit dilemma is de algorithme na iedere n iteraties opnieuw te herstarten. Men spreekt in dat geval van de "reset-versie" van de betreffende algorithme. In de praktijk blijkt (vgl. [2.6.1]) deze modificatie in de meeste gevallen betere resultaten te geven dan de ongemodificeerde algorithme (zonder herstarten). Het laatste woord is hier echter nog niet over gezegd. Voor theoretische doeleinden zoals voor het doen van convergentie uitspraken worden vrijwel steeds de reset-versies van de algorithmen beschouwd. (zie [2.6.12] - [2.6.15]).

2.6.27. Referenties

[2.6.1]: Zie [1.1.3]: Murray (1972).

[2.6.2]: Zie [2.2.7]: Kowalik en Osborne (1968).

[2.6.3]: Powell, M.J.D.: An efficient method for finding the minimum of a function of several variables without calculating derivatives, Computer J., 7 (1964), pp. 155-162.

[2.6.4]: Zangwill, W.I.: Minimizing a function without calculating derivatives, Computer J., 10 (1967), pp. 293-296.

[2.6.5]: Shah, B.V., Buehler, R.J. and Kempthorne, O.: Some algorithms for minimizing a function of several variables, J. SIAM, 12 (1964), pp. 74-92.

[2.6.6]: Hestenes, M.R. and Stiefel, E.L.: Methods of conjugate directions for solving linear systems, J. Res. Mat. Bur. of Standards, Section B, 49 (1952), pp. 409-436.

[2.6.7]: Fletcher, R. and Reeves, R.M.: Function minimization by conjugate gradients, Computer J., 1 (1964), pp. 149-154.

[2.6.8]: Pearson, J.D.: Variable metric methods for minimization Computer J., 12 (1969), pp. 171-178.

[2.6.9]: Stewart, G.W.: A modification of Davidon's minimization method to accept difference approximation of derivatives. J. ACM 14 (1967), pp. 72-83.

[2.6.10]: Zie [1.1.1]: Luenberger (1973).

[2.6.11]: Zie [1.1.2]: Jacoby, Kowalik and Pizzo (1972).

[2.6.12]: Mc. Cormick, G.P., and Pearson, J.D.: Variable metric methods and unconstrained optimization. in "Optimization" (R. Fletcher, Ed.) Academic Press, New York (1969).

[2.6.13]: Zie [2.1.5]: Polak (1971).

[2.6.14]: Cohen, A.I.: Rate of convergence of several conjugate gradient algorithms, SIAM J. Numer. Anal. 9 (1972), pp. 248-259.

[2.6.15]: Mc Cormick, G.P. and Ritter, K: Alternative proofs of the convergence properties of the conjugate gradient method. J. Opt. Theory & Appl. 13 (1974), pp. 497-518.

§ 2.7. Quasi-Newton -(of variabele-metriek-) methoden I: Algemene theorie

2.7.1. Zoals besproken in § 2.5 is de methode van Newton een zeer snel convergerende methode voor het minimaliseren van functies van meer variabelen (mits de methode convergeert). Een van de nadelen ervan is (vgl. pt 2.5.11) dat in iedere iteratiestap alle elementen van de (inverse van de) Hessiaan van de te minimaliseren functie moeten worden berekend. In principe kan dit ook met numerieke differentiatie in welk geval telkens n + 1 gradiënt+vectoren of $\frac{1}{2}n^2 + \frac{3}{2}n + 1$ functie waarden moeten worden geëvalueerd, die alle slechts worden gebruikt voor het bepalen van een enkele stap. De overweging dat de gradiënt- of functieëvaluaties efficiënter kunnen worden gebruikt in het minimaliseringsproces heeft ertoe geleid dat men minimaliseringsmethoden heeft ontwikkeld die van de gradiënt-informatie in opvolgende iteratiepunten gebruik maken voor het benaderen van de inverse van de Hessiaan. Inplaats van de formule voor de zoekrichting volgens de methode van Newton (vgl. (2.5.8))

$$d^{(k)} := -[G(x^{(k)})]^{-1} g^{(k)} \tag{2.7.1}$$

gebruikt men daartoe de analoge formule

$$d^{(k)} := -H^{(k)} g^{(k)} \tag{2.7.2}$$

waar $H^{(k)}$ een benadering representeert van de inverse van de Hessiaan

$$H^{(k)} \sim [G(x^{(k)})]^{-1}$$

De methoden die gebaseerd zijn op het gebruik van deze formule worden aangeduid met de verzamelnamen quasi-Newton methoden of wel variabele-metriek methoden. Deze tweede naam is ontleend aan de observatie dat een stap volgens formule (2.7.1) juist als bij de methode van Newton (vgl. pt 2.5.1) kan worden opgevat als een stap volgens de methode van de steilste helling in een getransformeerd coördinatenstelsel (d.i. met andere norm of metriek). De iteratief bepaalde en positief definiet veronderstelde matrix $H^{(k)}$ representeert deze telkens veranderende coördinaten-transformatie (d.i. veranderende metriek)

2.7.2. De meeste quasi-Newton methoden worden gekenmerkt door de eigenschap dat voor het verbeteren van de benadering $H^{(k)}$ van de inverse Hessiaan gebruik wordt gemaakt van een aanpassingsformule (Eng.: updating formule) van de karakteristieke vorm

$$H^{(k+1)} := H^{(k)} + C^{(k)} \tag{2.7.3}$$

waarin $C^{(k)}$ een correctie matrix voorstelt die veelal afhankelijk is van de stap $s^{(k)} := x^{(k+1)} - x^{(k)}$ en de verandering in de gradiënt $y^{(k)} := g^{(k+1)} - g^{(k)}$ in de voorgaande (k+1-de) iteratie. Deze correctie matrix $C^{(k)}$ is in de meeste gevallen een matrix van rang 1 of van rang 2. De diverse methoden onderscheiden zich voornamelijk in de vorm van deze correctie matrix en maken nagenoeg alle gebruik van de volgende standaard quasi-Newton algorithme (vgl. pt 2.1.5)

Standaard vorm quasi-Newton algorithmen

(o) zet $x^{(0)}$:= gegeven startpunt, $H^{(0)}$:= gegeven positief definiete beginmatrix (meestal $H^{(0)} := I$: eenheidsmatrix) en k: = 0.

(i) bereken de functiewaarde $f(x^{(k)})$ en de gradient $g^{(k)} := \nabla f(x^{(k)})$

(ii) ga na of $x^{(k)}$ optimaal is; zo ja, dan klaar, zo nee, dan:

(iii) a) als $k > 0$ bepaal $s^{(k-1)} := x^{(k)} - x^{(k-1)}$ en $y^{(k-1)} := g^{(k)} - g^{(k-1)}$ en daarmee de correctie matrix $C^{(k-1)}$ en de nieuwe benadering van de inverse van de Hessiaan

$$H^{(k)} := H^{(k-1)} + C^{(k-1)}$$

b) bepaal als zoekrichting

$$d^{(k)} := -H^{(k)} g^{(k)}$$

(iv) bepaal een staplengte (-factor) $\alpha^{(k)}$. Bijvoorbeeld uit

$$f(x^{(k)} + \alpha^{(k)} d^{(k)}) := \min \{ f(x^{(k)} + \alpha d^{(k)}) \mid \alpha \in \mathbb{R}^1_+ \}$$

(v) zet $x^{(k+1)} := x^{(k)} + \alpha^{(k)} d^{(k)}$, k:= k + 1 en ga terug naar stap (i)

Speciale notatie

2.7.3. Bij de bepaling van de nieuwe benaderingen $H^{(k+1)}$ van de inverse Hessiaan $[G(x^{(k+1)})]^{-1}$, welke hieronder in extenso besproken zal worden, wordt bij nagenoeg alle methoden uitsluitend gebruik gemaakt van grootheden die geindiceerd zijn met de iteratienummers (k+1) en (k) (of (k) en (k-1)), d.w.z. grootheden die betrekking hebben op de huidige (ofwel nieuwe) iteratiestap en op voorafgaande (of oude) iteratiestap. Ter vereenvoudiging van de notatie worden in de literatuur de indices (k) (of (k-1)) van de oude iteratiestap weggelaten en worden die van de nieuwe iteratiestap (k+1)(of(k)) vervangen door een $*$-symbool, d.w.z.

$$H^{(k)} \rightarrow H \qquad H^{(k+1)} \rightarrow H^{*} \tag{2.7.4}$$

De eerder gegeven formules (2.7.2) en (2.7.3) krijgen met deze notatie, bijvoorbeeld, de simpelere vorm

$$d := -Hg \tag{2.7.2'}$$

$$H^{*} := H + C \tag{2.7.3'}$$

M.u.v. op die plaatsen, waar verwarring zou kunnen ontstaan wordt hieronder van deze notatie gebruik gemaakt.

Quasi-Newton- en erfelijkheidsrelaties, Qn-eigenschap

2.7.4. Bij toepassing van een quasi-Newton algorithme voor de minimalisering van de (standaard) positief definiete kwadratische vorm

$$f(x) = \frac{1}{2} x^{T}Ax + b^{T}x + c \tag{2.7.5}$$

zou in het ideale geval (= exacte benadering) voor de benadering $H^{(k+1)}$ van de inverse Hessiaan $[G(x^{(k+1)})]^{-1}$ gelden.

$$H^{(k+1)} = [G(x^{(k+1)})]^{-1} = A^{-1}$$

en daarmee in het bijzonder

$$H^{(k+1)}[g^{(j+1)} - g^{(j)}] = A^{-1}[g^{(j+1)} - g^{(j)}] = x^{(j+1)} - x^{(j)} \tag{2.7.6}$$

of

$$H^{(k+1)}y^{(j)} = s^{(j)} \qquad j = 0, \ldots, n-1$$

In het algemeen is een dergelijke exacte benadering niet te realiseren zonder het equivalent van de berekening van de elementen van de Hessiaan $G(x^{(k+1)}) = A$. Wel kan eenvoudig worden bereikt dat de nieuwe benadering $H^* = H^{(k+1)}$ de eigenschap (2.7.6) bezit m.b.t. de verschilvector van de gradiënten in begin- en eindpunt van de laatste stap, d.i.

$$H^{(k+1)}y^{(k)} = H^{(k+1)}[g^{(k+1)} - g^{(k)}] = x^{(k+1)} - x^{(k)} = s^{(k)}$$

ofwel (2.7.7)

$$H^*y = s$$

Aan deze laatste relatie, die bekend staat als de quasi-Newton-relatie voldoen nagenoeg alle aanpassings- (of updating) formules voor quasi-Newton algorithmen. De quasi-Newton relatie (2.7.7) is dan ook het uitgangspunt bij uitstek voor de hierna volgende discussies.

2.7.5. Onder zekere voorwaarden kan met wat meer moeite ook worden bereikt dat de benadering $H^{(k+1)}$ van de inverse Hessiaan $[G(x^{(k+1)})]^{-1}$ zich bij toepassing bij de minimalisering van een positief definiete kwadratische vorm met Hessiaan A gedraagt als de echte inverse Hessiaan A^{-1} m.b.t. de gradiënt-verschil vectoren $y^{(j)}$ in elke voorgaande stappen van het iteratieproces, d.w.z. dat $H^{(k+1)}$ voldoet aan

$$H^{(k+1)}y^{(j)} = s^{(j)} \qquad j = 0, \ldots, k-1 \leq n-1 \qquad (2.7.8)$$

Deze relatie (of eigenschap) wordt in de literatuur (b.v.: [2.7.1]p. 95) wel aangeduid met de naam erfelijkheidsrelatie.

2.7.6. Voldoet een aanpassingsformule van een quasi-Newton algorithme bij toepassing op de minimalisering van een positief definiete kwadratische vorm aan zowel de quasi-Newton relatie (2.7.7) als aan de erfelijkheidsrelatie (2.7.8) dan zal na n stappen met lineair onafhankelijke vectoren $y^{(j)}$, $j = 0, n-1$, gelden dat

$$H^{(n)} = A^{-1} \tag{2.7.9}$$

Het minimum van de kwadratische vorm kan daarna in de (n + 1)-de stap direct worden gevonden met de Newton-stap

$$x^{(n+1)} := x^{(n)} - H^{(n)} g^{(n)} \tag{2.7.10}$$

Quasi-Newton algorithmen die bij de minimalisering van positief-definiete kwadratische vormen voldoen aan de drie genoemde voorwaarden:

(i) $H^{(k+1)} y^{(k)} = s^{(k)}$

(ii) $H^{(k+1)} y^{(j)} = s^{(j)} \qquad j = 0, \ldots, k-1$

(iii) $y^{(0)}, y^{(1)}, y^{(2)} \ldots$ lineair onafhankelijk

convergeren in ten hoogste (n + 1) stappen naar het minimum van de kwadratische vorm. Van deze algorithmen wordt gezegd dat zij "n-stappen-convergentie" of wel de "Qn-eigenschap" (vgl. pt 2.6.7) bezitten.

Standaardvorm van aanpassingsformules

2.7.7. Aan de in pt 2.7.4. besproken quasi-Newton-relatie zal worden voldaan indien voor de correctie matrix C in (2.7.3) een matrix gekozen wordt die voldoet aan de relatie

$$Cy = s - Hy \tag{2.7.11}$$

Hieraan wordt voldaan indien gebruik gemaakt wordt van een van de twee (rang 1 en rang 2) algemene aanpassingsformules gesuggereerd door Broyden [2.7.2]:

$$H^* := H + C = H + \frac{(s - Hy)z^T}{z^T y} \tag{2.7.12}$$

en

$$H^* := H + C = H + \frac{sv^T}{v^T y} - \frac{Hyw^T}{w^T y} \tag{2.7.13}$$

waarin z, v en w in principe willekeurige vectoren kunnen zijn die alleen

dienen te voldoen aan de voorwaarden

$$z^T y \neq 0,\ v^T y \neq 0 \text{ en } w^T y \neq 0 \qquad (2.7.14)$$

Door verschillende keuzen van de vectoren z, v en w kunnen een oneindig aantal variabele-metriek aanpassingsformules worden gegenereerd.

2.7.8. Op grond van de overweging dat matrix H^* een benadering is voor een symmetrische matrix (n.l. de inverse van de Hessiaan) liggen de keuzen

$$z = s - Hy \quad , \quad v = s \quad , \quad w = Hy$$

min of meer voor de hand. Hiermee resulteert de rang-1- aanpassingsformule

$$H^* := H + \frac{(s - Hy)(s - Hy)^T}{(s - Hy)^T y} \qquad (2.7.15)$$

welke bekend staat als "de"(symmetrische)<u>rang-1-aanpassingsformule</u> en de rang-2-aanpassingsformule

$$H^* := H + \frac{ss^T}{s^T y} - \frac{Hyy^T H}{y^T Hy} \qquad (2.7.16)$$

welke bekend staat als de <u>aanpassingsformule van Davidon - Fletcher - Powell</u> (ofwel de DFP-aanpassingsformule [2.7.7]). Deze beide vervullen een centrale rol in de theorie van de quasi-Newton-methoden. Bijzondere eigenschappen ervan worden besproken in § 2.8.

<u>Familie van aanpassingsformules van Huang</u>

2.7.9. Teneinde enige samenhang te brengen in de vele mogelijke aanpassingsformules construeerde Huang [2.7.9] een 3-parameter-familie van aanpassingsformules, waartoe vrijwel alle bekende aanpassingsformules behoren en waarvoor hij een aantal algemene eigenschappen kon bewijzen. Als uitgangspunt voor deze algemene klasse van aanpassingsformules koos Huang een algemenere vorm van de quasi-Newton-relatie dan (2.7.7), namelijk

$$H^* y = \rho s \qquad (2.7.17)$$

waarin ρ een willekeurig niet-negatief getal is (ρ mag dus ook de waarde 0 hebben). Tevens liet hij de mogelijkheid open voor niet-symmetrische aanpassingsformules, in verband waarmee hij de te genereren zoekrichting definieerde met behulp van de uitdrukking (vgl. (2.7.2))

$$d := - H^T g \tag{2.7.18}$$

Als algemene vorm van de aanpassingsformules van deze familie van Huang werd gekozen

$$H^* := H + \rho \frac{s(c_1 s + c_2 H^T y)^T}{(c_1 s + c_2 H^T y)^T y} - \frac{Hy(k_1 s + k_2 H^T y)^T}{(k_1 s + k_2 H^T y)^T y} \tag{2.7.19}$$

met als parameters ρ, $c_1/c_2 = \alpha$ en $k_1/k_2 = \beta$, met als restrictie k_1 en k_2 niet beide tegelijk aan 0 mogen zijn.

Symmetrische aanpassingsformules

2.7.10. Wordt de extra beperking opgelegd dat de aanpassingsformules symmetrisch moeten zijn dan resulteert de 2-parameter familie van symmetrische aanpassingsformules gekarakteriseerd door de formule

$$H^* := H + \rho \frac{ss^T}{s^T y} - \frac{Hyy^T H}{y^T Hy} - \gamma \left[\left(\frac{y^T Hy}{s^T y} \right) ss^T - sy^T H - Hys^T + \left(\frac{s^T y}{y^T Hy} \right) Hyy^T H \right] \tag{2.7.20a}$$

of equivalent door de formules

$$H^* := H + \rho \frac{ss^T}{s^T y} - \frac{Hyy^T H}{y^T Hy} - \gamma (s^T y)\ (y^T Hy) \left[\left(\frac{s}{s^T y} - \frac{Hy}{y^T Hy} \right) \left(\frac{s}{s^T y} - \frac{Hy}{y^T Hy} \right)^T \right] \tag{2.7.20b}$$

of

$$H^* := H + (\rho - \gamma (y^T Hy)) \frac{ss^T}{s^T y} - (1 + \gamma (s^T y)) \frac{Hyy^T H}{y^T Hy} + \gamma [sy^T H + Hys^T] \tag{2.7.20c}$$

of, nog anders,

$$H^* := [I + \gamma sy^T]\, H[I + \gamma ys^T] + \rho \frac{ss^T}{s^T y} - \\ + (1 + \gamma(s^T y)) \left[\gamma(y^T Hy) \frac{ss^T}{s^T y} + \frac{Hyy^T H}{y^T Hy} \right] \qquad (2.7.20d)$$

Vult men in bovenstaande formules voor ρ de waarde $\rho = 1$ in, d.w.z. beperkt men zich tot die formules die voldoen aan de ongemodificeerde quasi-Newton-relatie (2.7.7) dan resulteert de in de literatuur als 1-parameter-familie van Broyden bekend staande familie van aanpassingsformules voor quasi-Newton methoden.

2.7.11. De drie meest bekende representanten van aanpassingsformules uit de familie van Broyden zijn de twee eerder genoemde formules (2.7.15) en (2.7.16) met resp.

$$\gamma = \frac{-1}{(s - Hy)^T y} \quad : \text{rang-1-aanpassingsformule} \qquad (2.7.21)$$

$$\gamma = 0 \quad : \text{Davidon-Fletcher-Powell(of DFP) formule} \qquad (2.7.22)$$

en een in 1970 door diverse auteurs (Broyden [2.7.3], Fletcher [2.7.6] , Goldfarb [2.7.8] en Shanno [2.7.13])gepropageerde aanpassingsformule waarvoor

$$\gamma = - \frac{1}{s^T y} \qquad (2.7.23)$$

welke in de literatuur bekend staat als de aanpassingsformule van Broyden-Fletcher-(Goldfarb)-Shanno of BFS-formule.

$$H^* = H + \left(1 + \frac{y^T Hy}{s^T y}\right) \frac{ss^T}{s^T y} - \left(\frac{1}{s^T y}\right) [sy^T H + Hys^T] \qquad (2.7.24)$$

Om later (pt 2.8.12) te bespreken redenen staat deze formule ook bekend als de zg. complementaire DFP-formule. Verdere eigenschappen ervan zullen worden besproken in § 2.8.

2.7.12. De BFS-formule (2.7.24) vervult samen met de DFP-formule (2.7.16) een centrale rol in de meer recente theorie van de quasi-Newton methoden. Wordt de BFS-formule in navolging van (2.7.20b) geschreven in de vorm

$$H^* := H + \frac{ss^T}{s^Ty} - \frac{Hyy^TH}{y^THy} + (y^THy)\left(\frac{s}{s^Ty} - \frac{Hy}{y^THy}\right)\left(\frac{s}{s^Ty} - \frac{Hy}{y^THy}\right)^T \tag{2.7.25}$$

dan illustreert deze formulering duidelijk de voor de theorie belangrijke observatie van Fletcher [2.7.6] dat de BFS en de DFP-formules onderling slechts verschillen in een matrix van rang 1. Anders gezegd, tussen de BFS en de DFP-formules geldt de relatie

$$H^*_{BFS} = H^*_{DFP} + vv^T \tag{2.7.26}$$

waar

$$v = (y^THy)^{\frac{1}{2}}\left(\frac{s}{s^Ty} - \frac{Hy}{y^THy}\right) \tag{2.7.27}$$

met de eigenschap

$$v^Ty = 0 \tag{2.7.28}$$

Met behulp van de BFS- en DFP-formules kan ook de formule voor de gehele 1-parameter familie van Broyden (2.7.20b) worden herschreven als

$$H^* = H^*_{DFP} - \gamma(s^Ty)\,(H^*_{BFS} - H^*_{DFP}) \tag{2.7.29}$$

Stelt men

$$\phi = -\gamma(s^Ty) \tag{2.7.30}$$

dan resulteert de formulering

$$H^* = (1 - \phi)\,H^*_{DFP} + \phi H^*_{BFS} \tag{2.7.31}$$

Deze formulering van de 1-parameter familie van Broyden staat in de literatuur bekend als de 1-parameter familie (van aanpassingsformules van quasi-Newton methoden) van Fletcher.

Andere voorbeelden van aanpassingsformules van Huang

2.7.13. Andere, minder vaak toegepaste, leden van de familie van aanpassingsformules van Huang (vgl. pt 2.7.9) zijn o.a. de niet-symmetrische aanpassingsformules van Pearson (zie [2.7.9])

$$H^* := H + \frac{(s - Hy)y^T H}{y^T Hy} \tag{2.7.32}$$

en van McCormick (zie [2.7.9])

$$H^* := H + \frac{(s - Hy)s^T}{s^T y} \tag{2.7.33}$$

Ook de in de paragraaf over geconjugeerde-richtingen methoden besproken gradiënt projectie methode van Pearson

$$H^* = H - \frac{Hyy^T H}{y^T Hy} = H\left[I - \frac{yy^T H}{y^T Hy}\right] \tag{2.7.34}$$

alsook de geconjugeerde-gradiënt-methode gebaseerd op de Gram-Schmidt A-orthogonalisatieprocedure (vgl. (2.6.51) en (2.6.52))

$$H^* := H - \frac{Hys^T}{s^T y} = H\left[I - \frac{ys^T}{s^T y}\right] \tag{2.7.35}$$

blijken te kunnen worden geformuleerd als leden van de familie(van aanpassingsformules) van Huang. Voor deze geconjugeerde-gradiënt-methoden (2.7.34) en (2.7.35) geldt dat

$$\rho = 0 \tag{2.7.36}$$

met als consequentie dat na n lineair onafhankelijke zoekrichtingen

$$H^{(n)} = 0 \tag{2.7.37}$$

Dit resultaat illustreert het verschil tussen geconjugeerde gradiënt- en quasi-Newton methoden.

Exacte lijnminimalisering, geconjugeerde richtingen en Qn-eigenschap

2.7.14. Voor alle aanpassingsformules uit de familie van Huang geldt op grond van hun constructie dat zij voldoen aan de gegeneraliseerde quasi-Newton relatie (2.7.17)

$$H^{*}y = \rho s$$

Daarnaast kan, opnieuw voor alle leden van de familie van Huang, worden aangetoond (zie pt 2.7.17) dat de corresponderende quasi-Newton algorithmen (bij toepassing op een positief definiete kwadratische vorm met Hessiaan A mits gebruik wordt gemaakt van exacte lijnminimalisering voor de stapgrootte bepaling en mits de opvolgende H-matrices steeds positief definiet zijn)onderling A-geconjugeerde richtingen genereren

$$d^{(k+1)T}Ad^{(j)} = 0 \qquad j = 0, \ldots, k \tag{2.7.38}$$

en dat, onder dezelfde voorwaarden, de successievelijke gegenereerde H-matrices voldoen aan de erfelijkheidsrelaties

$$H^{(k+1)}y^{(j)} = \rho s^{(j)} \qquad j = 0, \ldots, k \tag{2.7.39}$$

In alle (niet-pathologische) gevallen waarbij uitgaande van een positief definiete*) beginmatrix $H^{(0)}$ ook alle opvolgende $H^{(k)}$ matrices positief definiet zijn, zijn de gegenereerde zoekrichtingen $d^{(k)}$ op grond van hun A-geconjugeerd zijn tevens lineair onafhankelijk met als gevolg dat hetzelfde geldt voor de verschilvectoren $y^{(k)}$. Aan alle drie in pt. 2.7.6 genoemde voorwaarden voor n-stappen convergentie van quasi-Newton algorithmen wordt dan voldaan. Mits exacte lijnminimalisering wordt toegepast

*) In het geval van niet-symmetrische $H^{(j)}$matrices moet gelden dat $\frac{1}{2}[H^{(j)} + H^{(j)T}]$positief definiet is.

voor de bepaling van de stapgrootte en mits gestart wordt met een positief-definiete beginmatrix $H^{(0)}$ *) en alle successievelijk gegenereerde H-matrices positief definiet blijven geldt dus dat alle quasi-Newton algorithmen met een aanpassingsformule behorend tot de familie van Huang (2.7.19) n-stappen convergentie ofwel de Qn-eigenschap bezitten.

2.7.15. Opgemerkt moet worden dat indien bij de minimalisering van een positief definiete kwadratische vorm gebruik wordt gemaakt van A-geconjugeerde zoekrichtingen in combinatie met lijnminimalisering voor de stapgroottebepaling dat in dat geval op grond van de in pt. 2.6.6 besproken eigenschap van geconjugeerde richtingen methoden volgt dat het minimum wordt bereikt in n stappen. Dit impliceert dat in het punt $x^{(n)}$ zowel geldt dat $g^{(n)} = 0$ als dat $H^{(n)} = A^{-1}$. Daarmee volgt voor de (n + 1)-de stap

$$x^{(n+1)} := x^{(n)} - A^{-1}0 = x^{(n)} \qquad (2.7.40)$$

Het minimum wordt dus een stap eerder (in n i.p.v. (n + 1) stappen) bereikt dan voorspeld voor quasi-Newton methoden in pt. 2.7.14.

2.7.16. Een tweede opmerking naar aanleiding van het besprokene in pt. 2.7.14 geldt het feit dat de moeilijkst te hanteren voorwaarde voor n-stappen-convergentie uit pt. 2.7.6 nl. dat de opvolgende $y^{(j)}$'s lineair onafhankelijk moeten zijn in 2.7.14 vervangen werd door de voorwaarde dat alle opvolgende H-matrices positief definiet*) moeten zijn. Deze laatste voorwaarde speelt ook een essentiële rol bij beschouwingen over de convergentie van de toepassingen van quasi-Newton methoden voor de minimalisering van algemene functies van meer variabelen. Voor de afgeleide van de restrictie $h^{(k)}(\alpha)$ van de functie f(x) langs de zoekrichting in het k-de iteratiepunt

$$h^{(k)}(\alpha) = f(x^{(k)} + \alpha d^{(k)}) \qquad (2.7.41)$$

geldt immers (met weglating van de iteratie index k)

$$h'(0) = g^T d = -g^T H g \qquad (2.7.42)$$

---

*) zie voetnoot p. 102

In het geval dat H positief definiet is zal het steeds mogelijk zijn om, zolang het minimum niet is bereikt d.i. $g \neq 0$, een van nul verschillende stapgrootte factor α en daarmee een nieuw iteratiepunt $x^* := x + \alpha d$ te bepalen waar de functiewaarde lager is.

2.7.17. Het bewijs van de in pt. 2.7.14 genoemde eigenschappen (2.7.39) en (2.7.38) kan worden geleverd met volledige inductie : er geldt voor k = 0, ..,n - 1, dat

$$H^{(k+1)}{}_y(k) = \rho s(k)$$

en

$$d^{(k+1)T}Ad(k) = - g^{(k+1)T}H(k+1)Ad(k) = - \frac{1}{\alpha(k)} g^{(k+1)T}H(k+1)As(k)$$

$$= - \frac{1}{\alpha(k)} g^{(k+1)T}H(k+1)_y(k) = \frac{-\rho}{\alpha(k)} g^{(k+1)T}s(k) = 0$$

en voor j = 0, 1, ..., k - 1, in de veronderstelling dat (2.7.39) en (2.7.38) gelden voor $H^{(k)}$ en $d^{(k)}$, dat

$$H^{(k+1)}{}_y(j) = H^{(j+1)}{}_y(j) + \sum_{\ell=j+1}^{k} \frac{s^{(\ell)}(c_1 s^{(\ell)} + c_2 H^{(\ell)T}y(\ell))T_y(j)}{(c_1 s^{(\ell)} + c_2 H^{(\ell)T}y(\ell))T_y(\ell)}$$

$$- \sum_{\ell=j+1}^{k} \frac{H^{(\ell)}y(\ell)(k_1 s^{(\ell)} + k_2 H^{(\ell)T}y(\ell))T_y(j)}{(k_1 s^{(\ell)} + k_2 H^{(\ell)T}y(\ell))T_y(\ell)}$$

$$= \rho s(j) + \sum_{\ell=j+1}^{k} \frac{s^{(\ell)}[(c_1 s^{(\ell)T}As(j)) + (c_2 \rho s^{(\ell)T}As(j))]}{(c_1 s^{(\ell)} + c_2 H^{(\ell)T}y(\ell))T_y(\ell)}$$

$$- \sum_{\ell=j+1}^{k} \frac{H^{(\ell)}y(\ell)[(k_1 s^{(\ell)T}As(j)) + (k_2 \rho s^{(\ell)T}As(j))]}{(k_1 s^{(\ell)} + k_2 H^{(\ell)T}y(\ell))T_y(\ell)}$$

$$= \rho s(j)$$

en

$$d^{(k+1)T}Ad(j) = - \frac{1}{\alpha(j)} g^{(k+1)T}H(k+1)As(j) = - \frac{\rho}{\alpha(j)} g^{(k+1)T}s(j)$$

$$= - \frac{\rho}{\alpha(j)} (g^{(j+1)} + \sum_{\ell=j+1}^{k} As^{(\ell)})T_s(j) = 0$$

Equivalentie van zoekrichtingen

2.7.18. Behalve het hierboven in pt. 2.7.14 weergegeven resultaat dat alle quasi-Newton algorithmen met aanpassingsformules uit de familie van Huang (2.7.19) bij minimalisering van positief definiete kwadratische vormen in het geval van exacte lijnminimalisering A-geconjugeerde zoekrichtingen genereren en daarmee de Qn-eigenschap bezitten, bewees Huang in 1970 ook het volgende veel opmerkelijker resultaat.

Stelling 2.7.18 (Huang [2.7.9]). Alle quasi-Newton algorithmen met aanpassingsformules uit de familie (2.7.19) (van Huang) genereren bij de minimalisering van een positief definiete kwadratische vorm als gebruik wordt gemaakt van exacte lijnminimalisering voor de stapgrootte bepaling en gestart wordt in hetzelfde startpunt $x^{(0)}$ met dezelfde positief definiete start matrix $H^{(0)}$ (zie voetnoot p. 102) exact dezelfde zoekrichtingen en daarmee als gevolg van de eenduidigheid van de lijnminimalisering exact dezelfde rij van iteratiepunten $\{x^{(0)}, x^{(1)}, \ldots, x^{(k)}, \ldots\}$.

Bewijs : Het bewijs van deze stelling vergt nogal veel rekenwerk. Een aantal tussenresultaten daarvan zijn op zichzelf ook interessant en om die reden volgt hieronder een summiere opsomming van de hoofdpunten van het bewijs: Van veel nut voor het bewijs zijn de scheve-projectie relatie

$$[I - \frac{ac^T}{a^Tc}]\, a = 0 \qquad (2.7.43)$$

en de daarmee af te leiden relatie

$$[I - \frac{(a+b)c^T}{(a+b)^Tc}]\, b = \frac{a^Tc}{(a+b)^Tc}\,[I - \frac{ac^T}{a^Tc}]\, b \qquad (2.7.44)$$

Toepassing van deze laatste relatie op het zoekrichtings-generatieproces met een aanpassingsformule van de familie van Huang geeft het resultaat

$$\begin{aligned} H^{*T}g^* &= [I - \frac{(k_1 s + k_2 H^T y)y^T}{(k_1 s + k_2 H^T y)^T y}]\, H^T g^* \\ &= [I - \frac{((k_1 + k_2/\alpha)s + k_2 H^T g^*)y^T}{((k_1 + k_2/\alpha)s + k_2 H^T g^*)^T y}]\, H^T g^* \end{aligned} \qquad (2.7.45)$$

$$= \left(\frac{(k_1 + k_2/\alpha)s^T y}{(k_1 + k_2/\alpha)s^T y + k_2 g^{*T} Hy}\right)[I - \frac{sy^T}{s^T y}]H^T g^*$$

Een consequentie van het feit dat ook geldt

$$H^{*T} g^* = - \frac{s^*}{\alpha^*} \tag{2.7.46}$$

is dat

$$H^T g^* = - \frac{((k_1 + k_2/\alpha)s^T y + k_2 g^{*T} Hy)}{(k_1 + k_2/\alpha)s^T y} \frac{s^*}{\alpha^*} + \frac{y^T H^T g^*}{s^T y} s \tag{2.7.47}$$

en dat

$$H^T y = -\left(\frac{1}{\alpha^*} + \frac{k_2/\alpha^*}{(k_1 + k_2/\alpha)} \frac{y^T H^T g^*}{s^T y}\right)s^* + \left(\frac{1}{\alpha} + \frac{y^T H^T g^*}{s^T y}\right) s \tag{2.7.48}$$

Meer in het bijzonder volgt daaruit dat

$$k_1 s + k_2 H^T y = \left((k_1 + \frac{k_2}{\alpha}) + \frac{k_2 y^T H^T g^*}{s^T y}\right)s -$$

$$+ \left(\frac{k_2/\alpha^*}{k_1 + k_2/\alpha}\right)\left((k_1 + \frac{k_2}{\alpha}) + \frac{k_2 y^T H^T g^*}{s^T y}\right)s^*$$

en daarmee met $s^{*T} y = 0$ dat

$$\left[I - \frac{(k_1 s + k_2 H^T y)y^T}{(k_1 s + k_2 H^T y)^T y}\right] = \left[I - \frac{sy^T}{s^T y} + \left(\frac{k_2/\alpha^*}{k_1 + k_2/\alpha}\right) \frac{s^* y^T}{s^T y}\right] \tag{2.7.49}$$

Voor alle vectoren p waarvoor $s^T p = 0$ volgt hiermee dat voor alle aanpassingsformules waarvoor $(k_1 + \frac{k_2}{\alpha}) \neq 0$ geldt dat

$$H^{*T} p = \left[I - \frac{sy^T}{s^T y} + \left(\frac{k_2/\alpha^*}{k_1 + k_2/\alpha}\right) \frac{s^* y^T}{s^T y}\right] H^T p$$

Herhaalde toepassing van deze relatie in formule (2.7.45) resulteert in de uitdrukking

$$H^{*T}g^{*} = \left(\frac{(k_1 + k_2/\alpha)s^T y}{(k_1 + k_2/\alpha)s^T y + k_2 g^{*T} H y}\right) \prod_{j=0}^{*-1} \left[I - \frac{s^{(j)} y^{(j)T}}{s^{(j)T} y^{(j)}} \right] H^{(0)T} g^{*} \tag{2.7.50}$$

ofwel in verband met het A-geconjugeerd zijn van de richtingen $s^{(j)}$

$$H^{*T}g^{*} = \left(\frac{(k_1 + k_2/\alpha)s^T y}{(k_1 + k_2/\alpha)s^T y + k_2 g^{*T} H y}\right) \left[I - \sum_{j=0}^{*-1} \frac{s^{(j)} y^{(j)T}}{s^{(j)T} y^{(j)}} \right] H^{(0)T} g^{*}$$

$$= \beta^{*} q^{*} \tag{2.7.51}$$

waarin β^{*} een scalaire grootheid is en q^{*} een zoekrichting die onafhankelijk is van de keuze van de parameters ρ, c_1/c_2 en k_1/k_2, namelijk

$$q^{*} = \left[I - \sum_{j=0}^{*-1} \frac{s^{(j)} y^{(j)T}}{s^{(j)T} y^{(j)}} \right] H^{(0)T} g^{*} \tag{2.7.52}$$

Dit completeert het bewijs. □

2.7.19. Naar het voorbeeld van Huang lukte het Dixon in 1972 een nog verdergaand resultaat te bewijzen en wel het volgende

<u>Stelling 2.7.19</u> (Dixon [2.7.5]). Onder de voorwaarde dat voor de stapgroottebepaling gebruik wordt gemaakt van een <u>eenduidige exacte lijnminimalisering</u>, d.w.z. een eenduidig algorithme voor de bepaling van een stapgrootte α zodat

$$\nabla^{T} f(x + \alpha d) d = g^{*T} d = 0$$

geldt dat alle quasi-Newton algorithmen met aanpassingsformules behorend tot de familie van aanpassingsformules (2.7.19) van Huang <u>met dezelfde parameterwaarde ρ</u> bij toepassing voor de minimalisering van <u>willekeurige differentieerbare functies</u>, indien gestart in hetzelfde startpunt $x^{(0)}$ met dezelfde

zoekrichtingen en daarmee (als gevolg van de eenduidigheid van de stapgroottebepaling) ook exact dezelfde successievelijke iteratiepunten $\{x^{(0)}, x^{(1)}, \ldots\}$ genereren.

Bewijs: Voor het nogal wat rekenwerk vergende bewijs moet worden verwezen naar [2.7.4] en [2.7.5].

Convergentie van quasi-Newton algorithmen

2.7.20. De hierboven weergegeven uitspraken over de equivalentie van de zoekrichtingen in quasi-Newton algorithmen impliceren dat in het geval gebruik wordt gemaakt van eenduidige exacte lijnminimalisering het er niet toe doet welke van de vele mogelijke aanpassingsformules wordt gebruikt in een algorithme. Dit heeft in de eerste plaats de interessante consequentie dat als bepaalde theoretische eigenschappen, zoals convergentieeigenschappen, kunnen worden aangetoond voor één aanpassingsformule, dezelfde eigenschappen direct ook gelden voor alle quasi-Newton algorithmen met aanpassingsformules uit dezelfde groep. Anderzijds volgt uit de genoemde uitspraken ook dat verschillen in het gedrag van de diverse algorithmen uitsluitend het gevolg zijn de numerieke onnauwkeurigheden (vooral die bij het lijnminimaliseringsproces) bij de practische implementatie van de algorithmen. Sommige aanpassingsformules blijken daarvoor gevoeliger dan andere.

2.7.21. De genoemde twee aspecten van de theoretische equivalentie van de zoekrichtingen van alle quasi-Newton algorithmen met aanpassingsformules uit de familie van Huang hebben er toe geleid dat er twee groepen van convergentie uitspraken over de toepassingen van quasi-Newton methoden voor de minimalisering van algemene differentieerbare of convexe functies kunnen worden onderscheiden in de literatuur. De eerste groep daarvan geldt voor die algorithmen die gebruik maken van eenduidige exacte lijnminimalisering (vgl. pt. 2.7.19) voor de stapgroottebepaling, de z.g. perfecte algorithmen. De andere groep betreft die algorithmen die geen of slechts bij benadering lijnminimalisering gebruiken voor de stapgroottebepaling. Convergentieuitspraken voor deze laatste categorie van algorithmen, de z.g. imperfecte algorithmen zijn veelal slechts geldig voor een zeer beperkte groep van algorithmen.

2.7.22. Een van de bekendste convergentieuitspraken voor de perfecte algorithmen is de in 1971 gepubliceerde convergentiestelling van Powell [2.7.10] voor de DFP-algorithme (vgl. ook [2.7.4]). Met een bijzonder intrigerend, bijna virtuoos bewijs was Powell in staat aan te tonen dat, in het geval de DFP-algorithme met exacte lijnminimalisering wordt toegepast voor de minimalisering van een twee maal differentieerbare convexe functie f(x) waarvoor de verzameling

$$S(x^{(0)}) = \{x \in \mathbb{R}^n \mid f(x) \leq f(x^{(0)})\}$$

gesloten en begrensd is en waarvoor geldt dat de norm van de Hessiaan begrensd is op deze verzameling, de door de algorithme gegenereerde rij $x^{(0)}, x^{(1)}, \ldots$ convergeert naar een stationair punt van f. Voor het geval van convergentie naar een lokaal minimum x_{min} van f waar de Hessiaan $G(x_{min})$ positief definiet is en onder de extra voorwaarde dat de Hessiaan voldoet aan een Lipschitz conditie

$$\|G(x) - G(x_{min})\| \leq L\|x - x_{min}\|$$

voor alle x in de verzameling $S(x^{(0)})$, bewees hij verder dat de convergentiesnelheid van de perfecte DFP-algorithme superlineair is.(Schuller en Stoer [2.7.12] lieten later zien dat de orde van convergentie onder nagenoeg dezelfde voorwaarden gelijk was aan $^n/_2$). Op grond van de equivalentie uitspraak van Dixon gelden deze convergentieeigenschappen ook voor alle perfecte quasi-Newton methoden met aanpassingsformules uit de familie van Huang (2.7.19) met de parameter ρ gelijk aan $\rho = 1$.

2.7.23. Over de convergentie van de imperfecte quasi-Newton algorithmen verschijnen er in de tegenwoordige optimaliserings-literatuur regelmatig nieuwe uitspraken van gevarieerde aard. Het onderzoek op dit gebied is nog volop in beweging. In de meeste gevallen is het resultaat dat onder relatief weinig beperkende voorwaarden superlineaire convergentie kan worden aangetoond. In sommige gevallen kan n - of (n + 1)-staps kwadratische convergentie (vgl. pt. 2.6.26) worden aangetoond.

Literatuur

2.7.24. Meer informatie en details over het besprokene in deze paragraaf kunnen worden gevonden in de volgende publicaties:

[2.7.1]: Zie [1.1.7] Adby en Dempster (1974).

[2.7.2]: Broyden, C.G.: Quasi-Newton methods and their application to function minimization. Math. Comp. 21 (1967),pp. 368-381.

[2.7.3]: Broyden, C.G.: The convergence of a class of doublerank minimization algorithms, Part I: General considerations, and Part II: The new algorithm, J. Inst. Math. & Appl., 6 (1970) pp. 76-90, 222-231.

[2.7.4]: Dixon, L.C.W.: Variable metric algorithms: Necessary and sufficient conditions for identical behavior of non-quadratic functions, J. Opt. Theory & Appl. 10 (1972),pp. 34-40.

[2.7.5]: Dixon, L.C.W.: Quasi-Newton algorithms generate identical points, Part I and Part II: The proofs of four new theorems, Math. Prop. 2 (1972) pp. 383-387 en 3 (1972) pp. 345-358.

[2.7.6]: Fletcher, R. : A new approach to variable metric algorithms, Computer J., 13 (1970) pp. 317-322.

[2.7.7]: Fletcher, R. and Powell, M.J.D. : A rapidly convergent descent method for minimization, Computer J. 6, pp. 163-168.

[2.7.8]: Goldfarb, D. : A family of variable metric methods derived by variational means, Math. Comp. 24 (1970) pp. 23-26.

[2.7.9]: Huang, H.Y. : Unified approach to quadratically convergent algorithms for function minimization. J. Opt. Theory & Appl. 6 (1970) pp. 269-282.

[2.7.10]: Powell, M.J.D. : On the convergence of the variable metric algorithm. J. Inst. Math. & Appl. 7 (1971) pp. 21-36.

[2.7.11]: Powell, M.J.D. : Some properties of the variable metric algorithm, in "Numerical methods for nonlinear optimization". F.A. Lootsma, (Ed) Academic Press, London,(1972) pp. 1-17.

[2.7.12]: Schuller, G und Stoer, J. : Ueber die Konvergenzordnung gewisser Rang-2-Verfahren zur Minimierung von Funktionen in Numerische Methoden bei Optimierungsaufgaben, Band 2, L. Collatz, W. Wetterling (Eds) Birkhauser Verlag, Basel, 1974, pp. 125-147.

[2.7.13] : Shanno, D.F.: Conditioning of quasi-Newton methods for function minimization, Math. Comp. 24 (1970)pp. 647-656.

§ 2.8. Quasi-Newton-methoden II: Speciale aanpassingsformules

I Rang-één formule

2.8.1. Een van de interessantste leden van de familie van aanpassingsformules voor quasi-Newton algorithmen van Huang (pt. 2.7.9) is de symmetrische rang-één-formule (2.7.15)

$$H^* := H + \frac{(s-Hy)(s-Hy)^T}{(s-Hy)^T y} \qquad (2.8.1)$$

Deze formule heeft de speciale voor de practijk zeer belangrijke eigenschap dat de bij de minimalisering van een positief definiete kwadratische vorm successievelijk gegenereerde H-matrices ook zonder lijnminimalisering voldoen aan de erfelijkheidsrelatie (2.7.8)

$$H^{(k+1)}y^{(j)} = s^{(j)} \quad j = 0, \ldots, k-1$$

Deze eigenschap kan worden aangetoond (vgl [2.8.1]) met volledige inductie: Als geldt dat $H^{(k)}y^{(j)} = s^{(j)}$ voor $j = 0, \ldots, k-1$ dan geldt ook dat

$$H^{(k+1)}y^{(j)} = H^{(k)}y^{(j)} + \frac{(s^{(k)} - H^{(k)}y^{(k)})\,(s^{(k)T}y^{(j)} - y^{(k)T}H^{(k)}y^{(j)})}{(s^{(k)T}y^{(k)} - y^{(k)T}H^{(k)}y^{(k)})}$$

$$= H^{(k)}y^{(j)} = s^{(j)} \quad j = 0, \ldots, k-1$$

omdat

$$s^{(k)T}y^{(j)} - y^{(k)T}H^{(k)}y^{(j)} = s^{(k)T}y^{(j)} - y^{(k)T}s^{(j)} = 0$$

2.8.2. Een direct gevolg van de eigenschap in pt. 2.8.1 is dat iedere quasi-Newton algorithme die gebruik maakt van de rang-één-formule onder de voorwaarde (vgl pt. 2.7.6) dat de successievelijk gegenereerde zoekrichtingen lineair onafhankelijk zijn zonder dat gebruik wordt gemaakt van lijnminimalisering n-staps-convergentie of wel de Qn-eigenschap bezit. Dit resultaat is voor de practijk van veel belang omdat de functie- (en evt. gradiënt-) evaluaties noodzakelijk voor al dan niet nauwkeurige lijnminimalisering bij toepassingen van quasi-Newton algorithmen veelal het grootste deel van het rekenwerk be-

tekenen. De vrijheid in de stapgrootte bepaling opent ook mogelijkheden voor de toepassing van deze quasi-Newton algorithmen bij minimaliserings problemen met beperkingen.

2.8.3. De rang-één-aanpassingsformule heeft naast de in het voorgaande punt genoemde positieve eigenschap ook een aantal negatieve eigenschappen die de balans vaak in de tegenovergestelde kant doen doorslaan. In het bijzonder geldt namelijk dat successievelijk gegenereerde H-matrices, ook bij de minimalisering van positief definiete kwadratische vormen niet noodzakelijk positief definiet blijven. Een berucht voorbeeld hiervan levert de situatie waarbij tijdens het iteratieproces de voor de hand liggende stapgrootte $\alpha = 1$

$$s = x^* - x = -Hg$$

juist het lijnminimum in de richting s oplevert, d.w.z.

$$g^{*T}s = 0$$

In dat geval geldt namelijk dat

$$H^*g^* = Hg^* + \frac{(s - Hy)(s - Hy)^T}{(s - Hy)^T y} g^*$$

$$= Hg^* + \frac{Hg^* g^{*T} Hg^*}{-g^{*T} Hg^*} = 0$$

De quasi-Newton algorithme stopt dan in het punt x^* waar $g^* \neq 0$.

2.8.4. Naast de kans op singulier worden van de H-matrix tijdens het iteratieproces bestaat ook de kans op onbegrensde correcties, nl. in het geval dat

$$(s - Hy)^T y \to 0$$

In het geval van toepassing op een positief definiete kwadratische vorm kan deze laatste voorwaarde worden herschreven als

$$y^T(A^{-1} - H)y \to 0$$

welke uitdrukking de moeilijkheid duidelijk illustreert. Het idee van de quasi-Newton algorithmen is immers om H-matrix een zo goed mogelijke benadering te laten zijn voor de inverse van Hessiaan A. Des te beter de benadering des te groter de kans op onbegrensde correcties.

2.8.5. Als gevolg van de in beide voorgaande punten gesignaleerde moeilijkheden wordt de symmetrische rang-één formule in de practijk niet op grote schaal gebruikt in quasi-Newton algorithmen voor minimaliseringsproblemen zonder beperkingen.Uit theoretisch oogpunt en in het geval van minimaliseringsproblemen met beperkingen blijft de rang-één aanpassingsformule van veel belang.

II. DFP-formule

2.8.6. Veruit het bekendste lid van de familie van aanpassingsformules van Huang is de DFP (of Davidon-Fletcher-Powell) aanpassingsformule (2.7.16)

$$H^* := H + \frac{ss^T}{s^Ty} - \frac{Hyy^TH}{y^THy} \qquad (2.8.2)$$

die ook wel bekend is als "de" variabele-metriek-formule. Ondanks het feit dat de DFP-formule op dit moment niet langer wordt beschouwd als de "beste" aanpassingsformule heeft de formule nog niets aan belang ingeboet voor de theorie als gevolg van de centrale plaats die zij inneemt in de recente ontwikkelingen van de quasi-Newton methoden. Dit is o.a. een gevolg van de omstandigheid dat de formule een aantal voor quasi-Newton algorithmen belangrijke eigenschappen heeft die deels ook gelden voor de andere aanpassingsformules en die voor de DFP-formule eenvoudig aan te tonen zijn.

2.8.7. Een van de belangrijkste eigenschappen van de DFP-formule is weergegeven in de volgende uitspraak

Lemma 2.8.7. Voor de H-matrices gegenereerd met de DFP-aanpassingsformule geldt dat indien (1) H positief definiet is en (2) $s^Ty > 0$ dan is ook H^* positief definiet.

Bewijs (vgl. [2.8.1]): Voor een willekeurige vector $v \in \mathbb{R}^n$ geldt dat

$$v^T H^* v = v^T H v + \frac{(v^T s)^2}{s^T y} - \frac{(v^T H y)^2}{y^T H y}$$

welke uitdrukking met $a = H^{\frac{1}{2}} v$ en $b = H^{\frac{1}{2}} y$ kan worden herschreven als

$$v^T H^* v = \frac{(a^T a)(b^T b) - (a^T b)^2}{(b^T b)} + \frac{(v^T s)^2}{s^T y}$$

Met de ongelijkheid van Cauchy-Schwartz volgt direct dat of

$$v^T H^* v > \frac{(v^T s)^2}{s^T y} \geq 0 \qquad \text{als } a \neq \beta b$$

of

$$v^T H^* v = \frac{(v^T s)^2}{s^T y} \qquad \text{als } a = \beta b$$

In het laatste geval geldt $v = \beta y$ waarmee

$$v^T H^* v = \beta^2 s^T y > 0$$

In beide gevallen geldt derhalve

$$v^T H^* v > 0$$

waarmee het bewijs van het positief definiet zijn van H^* is geleverd. □

2.8.8. Met behulp van dit lemma volgt onmiddellijk de volgende uitspraak.

Stelling 2.8.8 : Wordt de DFP-formule (2.8.2) gebruikt in een quasi-Newton algorithme voor de minimalisering van een tweemaal differentieerbare functie dan geldt dat de successievelijk gegenereerde H-matrices positief definiet zijn als voldaan wordt aan de voorwaarden :

1) de start matrix $H^{(0)}$ is positief definiet en
2) in iedere iteratieslag geldt

$$s^T y > 0 \tag{2.8.3}$$

Aan deze laatste voorwaarde wordt steeds voldaan indien
a) de Hessiaan van de te minimaliseren functie positief definiet is (of, equivalent, de functie strikt convex is) of
b) gebruik wordt gemaakt van exacte lijn-minimalisering en $\alpha > 0$.

Bewijs: Het eerste deel van deze uitspraak volgt onmiddellijk met volledige inductie en Lemma 2.8.7. Als de te minimaliseren functie tweemaal differentieerbaar en strikt convex is dan geldt met de middelwaarde stelling

$$s^T y = s^T(g^* - g) = s^T(G(x + \hat{\alpha}s))s > 0 \quad 0 < \hat{\alpha} < 1$$

Analoog als gebruik gemaakt wordt van lijnminimalisering en $\alpha > 0$ geldt

$$s^T y = s^T(g^* - g) = -s^T g = \alpha g^T H g > 0 \tag{2.8.4}$$

Dit completeert het bewijs. □

2.8.9. Een in verband met de hierboven gegeven stelling interessante relatie wordt gegeven in de volgende uitspraak (vgl. [2.8.3]).

Stelling 2.8.9 : Bij toepassing van de quasi-Newton algorithme met de DFP-aanpassingsformule (2.8.2) voor de minimalisering van een tweemaal differentieerbare convexe functie geldt indien gebruik gemaakt wordt van exacte lijn-minimalisering voor de stapgrootte bepaling en indien gestart wordt met een positief definiete beginmatrix $H^{(0)}$ dat

$$g^{*T}H^*g^* = g^{*T}Hg^*\left(\frac{g^T Hg}{g^{*T}Hg^* + g^T Hg}\right) \tag{2.8.5a}$$

of equivalent dat (als $g^* \neq 0$)

$$\frac{1}{g^{*T}H^*g^*} = \frac{1}{g^{*T}Hg^*} + \frac{1}{g^T Hg} \tag{2.8.5b}$$

Bewijs : Gebruik van de DFP-aanpassingsformule en exacte lijnminimalisering impliceert dat $g^{*T}Hg = 0$ waarmee

$$g^{*T}H^*g^* = g^{*T}Hg^* - \frac{g^*Hyy^THg^*}{y^THy}$$

$$= g^{*T}Hg^* - \frac{(g^{*T}Hg^*)^2}{g^{*T}Hg^* + g^THg}$$

$$= \frac{(g^{*T}Hg^*)(g^THg)}{g^{*T}Hg^* + g^THg}$$

Inversie van deze relatie geeft het tweede resultaat. □

Uit stelling 2.8.9 volgt dat onder de genoemde voorwaarden de rij

$$\{g^{(k)T}H^{(k)}g^{(k)}\}_{k=0}^{\infty}$$

monotoon daalt. Dit impliceert met

$$g^THg = -g^Td \tag{2.8.6}$$

dat ook de afgeleide van de functie in de zoekrichting (vgl. pt. 2.2.1) in absolute waarde eveneens monotoon daalt.

2.8.10. Een ander interessant aspect van de DFP-formule is dat de H-matrix in het geval van de toepassing binnen een perfect quasi-Newton algorithme voor de minimalisering van een positief definiete kwadratische vorm geschreven kan worden als de som van twee matrices met eigen aanpassingsformules, waarvan de een zorg draagt voor het genereren van A-orthogonale zoekrichtingen en de ander zorgt voor de opbouw van de inverse matrix A^{-1}. D.w.z. de DFP-formule kan worden geschreven in de vorm

$$H^* := P^* + C^* \tag{2.8.7}$$

met

$$P^* := P - \frac{Pyy^TP}{y^TPy} \tag{2.8.8}$$

en

$$C^* := C + \frac{ss^T}{s^T y} \tag{2.8.9}$$

met respectievelijk

$$P^{(0)} := H^{(0)} \qquad P^{(n)} := 0 \tag{2.8.10}$$

en

$$C^{(0)} := 0 \qquad C^{(n)} := A^{-1}$$

Vergelijking van de aanpassingsformule voor de matrix P met de aanpassingsformule voor de matrix H volgens de gradiënt-projectiemethode van Pearson (verg. (2.6.39))

$$H^* := H - \frac{Hyy^TH}{y^THy} = H\frac{(I - yy^TH)}{y^THy} \tag{2.8.11}$$

leert dat deze matrices juist aan elkaar gelijk zijn. In het beschouwde speciale geval, kwadratische vorm en exacte lijnminimalisering speelt de matrix C geen enkele rol bij de bepaling van de nieuwe zoekrichting. In verband met het "expanding subspaces" theorema (Stelling 2.6.6) voor geconjugeerde-richtingen-methoden geldt immers dat zowel dat

$$Cg^* = \left(\sum_{j=0}^{k-1} \frac{s^{(j)}s^{(j)T}}{s^{(j)T}y^{(j)}}\right)g^{(k+1)} = 0 \tag{2.8.12}$$

als

$$Cy = \left(\sum_{j=0}^{k-1} \frac{s^{(j)}s^{(j)T}}{s^{(j)T}y^{(j)}}\right)(g^{(k+1)} - g^{(k)}) = 0 \tag{2.8.13}$$

De equivalentie van de gegenereerde zoekrichtingen (vgl. pt. 2.7.18) is in dit speciale geval op een illustratieve manier aangetoond.

III. BFS-formule

2.8.11. De derde aanpassingsformule uit de familie van Huang die een centrale plaats inneemt in de theorie van de quasi-Newton methoden is zg. BFS- of Broyden-

Fletcher-Shanno-formule (2.7.24)

$$H^* := H + \left(1 + \frac{y^T H y}{s^T y}\right)\frac{ss^T}{s^T y} - \left(\frac{1}{s^T y}\right)[sy^T H + Hys^T] \qquad (2.8.14)$$

of (2.7.25)

$$H^* := H + \frac{ss^T}{s^T y} - \frac{Hyy^T H}{y^T Hy} + (y^T Hy)\left(\frac{s}{s^T y} - \frac{Hy}{y^T Hy}\right)\left(\frac{s}{s^T y} - \frac{Hy}{y^T Hy}\right)^T \qquad (2.8.15)$$

2.8.12. De BFS-formule staat ook bekend als de <u>complementaire DFP-formule</u> en wel i.v.m. de in de volgende stellingen vervatte relaties (vgl. [2.8.5]).

<u>Stelling 2.8.12</u> : Voor de inverse matrix B van de H-matrix gegenereerd in een quasi-Newton algorithme met behulp van de DFP-aanpassingsformule geldt de equivalente aanpassingsformule

$$B^* = B + \left(1 + \frac{s^T Bs}{y^T s}\right)\frac{yy^T}{y^T s} - \frac{1}{y^T s}[ys^T B + Bsy^T] \qquad (2.8.16)$$

waar

$$B^* = (H^*)^{-1} \qquad \text{en} \qquad B = H^{-1}$$

<u>Bewijs</u> : Het bewijs van de juistheid van deze uitdrukking volgt uit twee successievelijke toepassingen van de modificatieregel van Householder ([2.8.2], p. 134) voor de inverse van een matrix gecorrigeerd met een rang-één-correctie.

$$\left(A + \frac{1}{\alpha} uv^T\right)^{-1} = A^{-1} - \frac{A^{-1}uv^T A^{-1}}{\alpha + v^T A^{-1} u} \qquad (2.8.17)$$

op achtereenvolgens

$$\bar{H} = H + \frac{ss^T}{s^T y} \qquad (2.8.18)$$

en

$$H^* = \bar{H} - \frac{Hyy^T H}{y^T Hy} \qquad (2.8.19)$$

In het eerste geval resulteert

$$\bar{\bar{H}}^{-1} = H^{-1} - \frac{H^{-1}ss^TH^{-1}}{s^Ty + s^TH^{-1}s} \tag{2.8.20}$$

en in het tweede geval

$$(H^*)^{-1} = \bar{\bar{H}}^{-1} + \frac{\bar{\bar{H}}^{-1}Hyy^TH\bar{\bar{H}}^{-1}}{y^THy - y^TH\bar{\bar{H}}^{-1}Hy} \tag{2.8.21}$$

Substitutie van (2.8.20) in (2.8.21) geeft het gewenste resultaat in termen van H^{-1} en $(H^*)^{-1}$

$$(H^*)^{-1} = H^{-1} + \left(1 + \frac{s^TH^{-1}s}{s^Ty}\right)\frac{yy^T}{s^Ty} - \frac{1}{s^Ty}\left[H^{-1}sy^T + ys^TH^{-1}\right]$$

Herschrijven van deze uitdrukking met B^* en B completeert het bewijs. □

2.8.13. Door vervanging van de matrices B^* en B en de vectoren y en s door resp. de matrices H^* en H en de vectoren s en y in overeenstemming met de aan de quasi-Newton relatie equivalente betrekking

$$B^*s = y \tag{2.8.22}$$

volgt onmiddellijk de aan Stelling 2.8.12 analoge uitspraak.

<u>Stelling 2.8.13</u> : De inverse matrix H van de benaderingsmatrix B van de Hessiaan gegenereerd met de aan de DFP-formule verwante aanpassingsformule

$$B^* := B + \frac{yy^T}{y^Ts} - \frac{Bss^TB}{s^TBs} \tag{2.8.23}$$

wordt gegeven door de BFS-aanpassingsformule

$$H^* := H + \left(1 + \frac{y^THy}{s^Ty}\right)\frac{ss^T}{s^Ty} - \frac{1}{s^Ty}\left[sy^TH + Hys^T\right]$$

waar (2.8.24)

$$H^* = (B^*)^{-1} \qquad \text{en} \qquad H = B^{-1}$$

Bewijs : Met de eerder genoemde vervanging van de matrices B en B^* en de vectoren s en y door resp. de matrices H en H^* en de vectoren y en s verloopt het bewijs geheel analoog aan het bewijs van stelling 2.8.12. □

2.8.14. De relatie die bestaat tussen de DFP- en de BFS-formule wordt in de literatuur vaak aangeduid als een dualiteitsrelatie. Men zegt dan dat de BFS-formule de duale is van de DFP-formule en vice versa op grond van de duale omzettingen

$$\begin{aligned} B &\leftrightarrow H \\ y &\leftrightarrow s \end{aligned} \qquad (2.8.25)$$

Aangetoond kan worden dat met dezelfde omzettingsregels geldt dat de eerder besproken symmetrische rang-één-formule de duale is van zichzelf, d.w.z. voor de inverse $B = H^{-1}$ van de matrix H genereert met de aanpassingsformule (2.8.1)

$$H^* := H + \frac{(s-Hy)\,(s-Hy)^T}{(s-Hy)^T y}$$

geldt de aanpassingsformule

$$B^* := B + \frac{(y-Bs)(y-Bs)^T}{(y-Bs)^T s} \qquad (2.8.26)$$

2.8.15. De gelijkvormigheid van de aanpassingsformule (2.8.16) van de inverse B van de met BFS-formule gegenereerde matrix H maakt het mogelijk t.a.v. het positief definiet blijven van de matrix H gegenereerd met de BFS-formule dezelfde uitspraak te doen als voor de matrix H gegenereerd met de DFP-formule in Stelling 2.8.16.

Stelling 2.8.15 : Wordt de BFS-formule (2.8.14) gebruikt als aanpassingsformule in een quasi-Newton algorithme voor de minimalisering van een twee maal differentieerbare functie dan geldt dat de successievelijk gegenereerde H-matrices positief definiet zijn als voldaan wordt aan de voorwaarden
1) de startmatrix $H^{(0)}$ is positief definiet
2) in iedere iteratieslag geldt

$$s^T y > 0$$

Aan deze laatste voorwaarde wordt steeds voldaan indien
a) de Hessiaan van de te minimaliseren functie positief definiet is (of, equivalent, de functie strikt convex) of
b) gebruik wordt gemaakt van exacte lijnminimalisering en $\alpha > 0$.

Bewijs : Als $H^{(0)}$ positief definiet en $s^T y > 0$ in iedere iteratieslag dan volgt uit Stelling 2.8.13 en Stelling 2.8.8 dat de inverse matrix B van de met de BFS-formule gegenereerde matrix positief definiet blijft in alle opvolgende iteratieslagen. Hetzelfde geldt dan voor de H-matrix zelf. □

2.8.16. In het geval van het gebruik van de BFS-formule in een perfecte quasi-Newton algorithme, dan gelden een aantal interessante relaties die aanleiding geven tot een soortgelijke uitspraak als gedaan voor de DFP-formule in Stelling 2.8.9.

Stelling 2.8.16 : Bij gebruik van de BFS-aanpassingsformule (2.8.15) in een perfecte quasi-Newton algorithme (d.i. met lijnminimalisering voor de stapgrootte bepaling) voor de minimalisering van een tweemaal differentieerbare functie, kan onder de voorwaarden dat gestart wordt met een positief definiete beginmatrix $H^{(0)}$ en dat in iedere iteratieslag $\alpha > 0$ voor de zoekrichting worden geschreven

$$d^* = d_I^* + d_{II}^* \tag{2.8.27a}$$

waar

$$d_I^* = -\left[H + \frac{ss^T}{s^T y} - \frac{Hyy^T H}{y^T Hy}\right] g^* = (g^T Hg)\left(\frac{s}{s^T y} - \frac{Hy}{y^T Hy}\right) \tag{2.8.27b}$$

$$d_{II}^* = -(y^T Hy)\left(\frac{s}{s^T y} - \frac{Hy}{y^T Hy}\right)\left(\frac{s}{s^T y} - \frac{Hy}{y^T Hy}\right)^T g^* = g^{*T} Hg^* \left(\frac{s}{s^T y} - \frac{Hy}{y^T Hy}\right) \tag{2.8.27c}$$

en waarmee

$$d^* = (y^T Hy)\left(\frac{s}{s^T y} - \frac{Hy}{y^T Hy}\right) \tag{2.8.28}$$

De BFS-aanpassingsformule is in dit geval equivalent met de uitdrukking

$$H^* := H + \frac{ss^T}{s^T y} - \frac{Hyy^T H}{y^T Hy} + \frac{d^* d^{*T}}{y^T Hy} \qquad (2.8.29)$$

en in het bijzonder geldt de relatie (vgl. (2.8.5))

$$g^{*T} H^* g^* = g^{*T} Hg^* \qquad (2.8.30)$$

Bewijs : Juist als bij Stelling 2.8.9 is het bewijs voor een groot deel gebaseerd op de voor exacte lijnminimalisering geldende relatie $g^{*T} Hg = -\frac{1}{\alpha} g^{*T} s = 0$. Uitschrijven van de uitdrukking voor de zoekrichting geeft

$$d^* = -[H + \frac{ss^T}{s^T y} - \frac{Hyy^T H}{y^T Hy}] g^* - (y^T Hy)(\frac{s}{s^T y} - \frac{Hy}{y^T Hy})(\frac{s}{s^T y} - \frac{Hy}{y^T Hy})^T g^*$$

waarvan de eerste term leidt tot

$$-[H + \frac{ss^T}{s^T y} - \frac{Hyy^T H}{y^T Hy}] g^* = -Hg^* + Hy \frac{y^T Hg^*}{y^T Hy} = -(Hy + Hg) + Hy \frac{g^{*T} Hg^*}{y^T Hy}$$

$$= Hy \left(-1 + \frac{g^{*T} Hg^*}{y^T Hy} \right) - \frac{y^T Hg}{y^T s} s = g^T Hg \left(\frac{s}{s^T y} - \frac{Hy}{y^T Hy} \right)$$

en de tweede tot

$$-(y^T Hy)(\frac{s}{s^T y} - \frac{Hy}{y^T Hy})(\frac{s}{s^T y} - \frac{Hy}{y^T Hy})^T g^* = g^{*T} Hg^* (\frac{s}{s^T y} - \frac{Hy}{y^T Hy})$$

De in Stelling 2.8.9 afgeleide relatie (2.8.5a)(die geldt voor willekeurige H)

$$g^{*T}(H + \frac{ss^T}{s^T y} - \frac{Hyy^T H}{y^T Hy}) g^* = \frac{(g^{*T} Hg^*)(g^T Hg)}{g^{*T} Hg^* + g^T Hg}$$

samen met de uit (2.8.28) af te leiden relatie

$$\frac{(g^{*T} d^*)^2}{y^T Hy} = \frac{(g^{*T} Hg^*)^2}{g^{*T} Hg^* + g^T Hg}$$

leveren direct de laatste uitspraak (2.8.30) van de stelling. □

2.8.17. Opgemerkt moet worden dat de H in het voorgaande punt de H-matrix voorstelt die wordt gegenereerd met de BFS-formule. De vector d_I^* in (2.8.27)

$$d_I^* = -[H + \frac{ss^T}{s^Ty} - \frac{Hyy^TH}{y^THy}]g^*$$

is niet noodzakelijk de richting die zou worden gegenereerd (in dezelfde iteratiestap) in een perfecte quasi-Newton algorithme met de DFP-formule als aanpassingsformule. Wel geldt, op grond van de equivalentie uitspraak van Dixon (vgl. pt. 2.7.19) dat d_I^* een veelvoud is van de met een perfect quasi-Newton algorithme met de DFP-aanpassingsformule te genereren richting.

2.8.18. In het geval van de toepassing van de BFS-formule in een perfecte quasi-Newton algorithme voor de minimalisering van een positief definiete kwadratische vorm kan deze formule juist als de DFP-formule (vgl. pt. 2.8.10) worden herschreven als de som van twee matrices met ieder een eigen aanpassingsformule waarvan de een zorgt voor de generatie van A-geconjugeerde richtingen en de ander voor de opbouw van de inverse matrix van de Hessiaan. Uitgangspunt daarbij is de BFS-formule geschreven in de vorm (vgl. (2.7.20d))

$$H^* := [I - \frac{sy^T}{y^Ts}]\ H[I - \frac{ys^T}{s^Ty}] + \frac{ss^T}{s^Ty} \qquad (2.8.31)$$

Analoog aan het geval van DFP-formule (pt. 2.8.10) geldt dat onder de genoemde bijzondere omstandigheden hiervoor geschreven kan worden

$$H^* = Q^* + C^* \qquad (2.8.32)$$

waar

$$Q^* := [I - \frac{sy^T}{y^Ts}]\ Q[I - \frac{ys^T}{s^Ty}] \qquad (2.8.33)$$

$$C^* := C + \frac{ss^T}{s^Ty} \qquad (2.8.34)$$

met

$$Q^{(0)} := H^{(0)} \qquad Q^{(n)} := 0$$

en (2.8.35)

$$C^{(0)} := 0 \qquad C^{(n)} := A^{-1}$$

2.8.19. Een interessant aspect van het voorgaande is dat de A-geconjugeerde richtingen genererende matrix Q nauw gerelateerd blijkt aan de asymmetrische matrix H van de Gram-Schmidt A-orthogonalisatie-procedure welke werd gegeven door de uitdrukking (vgl. (2.7.35)).

$$H^* := H - \frac{Hys^T}{s^T y} = H[I - \frac{ys^T}{s^T y}] \tag{2.8.36}$$

De equivalentie van de zoekrichtingen gegenereerd door resp. de DFP-aanpassingsformule en de BFS-aanpassingsformule kan in verband hiermee in het speciale geval van de minimalisering van een positief definiete kwadratische vorm met een quasi-Newton algorithme met exacte lijnminimalisering worden afgeleid uit de equivalentie van de zoekrichtingen gegenereerd met resp. de projectie-algorithme van Pearson (vgl. (2.8.11), pt. 2.6.20) en de A-orthogonalisatie algorithme van Gram-Schmidt (pt. 2.6.24). De matrix C vervult onder deze speciale omstandigheden juist als in het geval bij de DFP-formule (vgl. pt. 2.8.10) geen rol bij de generatie van de zoekrichtingen.

2.8.20. Naar analogie van de uitdrukking (2.8.31) voor de BFS-formule had voor de DFP-formule (2.8.2) gebruik kunnen worden gemaakt van de schrijfwijze

$$H^* := [I - \frac{Hyy^T}{y^T Hy}]H[I - \frac{yy^T H}{y^T Hy}] + \frac{ss^T}{s^T y} \tag{2.8.37}$$

In overeenstemming daarmee had voor de zoekrichting genererende matrix P (2.8.8) in de DFP-formule geschreven kunnen worden

$$P^* := [I - \frac{Pyy^T}{y^T Py}]P[I - \frac{yy^T P}{y^T Py}] \tag{2.8.38}$$

Vergelijking van deze uitdrukking met die voor de zoekrichting genererende matrix Q (2.8.33) in de BFS-formule

$$Q^* := [I - \frac{sy^T}{s^T y}]Q[I - \frac{ys^T}{s^T y}] \qquad (2.8.139)$$

levert een extra illustratie van de in het voorgaande punt genoemde relatie die bestaat tussen de DFP-formule en de gradiënt-projectie-formule van Pearson enerzijds en de BFS-formule en Gram-Schmidt A-orthogonalisatie-formule anderzijds.

2.8.21. Als laatste opmerking in verband met de schrijfwijzen (2.8.31) voor de BFS-formule en (2.8.37) voor de DFP-formule zij vermeld dat door uitschrijven aangetoond kan worden dat de gehele convexe klasse van aanpassingsformules van Fletcher (pt. 2.7.12)

$$H^* = H^*_{DFP} + \phi(H^*_{BFS} - H^*_{DFP}) \qquad \phi \in [0,1]$$

kan worden weergegeven door de uitdrukking

$$H^* = [I - \frac{wy^T}{y^T w}] H [I - \frac{yw^T}{w^T y}] + \frac{ss^T}{s^T y} \qquad (2.8.40)$$

waar

$$\begin{aligned} w &= \frac{Hy}{y^T Hy} + \phi^{\frac{1}{2}}(\frac{s}{s^T y} - \frac{Hy}{y^T Hy}) \\ &= (1 - \phi^{\frac{1}{2}}) \frac{Hy}{y^T Hy} + \phi^{\frac{1}{2}} \frac{s}{s^T y} \end{aligned} \qquad (2.8.41)$$

met

$$\phi \in [0,1].$$

Deze laatste formules illustreren het speciale karakter van deze klasse van aanpassingsformules, waarvoor in de volgende paragraaf nog een andere bijzondere eigenschap zal worden aangetoond.

Literatuur

2.8.22. Voor meer details van het hierboven besprokene zij verwezen naar de litera-

tuur over quasi-Newton-methoden en in het bijzonder naar de volgende publikaties

[2.8.1] : Zie [1.1.1] Luenberger (1973)

[2.8.2] : Zie [1.1.3] Murray (1972)

[2.8.3] : Zie [2.7.6] Fletcher (1970)

[2.8.4] : Zie [2.7.11] Powell (1972)

[2.8.5] : Gill, P.E. and Murray, W.: Quasi-Newton methods for unconstrained optimization, J. Inst. Math. Appl. 9 (1972), pp. 91-108.

§ 2.9. Quasi-Newton-methoden III: Recente ontwikkelingen

2.9.1. De theorie van de quasi-Newton methoden zoals hiervoor gepresenteerd concentreerde zich voornamelijk rond de n-stappen-convergentie of Qn-eigenschap van deze methoden. Deze voor toepassingen op positief definiete kwadratische vormen belangrijke eigenschap hing samen met de generatie van A-geconjugeerde richtingen (in het geval van exacte lijnminimalisering) en met de successievelijke opbouw van de inverse van de Hessiaan. Bij toepassingen op niet-kwadratische functies en bij gebruik van andere technieken dan exacte lijnminimalisering voor stapgroottebepaling bleek vooral dit laatste aspect van het meeste belang: In n stappen wordt een benadering gegenereerd van de inverse van de Hessiaan waarmee een "quasi" Newton stap kan worden gezet. Voor cycli bestaande uit telkens n stappen, al dan niet voorafgegaan door een herstart (of "reset")kan met deze overwegingen superlineaire convergentie worden bewezen (voor de cycli). Eerst recent werd door o.a. Fletcher [2.9.3]en Oren en Luenberger [2.9.10] vooruitgang geboekt met een opzet om te pogen om in iedere iteratiestap (i.p.v. na n stappen) een zo groot mogelijke vermindering van de functiewaarde te bereiken en dat, bij voorkeur, met technieken voor de stapgrootte bepaling die minder functie-evaluaties vergen dan lijnminimalisering.

2.9.2. Centraal in de theorie achter deze ontwikkelingen staat de hieronder te bespreken uitspraak dat de verlaging van de functiewaarde in de k-de iteratie bij toepassing van een quasi-Newton algorithme nauw samenhangt met het quotiënt λ_n/λ_1 van de grootste gedeeld door kleinste eigenwaarde van de matrix

$$U^{(k)} := H^{(k)}G^{(k)} \tag{2.9.1}$$

Dit quotiënt is juist het conditiegetal (vgl. [2.9.6])t.o.v. de 2-norm van de matrix $U^{(k)}$,

$$c(U^{(k)}) := \|U^{(k)}\| \, \|(U^{(k)})^{-1}\| = \lambda_n^{(k)}/\lambda_1^{(k)} \tag{2.9.2}$$

In de theorie spelen naast de matrix $U^{(k)}$ vaak ook de daarvan (in de veronderstelling dat $H^{(k)}$ en $G^{(k)}$ beide positief definiet zijn) afgeleide matrices $Z^{(k)}$ en $K^{(k)}$ die, bij weglating van de indices (k), worden gedefinieerd door

$$Z := H^{\frac{1}{2}}GH^{\frac{1}{2}} := H^{-\frac{1}{2}}UH^{\frac{1}{2}} \tag{2.9.3}$$

en

$$K := G^{\frac{1}{2}}HG^{\frac{1}{2}} := G^{\frac{1}{2}}UG^{-\frac{1}{2}} \tag{2.9.4}$$

Op grond van de gelijkvormigheid van Z, K en U geldt

$$c(Z) = c(K) = c(U) \tag{2.9.5}$$

2.9.3. Illustratief voor de relatie tussen het conditiegetal van de matrix $U^{(k)}$ en de convergentiesnelheid van de toepassing van een quasi-Newton algorithme is de volgende uitspraak, die zelf weer een generalisatie is van een eerdere, soortgelijke uitspraak voor de methode van de steilste helling (zie pt. 2.4.8).

Stelling 2.9.3 : (Luenberger [2.9.1]) Bij toepassing van een perfecte quasi-Newton algorithme, (d.i. met exacte lijnminimalisering) voor de minimalisering van de positief definiet kwadratische vorm

$$f(x) = f(\hat{x}) + \frac{1}{2}(x - \hat{x})^T G(x - \hat{x}) \tag{2.9.6}$$

geldt (mits $H^{(k)}$ positief definiet) voor de vermindering van de functie in de (k+1)-de stap

$$f(x^{(k+1)}) - f(\hat{x}) \leq \left[\frac{c(U^{(k)}) - 1}{c(U^{(k)}) + 1}\right]^2 (f(x^{(k)} - f(\hat{x})) \tag{2.9.7}$$

waar $c(U^{(k)})$ het conditiegetal voorstelt van de matrix

$$U^{(k)} = H^{(k)}G$$

Bewijs : In het geval van exacte lijnminimalisering geldt in de(k+1)-de (met index (k+1) vervangen door een * en index (k) weglaten) iteratie

$$0 = g^{*T}d = (g + G(x^*-x))^T Hg = (g - \alpha GHg)^T Hg$$

waaruit volgt dat

$$\alpha = \frac{g^T Hg}{g^T HGHg}$$

Voor de functiewaarde in het nieuwe punt geldt daarmee

$$f(x^*) = f(\hat{x}) + \frac{1}{2}(x - \alpha Hg - \hat{x})^T G(x - \alpha Hg - \hat{x})$$

$$= f(\hat{x}) + \frac{1}{2}(x - \hat{x})^T G(x - \hat{x}) + \frac{1}{2}\alpha^2 g^T HGHg$$

$$- \alpha g^T HG(x - \hat{x})$$

$$= f(x) - \frac{1}{2}\frac{(g^T Hg)^2}{g^T HGHg}$$

met

$$f(x) - f(\hat{x}) = \frac{1}{2}(x - \hat{x})^T G(x - \hat{x}) = \frac{1}{2} g^T G^{-1} g$$

volgt dat

$$\frac{f(x) - f(x^*)}{f(x) - f(\hat{x})} = \frac{(g^T Hg)^2}{(g^T HGHg)(g^T G^{-1} g)}$$

of met en

$$p := H^{\frac{1}{2}} g \qquad Z := H^{\frac{1}{2}} G H^{\frac{1}{2}}$$

dat

$$\frac{f(x) - f(x^*)}{f(x) - f(\hat{x})} = \frac{(p^T p)^2}{(p^T Zp)(p^T Z^{-1} p)}$$

of equivalent dat

$$f(x^*) - f(\hat{x}) = (f(x) - f(\hat{x}))\left(1 - \frac{(p^T p)^2}{(p^T Zp)\ (p^T Z^{-1} p)}\right)$$

Met de reeds eerder genoemde (pt. 2.4.8) ongelijkheid van Kantorovich-Bergstrom

$$\frac{(p^Tp)^2}{(p^TZp)(p^TZ^{-1}p)} \geq \frac{4\lambda_1\lambda_n}{(\lambda_1 + \lambda_n)^2} = \frac{4c\,(Z)}{(1 + c(Z))^2} \qquad (2.9.8)$$

en het gelijk zijn van de conditiegetallen van de matrices $Z = H^{\frac{1}{2}}GH^{\frac{1}{2}}$ en $U = HG$ volgt het gewenste resultaat onmiddellijk. □

2.9.4. In pt. 2.7.14 werd opgemerkt (en in 2.7.17 aangetoond) dat alle H-matrices gegenereerd in quasi-Newton algorithmen met aanpassingsformules behorend tot de familie van Huang (2.7.19) in het geval van toepassing op positief definiete kwadratische functies en gebruik van exacte lijnminimalisering voldoen aan de erfelijkheids- en quasi-Newton relaties, d.i.

$$H^{(k+1)}y^{(j)} = \rho s^{(j)} \qquad j = 0, \ldots, k$$

Deze relaties impliceren in het gebruikelijke geval waarbij $\rho = 1$ dat de matrix $H^{(k+1)}G$ k+1 eigenwaarden λ_j bezit met de waarde $\lambda_j = 1$ en als corresponderende eigenvectoren de vectoren $s^{(j)}$. Immers

$$H^{(k+1)}y^{(j)} = H^{(k+1)}Gs^{(j)} = s^{(j)} \qquad j = 0, \ldots, k \qquad (2.9.9)$$

In iedere stap wordt telkens een nieuwe combinatie eigenwaarde-eigenvector gevonden met eigenwaarde 1 zodat na n stappen alle eigenwaarden van $H^{(n)}G$ gelijk zijn aan 1 en geldt (vgl. (2.7.9))

$$H^{(n)}G = I \qquad (2.9.10)$$

De onderlinge verhouding van de eigenwaarden λ_j van de matrix $U^{(k+1)} = H^{(k+1)}G$ en in het bijzonder de verhouding tussen de grootste en de kleinste eigenwaarde, d.w.z. het conditiegetal $c(U^{(k+1)})$, zal tijdens het iteratieproces veranderen en wel in het algemeen op een niet van te voren geheel duidelijke manier. De eerste start tot de studie van het gedrag van deze eigenwaarden werd gegeven door Fletcher in 1970 in [2.9.3] waarin hij aantoonde dat voor een beperkte klasse van aanpassingsformules uit de familie van Huang geldt dat de eigenwaarden van de matrices $H^{(k)}G$ monotoon convergeren naar 1. Daarna toonden Oren en Luenberger in [2.9.11] aan dat, door invoering van een extra schaling van de H-matrices uit de beperkte klasse van Fletcher, het mogelijk was te garanderen dat ook de conditiegetallen $c(U^{(k+1)})$ monotoon convergeren

naar 1. Zowel Fletcher als Oren en Luenberger gebruikten hun nieuw verworven inzichten voor het ontwikkelen van nieuwe quasi-Newton algorithmen die in een aantal gevallen efficienter zijn gebleken dan de tot dan gebruikelijke algorithmen.

Monotone convergentie van de eigenwaarden van de matrices $H^{(k)}G$ en de convexe klasse van aanpassingsformules van Fletcher

2.9.5. De theoretische basis voor de hierna te bespreken uitspraak van Fletcher over de monotone convergentie van de eigenwaarden van de matrices $H^{(k)}G$ wordt geleverd door het volgende Lemma van Loewner uit de matrix theorie dat in de Engelstalige literatuur ook bekendheid geniet als het "interlocking eigenvalue lemma".

Lemma 2.9.5.(Loewner [2.9.9]): Als $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$ de naar grootte gerangschikte eigenwaarden voorstellen van de symmetrische n x n matrix A en $\mu_1 \leq \mu_2 \leq \ldots \leq \mu_n$ de analoog gerangschikte eigenwaarden van de matrix $A + \sigma aa^T$ waar a een n-vector is, dan geldt dat

$$\begin{aligned} &\mu_1 \leq \lambda_1 \leq \mu_2 \leq \lambda_2 \leq \ldots \leq \mu_n \leq \lambda_n \quad \text{als} \quad \sigma < 0 \\ &\lambda_1 \leq \mu_1 \leq \lambda_2 \leq \mu_2 \leq \ldots \leq \lambda_n \leq \mu_n \quad " \quad \sigma > 0 \end{aligned} \tag{2.9.11}$$

Bewijs : Voor het bewijs moet worden verwezen naar matrix theorieboeken (zoals dat van Wilkinson [2.9.13] en Gantmacher [2.9.7]).

2.9.6. Met behulp van het Lemma van Loewner bewees Fletcher de volgende uitspraak

Stelling 2.9.6 : (Fletcher [2.9.3]) Voor iedere H-matrix gegenereerd met een convexe combinatie van de DFP- en de BFS-aanpassingsformules, d.i. met een aanpassingsformule van de vorm (2.7.31)

$$H_\phi^* = (1 - \phi)\, H_{DFP}^* + \phi H_{BFS}^* \tag{2.9.12}$$

waar

$$\phi \in [0,1]$$

geldt bij toepassing in een quasi-Newton algorithme voor de minimalisering van een positief definiete kwadratische vorm met Hessiaan G dat de naar grootte gerangschikte eigenwaarden van de matrix HG monotoon convergeren naar 1, d.w.z. als $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$ en $\lambda_1^* \leq \lambda_2^* \leq \ldots \leq \lambda_n^*$ de eigenwaarde zijn van resp. $H_\phi G$ en $H_\phi^* G$ dan geldt

$$|\lambda_j^* - 1| \leq |\lambda_j - 1| \qquad j = 1, \ldots, n \tag{2.9.13}$$

Bewijs : Met behulp van de transformaties

$$z := G^{\frac{1}{2}} s \qquad\qquad K_\phi := G^{\frac{1}{2}} H_\phi G^{\frac{1}{2}} \tag{2.9.14}$$

en gebruikmaking van de relaties

$$y = G^{\frac{1}{2}} z \qquad\qquad s = G^{-\frac{1}{2}} z \tag{2.9.15}$$

volgt als aanpassingsformule voor de matrix K_0 corresponderend met de DFP-aanpassingsformule (waarvoor $\phi = 0$) de uitdrukking

$$\begin{aligned} K_0^* &:= K_0 - \frac{K_0 z z^T K_0}{z^T K_0 z} + \frac{z z^T}{z^T z} \\ &:= \bar{K}_0 + \frac{z z^T}{z^T z} \end{aligned} \tag{2.9.16}$$

Voor de onderlinge relatie van de eigenwaarden $\bar{\lambda}_1 \leq \bar{\lambda}_2 \leq \ldots \leq \bar{\lambda}_n$ van de matrix $\bar{K}_0$ en de eigenwaarden $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$ van de matrix K_0 geldt op grond van het voorgaand Lemma 2.9.5 dat

$$0 = \bar{\lambda}_1 < \lambda_1 \leq \bar{\lambda}_2 \leq \lambda_2 \leq \ldots \leq \bar{\lambda}_n \leq \lambda_n$$

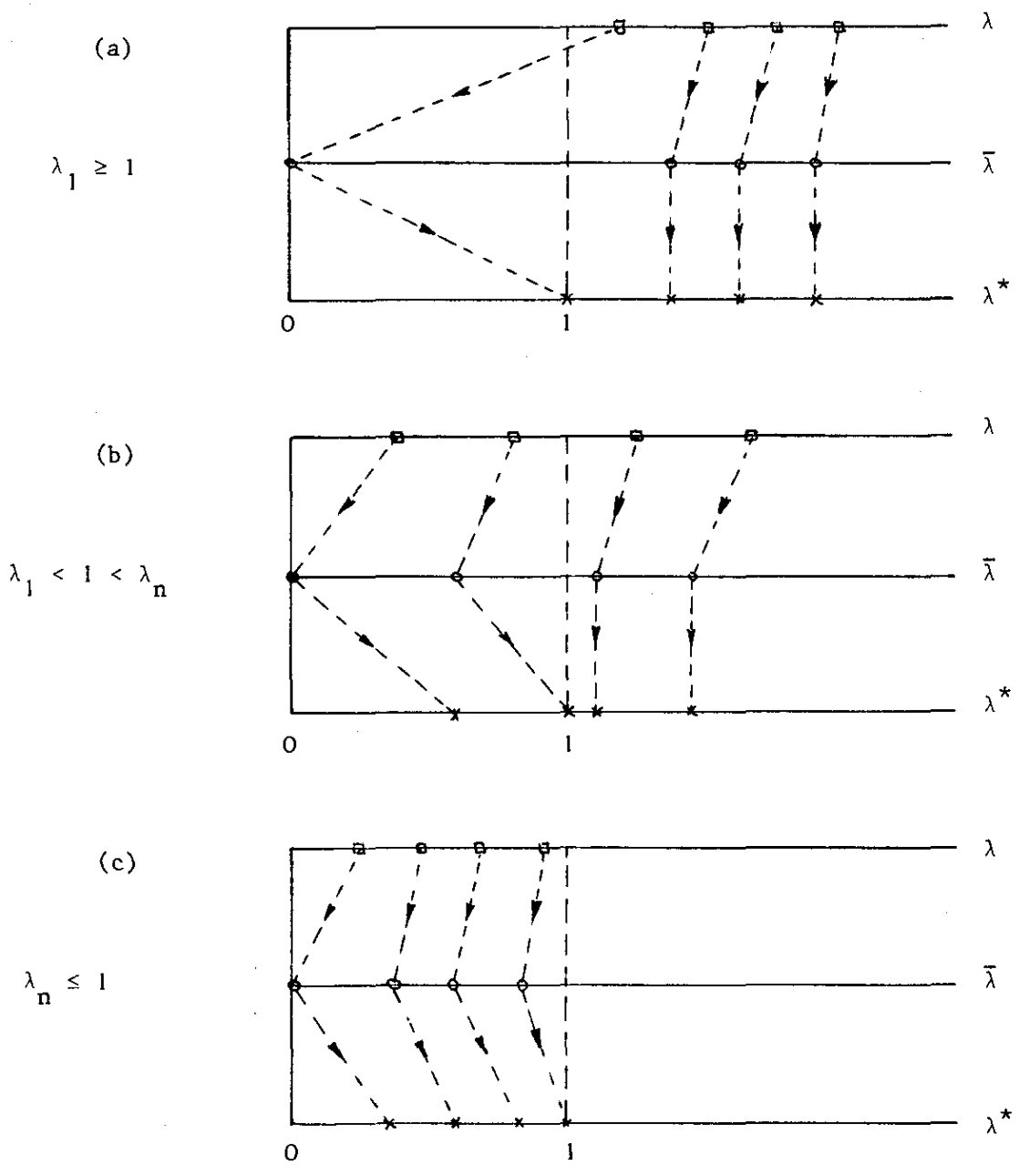
waarbij z de eigenvector is die correspondeert met de eigenwaarde $\bar{\lambda}_1 = 0$ en de overige eigenvectoren het crthogonale complement van z opspannen. Toevoeging van de rang-één-matrix zz^T/z^Tz aan de matrix $\bar{K}_0$ verandert de eigenwaarde 0 voor de eigenvector z van 0 in 1 en laat de overige eigenwaarden onveranderd. Deze veranderingen van de eigenwaarden zijn voor de drie te onderscheiden gevallen

(a) $\lambda_1 \geq 1$

(b) $\lambda_1 < 1 < \lambda_n$

(c) $\lambda_n \leq 1$

geschetst in Figuur 2.9.6



Figuur 2.9.6.: Verloop van de in grootte gerangschikte eigenwaarden van de matrix $K_0 := G^{\frac{1}{2}} H_0 G^{\frac{1}{2}}$

Met behulp van deze schetsen kunnen de voor het bewijs benodigde ongelijkheden direct worden opgeschreven. Bijvoorbeeld in geval (b) geldt, in de veronderstelling dat er i eigenwaarden kleiner of gelijk zijn aan 1 en n-i eigenwaarden groter dan 1, dat

$$0 = \bar{\lambda}_1 < \lambda_1 \leq \lambda_1^* = \bar{\lambda}_2 \leq \lambda_2 \leq \ldots \leq \lambda_{i-1} \leq \lambda_{i-1}^* = \bar{\lambda}_i \leq \lambda_i \leq 1 = \lambda_i^*$$

$$1 \leq \lambda_{i+1}^* = \bar{\lambda}_{i+1} \leq \lambda_{i+1} \leq \ldots \leq \lambda_n^* = \bar{\lambda}_n \leq \lambda_n \qquad (2.9.17)$$

Dit completeert de essentie van het bewijs voor $\phi = 0$.
Voor de inverse matrix B_1 van de matrix H_1 gegenereerd met de BFS-formule (waarvoor $\phi = 1$) geldt volgens Stelling 2.8.13 de aanpassingsformule

$$B_1^* := B_1 - \frac{B_1 s s^T B_1}{s^T B_1 s} + \frac{y y^T}{y^T s} \qquad (2.9.18)$$

De transformaties

$$z := G^{-\frac{1}{2}} y \qquad M_1 := G^{-\frac{1}{2}} B_1 G^{-\frac{1}{2}} \qquad (2.9.19)$$

en de relaties

$$s = G^{-\frac{1}{2}} z \qquad y = G^{\frac{1}{2}} z \qquad (2.9.20)$$

leveren op analoge wijze als boven als aanpassingsformule voor de matrix M_1

$$M_1^* := M_1 - \frac{M_1 z z^T M_1}{z^T M_1 z} + \frac{z z^T}{z^T z} \qquad (2.9.21)$$

Herhaling van de boven gegeven argumentatie levert voor de eigenwaarden $\mu_1^* \leq \mu_2^* \leq \ldots \leq \mu_n^*$ van M_1^* en $\mu_1 \leq \mu_2 \leq \ldots \leq \mu_n$ van M_1 de relaties

$$|1 - \mu_j^*| \leq |1 - \mu_j| \qquad j = 1, \ldots, n$$

Aangezien de eigenwaarden μ_j^* en μ_j juist de inversen zijn van de eigenwaarden

λ^*_{n-j} en λ_{n-j} van de matrices K_1^* en K_1 (2.9.14) omdat

$$(K_1^*)^{-1} = G^{-\frac{1}{2}}H_1^*G^{-\frac{1}{2}} \qquad K_1^{-1} = G^{-\frac{1}{2}}H_1^{-1}G^{-\frac{1}{2}}$$

volgt onmiddellijk dat

$$\left|1 - \frac{1}{\lambda^*_{n-j}}\right| \leq \left|1 - \frac{1}{\lambda_{n-j}}\right|$$

waaruit bij uitwerken volgt dat

$$\begin{aligned} &\lambda^*_{n-j} \leq \lambda_{n-j} \quad \text{als} \quad \lambda^*_{n-j} \geq 1 \\ &\lambda^*_{n-j} \geq \lambda_{n-j} \quad \text{als} \quad \lambda^*_{n-j} < 1 \end{aligned} \tag{2.9.22}$$

of equivalent

$$\left|1 - \lambda^*_{n-j}\right| \leq \left|1 - \lambda_{n-j}\right| \qquad j = 1, \ldots, n$$

Dit completeert het bewijs voor $\phi = 1$.
Voor de matrix $K^*_\phi = G^{\frac{1}{2}}H^*_\phi G^{\frac{1}{2}}$ met willekeurige $\phi \in [0,1]$ kan met behulp van de transformatie

$$w = G^{\frac{1}{2}}v \tag{2.9.23}$$

geschreven worden

$$\begin{aligned} K^*_\phi &:= G^{\frac{1}{2}}(H^*_0 + \phi vv^T)G^{\frac{1}{2}} \\ &:= K^*_0 + \phi ww^T \\ &:= K^*_1 - (1-\phi)ww^T \end{aligned} \tag{2.9.24}$$

Voor ieder van eigenwaarden $\lambda^*_{1,\phi} \leq \lambda^*_{2,\phi} \leq \ldots \leq \lambda^*_{n,\phi}$ van K^*_ϕ met $\phi \in [0,1]$ geldt op grond van deze schrijfwijze en het Lemma 2.9.5 dat

$$\lambda^*_{j,0} \leq \lambda^*_{j,\phi} \leq \lambda^*_{j,1} \tag{2.9.25}$$

Combinatie van deze ongelijkheid met de ongelijkheden (2.9.17) voor $\lambda^*_{j,0}$ voor die j waarvoor $\lambda^*_{j,\phi} \leq 1$ en met de gevonden ongelijkheden (2.9.22) voor $\lambda^*_{j,1}$ voor de overige j resulteert in de gezochte conclusie dat voor de eigenwaarden $\lambda^*_{j,\phi}$ van de matrix $K^*_\phi = G^{\frac{1}{2}} H^*_\phi G^{\frac{1}{2}}$ voor alle waarden van $\phi \in [0,1]$ geldt

$$|1 - \lambda^*_{j,\phi}| \leq |1 - \lambda_j| \qquad j = 1, \ldots, n$$

waar λ_j de j-de eigenwaarde (in grootte) is van de matrix $K = G^{\frac{1}{2}} H G^{\frac{1}{2}}$ $(= K_0 = K_1 = K_\phi)$. De gelijkheid van de eigenwaarden van de matrices $K_\phi = G^{\frac{1}{2}} H_\phi G^{\frac{1}{2}}$ en $U_\phi = H_\phi G$, resp. $K^*_\phi = G^{\frac{1}{2}} H^*_\phi G^{\frac{1}{2}}$ en $U^*_\phi = H^*_\phi G$ completeert het bewijs. □

2.9.7. Fletcher [2.9.3] toonde met behulp van een tegenvoorbeeld aan dat de conditie (2.9.12)

$$\phi \in [0,1]$$

een noodzakelijke conditie is voor de monotonie van de eigenwaarden. Hij beschouwde daartoe het geval waarbij

$$G = \begin{pmatrix} 1+\varepsilon & \sqrt{\varepsilon} \\ \sqrt{\varepsilon} & \varepsilon \end{pmatrix} \qquad H = I \qquad z = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{2.9.26}$$

De eigenwaarden van de matrix K = G zijn in dit geval gelijk aan resp. η en $1 + 2\varepsilon - \eta$ waar

$$\eta = \frac{1}{2}\left(1 + 2\varepsilon - \sqrt{1 + 4\varepsilon}\right) = O(\varepsilon^2) \tag{2.9.27}$$

Geverifieerd kan worden dat voor de gevallen $\phi = -\varepsilon$ en $\phi = 1 + \varepsilon$ resp. geldt

$$K^*_{-\varepsilon} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \text{ en } K^*_{1+\varepsilon} = \begin{pmatrix} 1+2\varepsilon & 0 \\ 0 & 1 \end{pmatrix}$$

De kleinste eigenwaarde van $K^*_{-\varepsilon}$ is gelijk aan 0 zodat

$$\lambda^*_1 = 0 < \eta = \lambda_1$$

Analoog geldt dat de grootste eigenwaarde van $K^*_{1+\varepsilon}$ gelijk is aan $1 + 2\varepsilon$ zodat

$$\lambda^*_2 = 1 + 2\varepsilon > 1 + 2\varepsilon - \eta = \lambda_2$$

Een negatieve ϕ verkleint in dit geval de kleinste eigenwaarde en een ϕ-waarde groter dan 1 vergroot de grootste eigenwaarde. (Het conditiegetal wordt daarbij ∞ in het eerste geval en gelijk aan $1 + 2\varepsilon$ in het tweede geval. In dit laatste geval verslechtert het conditiegetal dus niet)

2.9.8. De rang-één formule (pt. 2.8.1) is een voorbeeld van een aanpassingsformule die niet tot de convexe klasse van Fletcher behoort. Immers de ϕ-waarde die er mee correspondeert is gelijk aan (vgl. pt. 2.7.11)

$$\phi_{RI} = \frac{s^T y}{(s - Hy)^T y} = \frac{s^T y}{s^T y - y^T H y}$$

ofwel (2.9.28)

$$\phi_{RI} = \left(1 - \frac{y^T H y}{s^T y}\right)^{-1}$$

Deze laatste uitdrukking impliceert dat afhankelijk van de waarde van de breuk

$$\frac{y^T H y}{s^T y} = \frac{y^T H y}{y^T G^{-1} y} \qquad (2.9.29)$$

voor de rang-één-aanpassingsformule geldt

$$1 < \frac{y^T H y}{y^T G^{-1} y} < \infty \quad : \quad \phi < 0$$

$$0 < \frac{y^T H y}{y^T G^{-1} y} < 1 \quad : \quad \phi > 1 \qquad (2.9.30)$$

In het eerste geval geldt dat een aantal eigenwaarden van de benadering H groter zijn dan de corresponderende eigenwaarden van de inverse Hessiaan G^{-1}. De negatieve ϕ-waarde van de rang-één-aanpassingsformule bewerkt in dat geval een gunstige verkleining van de eigenwaarden van de matrix H^* t.o.v. die van de matrix H. In het tweede geval geldt het tegenovergestelde en bewerkt de waarde $\phi > 1$ een gunstige vergroting van de eigenwaarden van de matrix H^*.

Algorithme van Fletcher

2.9.9. Fletcher [2.9.3] gebruikte het door hem verkregen en hierboven weergegeven inzicht in het gedrag van de eigenwaarden van de matrix $U^{(k)} = H^{(k)}G$ voor de constructie van een quasi-Newton algorithme waarin afwisselend gebruik gemaakt wordt van zowel de DFP-als de BFS-aanpassingsformule en waarbij wordt afgezien van lijnminimalisering voor de bepaling van de stap-grootte. De keus van de aanpassingsformule (in stap (iii) van de standaard quasi-Newton algorithme (vgl. pt. 2.7.2))maakte Fletcher naar analogie met de rang-één formule afhankelijk van de waarde van het quotiënt y^THy/s^Ty en wel volgens de volgende regel

$$\begin{aligned} &\text{als } \frac{y^THy}{s^Ty} > 1 \quad \text{kies} \quad \phi = 0 \quad \text{(DFP-formule)} \\ &\text{"} \quad \frac{y^THy}{s^Ty} \leq 1 \quad \text{"} \quad \phi = 1 \quad \text{(BFS-formule)} \end{aligned} \tag{2.9.31}$$

2.9.10. Inplaats van het gebruik van lijnminimalisering voor de stapgrootte bepaling (stap (iv) van de standaard quasi-Newton algorithme) suggereerde Fletcher om voor $\alpha^{(k)}$ na de eerste n stappen steeds de waarde $\alpha^{(k)} = 1$ te kiezen en afhankelijk van de werkelijke afname van de functiewaarde Δf relatief t.o.v. een theoretisch mogelijke afname $\nabla^T fs = g^Ts$ deze stapgrootte factor met een constante factor te verkleinen. In de eerste n stappen wordt α i.v.m. nog niet aangepaste schaling van de H-matrix in eerste instantie gelijk gekozen aan de uiteindelijke α-waarde in de voorgaande stap. De stapgroottebepaling verloopt dan volgens de volgende algorithme (vgl. pt. 2.2.11)

(0) zet $\alpha^{(k)} := 1$ als $k = 0$ of $k \geq n$ en

$\alpha^{(k)} := \alpha^{(k-1)}$ anders

(i) zet $x^{(k+1)} := x^{(k)} - \alpha^{(k)} H^{(k)} g^{(k)}$ en evalueer

$\Delta f^{(k)} := f(x^{(k+1)}) - f(x^{(k)})$, $s^{(k)} := x^{(k+1)} - x^{(k)}$ en

$g^{(k)T} s^{(k)} = \nabla^T f(x^{(k)}) s^{(k)}$

(ii) als $\Delta f^{(k)} / g^{(k)T} s^{(k)} > 10^{-4}$ dan klaar; zo niet :

(iii) zet $\alpha^{(k)} := \alpha^{(k)}/10$ en ga terug naar stap (i)

2.9.12. Ter verhoging van de efficiëntie in het geval de afgeleide van de objectfunctie langs de lijn eenvoudig te berekenen is verving Fletcher zelf stap (iii) van de hier gegeven algorithme door

(iii) bepaald het minimum $\bar{\alpha}$ van een cubische benadering van $f(x^{(k)} + \alpha d^{(k)})$ en zet $\alpha^{(k)} := \max [\alpha^{(k)}/10, \bar{\alpha}]$

Bij de numerieke experimenten uitgevoerd door Himmelblau [2.9.8] kwam deze algorithme van Fletcher (volgens Himmelblau vooral als gevolg van deze manier van stap-grootte bepaling) als een van de meest efficiënte quasi-Newton algorithmen naar voren.

Monotone convergentie van de conditiegetallen van de matrices $H^{(k)}G$

2.9.13. De monotone convergentie van de eigenwaarden van de matrices $U^{(k)} = H^{(k)}G$ naar 1 geïntroduceerd door Fletcher impliceert niet altijd dat ook de corresponderende conditiegetallen er beter (d.i. kleiner) op worden. In het geval dat b.v. alle eigenwaarden in elkaars nabijheid liggen en daarbij aanzienlijk groter (of kleiner) zijn dan 1 (vgl. Figuur 2.9.6) betekent de introductie van een nieuwe eigenwaarde gelijk aan 1, zoals bij gebruik van een aanpassingsformule uit de klasse van Fletcher, een aanzienlijke verslechtering van het conditiegetal. In de volgende stappen herstelt de algorithme zich wel weer enigszins doch het kan een relatief groot aantal iteratie stappen kosten voor het conditiegetal van

de matrix $H^{(k)}G$ kleiner wordt dan het conditiegetal van $H^{(0)}G = G$. De één-staps convergentiesnelheid van de quasi-Newton-methode is gedurende die periode in het iteratieproces zelfs slechter dan de methode van de steilste helling. Een eenvoudig uit het bewijs van Stelling 2.9.6 en de Figuur 2.9.6 te destilleren voorwaarde dat geen verslechtering van het conditiegetal optreedt is de voorwaarde dat voor de grootste en de kleinste eigenwaarden van de matrix $H^{(0)}G$ geldt

$$\lambda_1 \leq 1 \leq \lambda_n \tag{2.9.32}$$

Wordt aan deze voorwaarde in iedere stap van het iteratieproces voldaan dan convergeren ook de conditiegetallen van de matrices $U^{(k)} = H^{(k)}G$ monotoon naar 1.

2.9.14. Aan de genoemde voorwaarde (2.9.32) kan in iedere stap van het iteratieproces worden voldaan door toepassing van schaling. Een elegante procedure hiervoor werd naar voren gebracht door Oren en Luenberger [2.9.11]. Deze construeerden een nieuw type aanpassingsformule voor quasi-Newton methoden gebaseerd op de convexe klasse van aanpassingsformules van Fletcher (2.9.12) met daarin als extra parameter opgenomen een positieve schaalfactor γ. In formule vorm

$$H^* := (H - \frac{Hyy^TH}{y^THy} + \phi vv^T)\gamma + \frac{ss^T}{s^Ty} \tag{2.9.33}$$

waar als voorheen (vgl. (2.7.27))

$$v = (y^THy)^{\frac{1}{2}} (\frac{s}{s^Ty} - \frac{Hy}{y^THy}) \tag{2.9.34}$$

Voor $\gamma = 1$ is deze formule gelijk aan de formule voor de familie van symmetrische aanpassingsformules van Fletcher (2.7.31). In het algemeen ($\gamma > 0$) is de formule met

$$\overline{H} := \gamma H \tag{2.9.35}$$

te schrijven als

$$H^* := \bar{H} - \frac{\bar{H}yy^T\bar{H}}{y^T\bar{H}y} + \phi\bar{v}\bar{v}^T + \frac{ss^T}{s^Ty} \tag{2.9.36}$$

waar

$$\bar{v} := (y^T\bar{H}y)^{\frac{1}{2}} \left(\frac{s}{s^Ty} - \frac{\bar{H}y}{y^T\bar{H}y}\right) \tag{2.9.37}$$

Met behulp van deze formulering kan eenvoudig worden aangetoond, omdat alle eigenwaarden van de matrix $\bar{H}G$ juist een factor γ van de eigenwaarden van de matrix HG verschillen, dat aan de voorwaarde (2.9.32) voor monotone convergentie van de conditiegetallen van de matrices $H^{(k)}G$ kan worden voldaan door γ zodanig te kiezen dat geldt

$$\bar{\lambda}_1 = \gamma\lambda_1 \leq 1 \leq \gamma\lambda_n = \bar{\lambda}_n \tag{2.9.38}$$

of, equivalent

$$1/\lambda_n \leq \gamma \leq 1/\lambda_1 \tag{2.9.39}$$

waar λ_1 en λ_n resp. de kleinste en de grootste eigenwaarden zijn van de matrix HG. Voor quasi-Newton algorithmen die gebruik maken van de aanpassingsformule (2.9.33) met $\phi \in [0,1]$ en $\gamma \in [1/\lambda_n, 1/\lambda_1]$ werd door Oren en Luenberger [2.9.11] de naam zelfschalende variabele metriek - of SSVM - (self-scaling-variable metric) algorithmen geïntroduceerd.

2.9.15. Voor het controleren van de voorwaarde voor de schaalfactor γ in iedere iteratiestap zouden telkens de extreme eigenwaarden van de matrices $H^{(k)}G$ moeten worden bepaald. In plaats daarvan wordt in de practijk echter gebruik gemaakt van de in het volgende lemma gegeven afschattingen.

Lemma 2.9.15. Onder voorwaarde dat voldaan wordt aan de (in de toepassing van quasi-Newton algorithmen gebruikelijke) relaties

$$y = Gs \qquad s = -\alpha Hg \qquad s^Ty > 0 \tag{2.9.40}$$

waarin H en G positief definiete matrices voorstellen geldt voor alle $\eta \in [0,1]$ de relatie

$$1/\lambda_n \leq (1 - \eta)\left(\frac{s^T y}{y^T H y}\right) + \eta\left(\frac{g^T s}{g^T H y}\right) \leq 1/\lambda_1 \qquad (2.9.41)$$

waar λ_1 en λ_n resp. de kleinste en de grootste eigenwaarden zijn van de matrix HG.

Bewijs : De uitdrukking (2.9.41) heeft betrekking op alle punten in het interval $[\frac{s^T y}{y^T H y}, \frac{g^T s}{g^T H y}]$. Het is daarom voldoende om de betreffende ongelijkheid te bewijzen voor de extreme punten van het interval. Met de definities (2.9.14)

$$z := G^{\frac{1}{2}} s \qquad K := G^{\frac{1}{2}} H G^{\frac{1}{2}}$$

en de daarvan afgeleide relaties (2.9.15)

$$s = G^{-\frac{1}{2}} z \qquad y = G^{\frac{1}{2}} z$$

volgt dat

$$\frac{s^T y}{y^T H y} = \frac{z^T G^{-\frac{1}{2}} G^{\frac{1}{2}} z}{z^T G^{\frac{1}{2}} H G^{\frac{1}{2}} z} = \frac{z^T z}{z^T K z} \qquad (2.9.42)$$

De uit de theorie van positief definiete kwadratische vormen bekende ongelijkheid

$$1/\lambda_n \leq \frac{z^T z}{z^T K z} \leq 1/\lambda_1 \qquad (2.9.43)$$

waar λ_1 en λ_n resp. de kleinste en grootste eigenwaarden zijn van de matrix K. De gelijkheid van de eigenwaarden van de gelijkvormige matrices K en U = HG leveren daarna direct het gewenste resultaat.

Met de eerdergenoemde definities en relaties en de observatie dat als $\alpha \neq 0$

$$g = -\frac{1}{\alpha} H^{-1}s \tag{2.9.44}$$

volgt dat

$$\frac{g^T s}{g^T Hy} = \frac{s^T H^{-1} s}{s^T y} = \frac{z^T G^{-\frac{1}{2}} H^{-1} G^{-\frac{1}{2}} z}{z^T G^{-\frac{1}{2}} G^{\frac{1}{2}} z} = \frac{z^T K^{-1} z}{z^T z} \tag{2.9.45}$$

De uit de matrix theorie bekende ongelijkheid

$$1/\lambda_n \leq \frac{z^T K^{-1} z}{z^T z} \leq 1/\lambda_1 \tag{2.9.46}$$

levert daarna juist als in het voorgaande geval het gewenste resultaat. □

Opm. : Uit de met de ongelijkheid van Cauchy-Schwartz (na de definitie van vectoren $u := K^{-\frac{1}{2}} z$ en $v := K^{\frac{1}{2}} z$) te bewijzen ongelijkheid

$$\left(\frac{z^T K^{-1} z}{z^T z}\right) \left(\frac{z^T K z}{z^T z}\right) \geq 1 \tag{2.9.47}$$

volgt nog dat steeds geldt

$$\frac{s^T y}{y^T Hy} \leq \frac{g^T s}{g^T Hy} \tag{2.9.49}$$

Zelfschalende variabele metriek (of SSVM)- algorithmen van Oren en Luenberger

2.9.16. Met het resultaat van het voorgaande lemma in gedachte definieerde Oren [2.9.10] de volgende twee-parameter aanpassingsformule voor zelfschalende variabele-metriek algorithmen (vgl. (2.9.33))

$$H^* := (H - \frac{Hyy^T H}{y^T Hy} + \phi vv^T)\gamma + \frac{ss^T}{s^T y} \tag{2.9.50}$$

waar

$$v := (y^T Hy)^{\frac{1}{2}} \left(\frac{s}{s^T y} - \frac{Hy}{y^T Hy}\right) \tag{2.9.51}$$

$$\gamma := (1 - \eta) \frac{s^T y}{y^T H y} + \eta \frac{g^T s}{g^T H y} \tag{2.9.52}$$

en met als voorwaarden dat zowel

$$\phi \in [0,1]$$

als (2.9.53)

$$\eta \in [0,1]$$

2.9.17. Oren [2.9.10] bewees dat gebruik van deze aanpassingsformule in een quasi-Newton algorithme, indien gestart wordt met een positief definiete beginmatrix $H^{(0)}$ en indien in iedere iteratie voldaan wordt aan de eis dat $s^T y > 0$, resulteert in de volgende bijzondere eigenschappen:

a) de successievelijk gegenereerde H-matrices zijn alle positief definiet

b) bij minimalisering van een positief definiete kwadratische vorm met Hessiaan G geldt voor de conditiegetallen van de successievelijk gegenereerde matrices $U^{(k)} = H^{(k)}G$ dat

$$c(U^*) \leq c(U) \tag{2.9.54}$$

c) bij minimalisering van een positief definiete kwadratische vorm en gebruik van lijnminimalisering voor de stapgroottebepaling zijn de successievelijk gegenereerde richtingen onderling G-geconjugeerd en geldt op grond daarvan dat de algorithme de Q_n-eigenschap bezit. In dit geval geldt in het algemeen (tenzij $\gamma = 1$ in alle opvolgende stappen) in tegenstelling tot de meeste andere bekende quasi-Newton aanpassingsformules dat (vgl. (2.9.10))

$$H^{(n)}G \neq I \tag{2.9.55}$$

2.9.18. De keuzevrijheid t.a.v. de parameters η (of γ) en ϕ leidt onmiddellijk tot vraag naar mogelijk optimale combinaties. Door Oren en Spedicato [2.9.12] werd hiernaar een onderzoek ingesteld waarbij zij als optimaliteitscriterium hanteerden het criterium van een zo groot mogelijke afname van een door hem gevonden bovengrens voor het conditiegetal. Afhankelijk van de ligging van

de getallen (vgl. pt. 2.9.15) :

$$k_1 = \frac{s^T y}{y^T H y} \qquad k_2 = \frac{g^T s}{g^T H y} \tag{2.9.56}$$

t.o.v. het getal 1 vonden zij als optimale strategie:

$$\begin{aligned} k_2 \leq 1 \quad &: \eta = 1 \quad (\gamma = k_2) \quad \phi = 0 \\ k_1 < 1 < k_2 \quad &: \eta = \frac{1 - k_1}{k_2 - k_1} \quad (\gamma = 1) \quad \phi = k_1 \frac{k_2 - 1}{k_2 - k_1} \\ k_1 \geq 1 \quad &: \eta = 0 \quad (\gamma = k_1) \quad \phi = 1 \end{aligned} \tag{2.9.57}$$

Voor de argumentatie van deze keuze zij verwezen naar het artikel van Oren en Spedicato [2.9.12] . Het eerste effect ervan is dat voor γ steeds een waarde gekozen wordt die binnen de toelaatbare grenzen zo dicht mogelijk bij 1.0 ligt. De corresponderende ϕ-waarden stemmen voor een deel overeen met de voor de algorithme van Fletcher aanbevolen waarden (vgl. pt. 2.9.9)

2.9.19. Oren en Spedicato [2.9.12] onderzochten de in het voorgaande punt genoemde strategie samen met een drietal andere soortgelijke (op de directe keuze van γ gebaseerde) strategieën nl.

$$\text{I} \qquad \gamma = \sqrt{k_1 k_2} \qquad \phi = (1 + \sqrt{\frac{k_2}{k_1}})^{-1} \tag{2.9.58}$$

$$\text{II} \qquad \gamma = k_1 k_2 \qquad \phi = \frac{1}{2} \tag{2.9.59}$$

$$\begin{aligned} \text{III} \qquad k_2 \leq 1 \quad &: \gamma = k_2 \quad \phi = 0 \\ k_1 \leq 1 \leq k_2 \quad &: \gamma = 1 \quad \phi = \frac{k_2 - 1}{k_2 - k_1} \\ k_1 \geq 1 \quad &: \gamma = k_1 \quad \phi = 1 \end{aligned} \tag{2.9.60}$$

De eerste van deze drie is optimaal in dezelfde zin als de strategie in het voorgaande punt. De laatste strategie verschilt slechts in de waarde van ϕ in het geval dat $k_1 \leq 1 \leq k_2$. Het resultaat van de numerieke experimenten was dat geen significante verschillen werden gevonden in het convergentiegedrag bij gebruik van de verschillende strategieën.

2.9.20. Naast de numerieke experimenten voor het vergelijken van verschillende parameter-combinaties werden ook numerieke experimenten uitgevoerd ter vergelijking met andere quasi-Newton-methoden. Bovendien werd geëxperimenteerd met verschillende gradaties van onnauwkeurige lijnminimalisering. De conclusies van deze tests waren dat de zelfschalende-variabelemetriek algorithmen duidelijk te prefereren zijn boven andere quasi-Newton algorithmen bij problemen met grotere aantallen variabelen en, in de tweede plaats, dat het achterwege laten van nauwkeurige lijnminimalisering de efficiëntie van de SSVM algorithmen (in termen van functie evaluaties) aanzienlijk groter maakt.

Numeriek stabielere quasi-Newton algorithmen

2.9.21. Een geheel anders gerichte recente ontwikkeling op het gebied van quasi-Newton-methoden werd in 1972 geïnitieerd door Gill en Murray in [2.9.5]. Ter verbetering van de numerieke stabiliteit suggereerden deze namelijk om in de standaard quasi-Newton algorithme (vgl. pt. 2.7.2) de gebruikelijke bepaling van de zoekrichting met behulp van het matrix-vector product

$$d^{(k)} := -H^{(k)} g^{(k)}$$

te vervangen door een bepaling van dezelfde zoekrichting als oplossing van de vergelijking

$$B^{(k)} d^{(k)} = -g^{(k)} \qquad (2.9.61)$$

waar $B^{(k)}$ een benadering van de Hessiaan zelf voorstelt in plaats van een benadering van de inverse Hessiaan

$$B^{(k)} = [H^{(k)}]^{-1} \qquad (2.9.62)$$

In plaats van aanpassingsformules voor de H-matrices dient dan gebruik gemaakt te worden van de duale formuleringen van deze (vgl. pt. 2.8.14) als aanpassingsformules voor de B-matrices. Gill en Murray [2.9.5] toonden aan dat deze aanpassingsformules in nagenoeg alle gevallen geschreven kunnen worden in de vorm

$$B^* := B + \pi_1 z z^T + \pi_2 w w^T \tag{2.9.63}$$

Een grotere numerieke stabiliteit kan worden verkregen door de noodzakelijke oplossing van de lineaire vergelijking (2.9.61) uit te voeren met behulp van matrix-decompositie technieken, zoals die van Choleski (vgl. pt. 2.5.19) : Na decompositie van de (positief definiete) matrix B als een product van een onderdriehoeksmatrix L, een diagonaal matrix D en de bovendriehoeksmatrix L^T

$$B = LDL^T \tag{2.9.64}$$

kan de lineaire vergelijking (2.9.61)

$$Bd = -g$$

eenvoudig worden opgelost door successievelijke oplossing van de twee driehoekstelsels

$$Lv = -g \tag{2.9.65}$$

en

$$L^T d = D^{-1} v \tag{2.9.66}$$

2.9.22. In de practijk is het niet noodzakelijk om in iedere stap een matrix decompositie uit te voeren. Gill en Murray [2.9.5] toonden aan dat de bijzondere vorm (2.9.63) van de aanpassingsformule, die kan worden opgevat als het resultaat van twee successievelijke rang-één correcties

$$\tilde{L}\tilde{D}\tilde{L}^T := LDL^T + \pi_1 z z^T \tag{2.9.67}$$

en

$$L^* D^* L^{*T} := \tilde{L}\tilde{D}\tilde{L}^T + \pi_2 w w^T \tag{2.9.68}$$

het mogelijk maakt om in plaats van de matrix B direct de factoren D en L aan te passen. Op die manier wordt gerealiseerd dat zonder extra rekenwerk een aanzienlijke verhoging van de nauwkeurigheid en een

betere numerieke controle van het rekenproces in quasi-Newton algorithmen wordt bereikt.

2.9.23. Naast het voordeel van een grotere numerieke stabiliteit heeft de alternatieve implementatie van de quasi-Newton algorithmen van Gill en Murray ook voordelen in het geval van het gebruik van quasi-Newton algorithmen voor minimaliseringsproblemen met beperkingen. Bij deze problemen is namelijk de kennis van een benadering van de Hessiaan van groter belang dan de kennis van de benadering van de inverse Hessiaan. Hierop wordt in het volgende hoofdstuk nader ingegaan. Een hiermee samenhangend voordeel van deze alternatieve implementatie van de quasi-Newton algorithmen is dat in de daarop gebaseerde rekenmachine programma's gebruik gemaakt kan worden van dezelfde (geavanceerde) lineaire-algebra-procedures als die welke hun toepassing met nevenvoorwaarden en interessant genoeg ook programma's voor lineaire programmering.

Quasi-Newton algorithmen met numerieke afgeleiden

2.9.24. Alle hierboven besproken quasi-Newton algorithmen gaan uit van het bekend zijn van de gradiënt van de te minimaliseren functie. In de practijk betekent dit dat deze gradiënten of berekend moeten worden met behulp van analytische uitdrukkingen of benaderd moeten worden met behulp van numerieke differentiatie. Teneinde de numerieke fouten in dit laatste geval zo klein mogelijk te houden werd door Stewart [2.9.4] een speciale versie van de Davidon-Fletcher-Powell algorithme ontwikkeld, waarin gebruik wordt gemaakt van de aanwezige informatie in de matrix H over de Hessiaan voor de bepaling van een "optimale" stapgrootte voor numerieke differentiatie, waarbij "optimaal" geïnterpreteerd wordt als het beste compromis m.b.t. de afrondings- en afkapfouten bij numerieke differentiatie. In de practijk blijkt dat het bij het merendeel van de "gewone" optimaliseringsproblemen niet noodzakelijk is om dit soort ingewikkelde procedures voor de bepaling van de stapgrootte voor numerieke differentiatie toe te passen. Een met enig inzicht gekozen vaste stapgrootte voldoet in de meeste gevallen minstens even goed en is volgens diverse auteurs (vgl. [2.9.2] p.116) zelfs te prefereren boven een mogelijk verkeerd aangepaste stapgrootte als die van Stewart. In de practijk blijkt het convergentiegedrag van de quasi-Newton algorithmen nauwelijks beïnvloed te worden door het gebruik van numerieke afgeleiden in plaats uit analytische uitdrukkingen berekende afgeleiden.

Literatuur

2.9.25. Meer details over de in deze paragraaf besproken ontwikkelingen kunnen worden aangetroffen in de volgende publikaties.

[2.9.1] : Zie [1.1.1] Luenberger (1973).

[2.9.2] : Zie [1.1.3] Murray (1972).

[2.9.3] : Zie [2.7.6] Fletcher (1970).

[2.9.4] : Zie [2.6.9] Stewart (1967).

[2.9.5] : Zie [2.8.5] Gill and Murray (1972).

[2.9.6] : Zie [2.5.15] Stoer (1972).

[2.9.7] : Gantmacher, F.R.: The theory of matrices, Vol. I, Chelsea Publ. Cy, New York (1959).

[2.9.8] : Himmelblau, D.M.: Applied nonlinear programming, McGraw-Hill New York (1972).

[2.9.9] : Loewner, C.: Ueber monotore Matrix Funktionen, Math. Zeitschr. 38 (1934) pp. 177-216.

[2.9.10]: Oren, S.S.: Self-scaling variable metric algorithms without line search for unconstrained optimization. Maths Comp. 27 (1973) pp. 873-885.

[2.9.11]: Oren, S.S. en Luenberger, D.G.: Self-scaling variable metric (SSVM) algorithms. Part I and II,Management Science, 20 (1974) pp. 845-874.

[2.9.12]: Oren, S.S. en Spedicato, E.: Optimal condtioning of self-scaling variable metric algorithms. Math. Progr. 10 (1976) pp. 70-90.

[2.9.13]: Wilkinson, J.H.: The algebraic eigenvalue problem. Oxford University Press, London (1965)

§ 2.10. Minimaliseren van sommen van kwadraten

2.10.1. In de practijk van vele rekencentra blijkt dat een zeer groot deel van de aangeboden onbeperkte minimaliseringsproblemen voortkomt uit toepassingen van "kleinste-kwadraten"-procedures voor de oplossing van curve-fitting- en parameterschattingsproblemen en uit toepassingen van dezelfde procedures voor de oplossing van(mogelijk overgedetermineerde) stelsels niet-lineaire vergelijkingen. In het eerste geval, d.i. bij curve-fittingen en parameterschattingsproblemen leiden deze kleinste-kwadraten procedures tot minimaliseringsproblemen van de vorm (1.1.6)

$$\min\{ \sum_{t=1}^{m} [(y_t - m_t(x))/\sigma_t]^2 \mid x \in \mathbb{R}^n\} \qquad (2.10.1)$$

waarin

y_t : geobserveerde waarde van een grootheid
$m_t(x)$: theoretische waarde van dezelfde grootheid in afhankelijkheid van de vector van parameters x
σ_t : gewichtsfactor voor de t-de waarneming

Definitie van een vectorfunctie $f : \mathbb{R}^n \to \mathbb{R}^m$ met als componenten

$$f_t(x) := (y_t - m_t(x))/\sigma_t \qquad t = 1, \ldots, m \qquad (2.10.2)$$

maakt het mogelijk de objectfunctie van het minimaliseringsprobleem (2.10.1)

$$F(x) := \sum_{t=1}^{m} [(y_t - m_t(x))/\sigma_t]^2$$

te schrijven in de vorm

$$F(x) := \sum_{t=1}^{m} f_t^{\,2}(x) = f^T(x)f(x) = \|f(x)\|^2 \qquad (2.10.3)$$

De componenten $f_t(x)$ worden vaak aangeduid als de residuen; de objectfunctie $F(x)$ heet dan de som van de kwadraten van de residuen. De standaardvorm van het niet-lineair kleinste-kwadratenprobleem zoals dat in deze para-

graaf zal worden behandeld wordt daar mee

$$\min \{f^T(x)f(x) \mid x \in \mathbb{R}^n\} \tag{2.10.4}$$

Deze standaardformulering is dezelfde als die welke resulteert bij de toepassing van kleinste-kwadraten-procedures voor de oplossing van stelsels niet-lineaire vergelijkingen van de vorm

$$f_t(x) = 0 \qquad t = 1, \ldots, m \tag{2.10.5}$$

In dit geval wordt de te minimaliseren objectfunctie

$$F(x) := \sum_{t=1}^{m} f_t^2(x) = f^T(x)f(x) = \| f(x) \|^2$$

en het probleem juist gelijk aan probleem (2.10.4). Bij deze laatste toepassing zowel als bij de bovengenoemde curve-fitting-problemen is het gebruikelijk dat $m \geq n$. Hiervan zal in deze paragraaf steeds worden uitgegaan.

2.10.2. In principe kunnen voor het minimaliseren van de kwadraatsom (2.10.4) de algemene minimaliseringsmethoden uit de voorgaande paragrafen worden gebruikt. Van belang in dat geval is in de eerste plaats de <u>gradiënt</u> van de kwadraatsom

$$\nabla F(x) := 2J^T(x)f(x) \tag{2.10.6}$$

waarin $J(x)$ de <u>functionaalmatrix</u> of <u>Jacobiaan</u> is van de vectorfunctie $f(x)$, d.w.z. de $m \times n$-matrix met als elementen

$$J_{tj}(x) := \frac{\partial f_t}{\partial x_j}(x) \tag{2.10.7}$$

en in de tweede plaats de <u>Hessiaan</u> van de kwadraatsom

$$\nabla^2 F(x) := G(x) := 2\left[\left(\frac{\partial J^T}{\partial x}(x)\right)f(x)\right] + 2J^T(x)J(x) \tag{2.10.8}$$

waar $\left(\frac{\partial J}{\partial x}(x)\right)$ een $m \times n \times n$-matrix voorstelt met als elementen de tweede afgeleiden van de componenten van de vectorfunctie f_t

$$\left(\frac{\partial J}{\partial x}(x)\right)^{t}_{ij} = \frac{\partial^2 f_t}{\partial x_i \partial x_j}(x) \qquad (2.10.9)$$

en $[\frac{\partial J}{\partial x}^{T}(x)f(x)]$ een n x n matrix met als elementen

$$[\frac{\partial J}{\partial x}^{T}(x)f(x)]_{ij} = \sum_{t=1}^{m} \frac{\partial^2 f_t}{\partial x_i \partial x_j}(x) f_t(x) \qquad (2.10.10)$$

In het geval dat f(x) klein is (d.i. $f(x) \approx 0$) en/of in het geval dat een groot deel van de component-functies $f_t(x)$ (bijna) lineair zijn, zijn de elementen van deze laatste matrix eveneens klein en geldt dat de Hessiaan van de kwadraatsom redelijk kan worden benaderd door de uitsluitend op eerste afgeleiden gebaseerde matrix

$$B(x) := 2J^{T}(x)J(x) \approx \nabla^2 F(x) \qquad (2.10.11)$$

Op deze benadering zijn vrijwel alle hierna te bespreken speciale methoden voor de minimalisering van kwadraatsommen gebaseerd.

Methode van Gauss-Newton

2.10.3. Zoals opgemerkt door diverse auteurs (vgl. [2.10.5],[2.10.12] en [2.10.15] blijken algemene minimaliseringsmethoden bij het minimaliseren van kwadraatsommen vaak minder efficiënt dan de speciale kwadraatsomminimaliseringsmethoden die gebruik maken van de speciale vorm van de objectfunctie F(x). Dit laatste betekent in de meeste gevallen dat op een of andere manier gebruik wordt gemaakt van de mogelijkheid om de Hessiaan G(x) (2.10.8) van de objectfunctie te benaderen door de matrix B(x) (2.10.11). De basismethode in dit verband is de methode waarbij gebruik gemaakt wordt van dezelfde algorithme als bij de methode van Newton (pt. 2.5.4) met dat verschil dat de benadering B(x) de plaats inneemt van de Hessiaan G(x). Deze basis methode werd in zijn originele vorm (d.i. zonder lijnminimalisering) voor het eerst toegepast door Friedrich Gauss (1777-1855) en staat om die reden bekend als de methode van Gauss-Newton. Een andere naam ervoor is om hierna in pt. 2.10.4 te bespreken redenen de gegeneraliseerde kleinste -kwadratenmethode. De algorithme van de methode heeft de volgende standaardvorm (vgl. pt. 2.5.4)

Algorithme van de Methode van Gauss-Newton

(0) zet $x^{(0)}$:= gegeven startpunt en k: = 0;

(i) bepaal de residuvector $f^{(k)} := f(x^{(k)})$, de Jacobiaan $J^{(k)} := J(x^{(k)})$ en de gradiënt (2.10.6)

$$\nabla F(x^{(k)}) := 2J^{(k)T} f^{(k)}$$

(ii) ga na of $x^{(k)}$ optimaal is; zo ja dan klaar; zo nee, dan

(iii) bepaal een zoekrichting $d^{(k)}$ uit

$$B(x^{(k)})d^{(k)} = -\nabla F(x^{(k)})$$

of equivalent uit

$$J^{(k)T}J^{(k)}d^{(k)} = -J^{(k)T}f^{(k)} \qquad (2.10.12)$$

(iv) bepaal een staplengte(-factor) $\alpha^{(k)}$; Bijvoorbeeld als

$$\alpha^{(k)} := 1$$

of zo dat

$$F(x^{(k)} + \alpha^{(k)}d^{(k)}) < F(x^{(k)})$$

of zo dat

$$F(x^{(k)} + \alpha^{(k)}d^{(k)}) = \min \{F(x^{(k)} + \alpha d^{(k)}) \mid \alpha \in \mathbb{R}_+^1\}$$

(v) zet $x^{(k+1)} := x^{(k)} + \alpha^{(k)}d^{(k)}$, k:= k + 1 en ga terug naar stap (i).

2.10.4. Een tweede, veel gebruikte aanpak om dezelfde methode van Gauss-Newton af te leiden heeft als uitgangspunt de locale benadering in de (k+1)-de iteratiestap van de functies $f_t(x)$ door lineaire functies $\ell_t^{(k)}(x)$ van de vorm

$$\ell_t^{(k)}(x) := f_t(x^{(k)}) + \sum_{j=1}^{n} J_{tj}^{(k)}(x_j - x_j^{(k)}) \qquad t = 1, \ldots m$$

waarin (vgl. (2.10.7)) (2.10.13)

$$J_{tj}^{(k)} := \frac{\partial f_t}{\partial x_j}(x^{(k)})$$

Als benadering voor F(x) volgt daarmee

$$F(x) \approx \sum_{t=1}^{m} [\ell_t^{(k)}(x)]^2 =: \phi^{(k)}(x) \qquad (2.10.14)$$

waarin voor $\phi^{(k)}(x)$ met de definitie

$$d := x - x^{(k)} \qquad (2.10.15)$$

in vector notatie geschreven kan worden

$$\begin{aligned} \phi^{(k)}(x) &:= \| \ell^{(k)}(x) \|^2 \\ &:= \| f^{(k)} + J^{(k)} d \|^2 \\ &:= f^{(k)T} f^{(k)} + 2 f^{(k)T} J^{(k)} d + d^T J^{(k)T} J^{(k)} d \end{aligned} \qquad (2.10.16)$$

Een betere benadering $x^{(k+1)}$ voor de oplossing van het niet-lineaire kleinste kwadratenprobleem in de (k+1)-de iteratiestap kan nu worden gevonden door de oplossing $d^{(k)}$ te bepalen van het lineaire kleinste-kwadraten-probleem

$$\min\{ \| f^{(k)} + J^{(k)} d \|^2 \mid d \in \mathbb{R}^n \} \qquad (2.10.17)$$

en daarna $x^{(k+1)}$ gelijk te stellen aan

$$x^{(k+1)} := x^{(k)} + \alpha^{(k)} d^{(k)} \qquad (2.10.18)$$

2.10.5. Voor de oplossing van het lineaire kleinste-kwadraten-probleem bestaan een aantal verschillende methoden, (zie [2.10.9]) waarvan de twee meest bekende zijn die welke gebruik maakt van de z.g. normaalvergelijkingen en die welke gebruik maakt van orthogonale transformaties. Bij de eerste methode wordt $d^{(k)}$ bepaald uit de oplossing van de vergelijkingen die ontstaan door het nulstellen van de gradiënt (t.o.v. d) van de objectfunctie in (2.10.17). Deze vergelijkingen die bekend zijn onder de naam normaalvergelijkingen zijn voor het lineair kleinste-kwadraten-probleem (2.10.17) juist gelijk aan de eerder gevonden vergelijkingen (2.10.12)

$$J^{(k)T}J^{(k)}d = -J^{(k)T}f^{(k)}$$

Deze aanpak met behulp van de normaalvergelijkingen blijkt dus equivalent met de in het voorgaande punt besproken aanpak gebaseerd op de benadering van de methode van Newton. Voor de actuele oplossing van de normaalvergelijkingen kan gebruik worden gemaakt van de in pt. 2.5.19 besproken methode van Choleski.

2.10.6. De tweede veel toegepaste methode om de oplossing $d^{(k)}$ van het lineaire kleinste-kwadraten-probleem (2.10.17) te bepalen is gebaseerd op het gebruik van orthogonale transformaties met behulp van QR-decompositie (zie [2.10.9]). Het idee van deze oplosmethode is dat als Q een orthogonale m x m matrix ($Q^TQ = I$) is dan geldt (bij weglating van de indices (k) in (2.10.16)) dat

$$\| f + Jd \|^2 = \| Q^Tf + Q^TJd \|^2 \qquad (2.10.19)$$

In het geval dat (in de veronderstelling dat $m \geq n$ en J volle rang heeft) er matrices Q en R bepaald kunnen worden zodanig dat

$$J = QR = [Q_1 \mid Q_2] \begin{bmatrix} R \\ 0 \end{bmatrix} \begin{matrix} \}n \\ \}m-n \end{matrix} \qquad (2.10.20)$$

waarin R een bovendriehoeksmatrix van rang n is,

$$R = \left\{ \begin{array}{|c} \diagdown \\ \hline \end{array} \right\} \qquad (2.10.21)$$

dan geeft uitwerken van (2.10.19)

$$\| Q^Tf + Q^TJd \|^2 = \left\| \begin{pmatrix} Q_1^Tf \\ Q_2^Tf \end{pmatrix} + \begin{pmatrix} Rd \\ 0 \end{pmatrix} \right\|^2 \qquad (2.10.22)$$

$$= \| Q_1^Tf + Rd \|^2 + \| Q_2^Tf \|^2$$

De oplossing d die de lineaire kwadratensom (2.10.19) minimaliseert kan dan bepaald worden als oplossing van het driehoekstelsel

$$Rd = -Q_1^T f \tag{2.10.23}$$

2.10.7. Zodra de matrices Q en R zijn bepaald - en dat blijkt in de practijk relatief eenvoudig te kunnen worden uitgevoerd met behulp van z.g. Householder-transformaties (zie [2.10.9]) - is het oplossen van deze laatste vergelijking bijzonder eenvoudig. Het resultaat is een oplosmethode die numeriek stabieler is dan die welke gebruik maakt van de normaalvergelijkingen en wel in het geval dat $m \approx n$ met ongeveer dezelfde hoeveelheid rekenwerk en in het geval dat $m >> n$ met ongeveer de dubbele hoeveelheid rekenwerk (Het aantal vermenigvuldigingen is (zie [2.10.6] p. 37) bij gebruik van de QR-decompositie van de orde $mn^2 - 1/3n^3 + O(n^2)$ en bij gebruik van de normaalvergelijkingen en Choleski-decompositie van de orde $\frac{1}{2} mn^2 + \frac{1}{6} n^3 + O(n^2)$). Beide besproken methoden worden bij practische toepassingen van de Gauss-Newton algorithme gebruikt.

2.10.8. Gewoonlijk leidt de methode van Gauss-Newton als het probleem goed geconditioneerd is snel tot een oplossing. Bij benadering zal de convergentie juist als bij de methode van Newton zelf kwadratisch zijn. De methode heeft echter ook een aantal nadelen, t.w.:

(1) Aangezien de fout in de benadering van $G^{(k)}$ door $B^{(k)}$ (vgl. pt. 2.10.2) in de omgeving van de oplossing x^* afhankelijk is van de waarde van $F(x)$ in x^* is de convergentie-snelheid kleiner naarmate $F(x^*)$ groter is

(2) Indien de semi-positief definiete matrix $J^{(k)T}J^{(k)}$ singulier is, is het niet mogelijk de zoekrichting $d^{(k)}$ volgens de hiervoor gegeven voorschriften te genereren. Een voor de hand liggend alternatief is om in dat geval verder te zoeken langs de negatieve gradiënt, d.i.

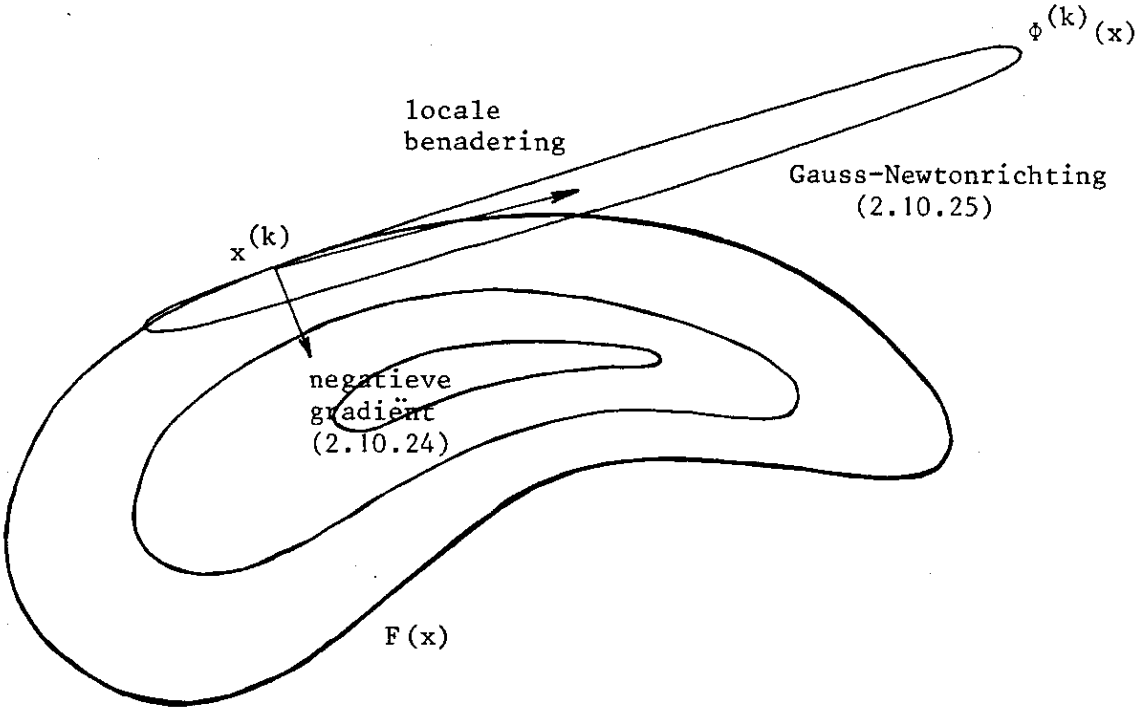
$$d^{(k)} := -2J^{(k)T}f^{(k)} \tag{2.10.24}$$

(3) In de practijk komt het nogal eens voor dat de volgens de Gauss-Newton algorithme regulier gegenereerde zoekrichtingen

$$d^{(k)} := -[J^{(k)T}J^{(k)}]^{-1}J^{(k)T}f^{(k)} \tag{2.10.25}$$

vrijwel loodrecht staan op de gradiënten (2.10.24). De methode convergeert dan slecht.

2.10.9. Een illustratie van het in dit laatste punt genoemde nadeel wordt gegeven in Figuur 2.10.9 waar de situatie is geschetst waar een algemene functie locaal wordt benaderd door een (minder goed geconditioneerde) kwadratische vorm



Figuur 2.10.9. Locale benadering van een functie door een positief definiete kwadratische vorm.

In deze figuur kan men zien dat de zoekrichting van Gauss-Newton bijna evenwijdig loopt aan de hoogtelijnen, dus ongeveer loodrecht op de negatieve gradiënt. In de practijk is het geen uitzondering dat door de

slechte locale benadering en/of door afrondfouten de met de methode van Gauss-Newton gegenereerde zoekrichting zodanig wordt, dat langs deze richting helemaal geen punten met functiewaardeverlaging kan worden gevonden.

De methoden die verder in deze paragraaf worden behandeld zijn alle gebaseerd op de methode van Gauss-Newton, doch proberen de genoemde nadelen van deze methode op te heffen door de component van de zoekrichting in de richting van de negatieve gradiënt te vergroten.

Methode van Marquardt

2.10.10. Ter verbetering van de in de voorgaande punten besproken, soms slechte convergentiegedrag van de Gauss-Newton algorithme suggereerde Marquardt [2.10.11] een practische algorithme gebaseerd op het reeds eerder (pt. 2.5.14) genoemde idee van Levenberg [2.10.3]. Deze algorithme bestaat (in analogie met de in pt. 2.5.14 besproken, doch historisch later ontwikkelde algorithme van Goldfeld, Quandt en Trotter)daaruit dat de zoekrichting $d^{(k)}$ in plaats van uit (2.10.12) bepaald wordt uit de vergelijking

$$[J^{(k)T}J^{(k)} + \mu^{(k)}I]\, d^{(k)} = -J^{(k)T}f^{(k)}, \qquad (2.10.26)$$

waarin $\mu^{(k)}$ een nog te kiezen niet-negatieve parameter is, en dat afgezien wordt van lijnminimalisering in die zin dat in iedere stap gekozen wordt voor de stapgrootte(-factor)

$$\alpha^{(k)} := 1 \qquad (2.10.27)$$

Afhankelijk of wel of niet voldaan wordt aan de eis van een strikte functiewaardevermindering

$$F(x^{(k)}) - F(x^{(k)} + d^{(k)}) > \varepsilon \qquad (2.10.28)$$

wordt de parameter $\mu^{(k)}$ verkleind voor de volgende iteratiestap

$$\mu^{(k+1)} := \mu^{(k)}/\upsilon \qquad \upsilon > 1 \qquad (2.10.29)$$

dan wel wordt $\mu^{(k)}$ vergroot

$$\mu^{(k)} := \mu^{(k)} \qquad \nu > 1 \tag{2.10.30}$$

en wordt opnieuw nagegaan of met deze grotere parameter waarde wel voldaan wordt aan de voorwaarde (2.10.28).

2.10.11. De uit vergelijking (2.10.26) bepaalde zoekrichting (of beter correctie) $d^{(k)}(\mu)$ heeft beschouwd als functie van $\mu \geq 0$ de volgende eigenschappen (vgl. [2.10.11] en pt. 2.5.14):

(i) $d^{(k)}(\mu)$ is de oplossing van het onbeperkte minimaliseringsprobleem

$$\min \{ \| J^{(k)} d + f^{(k)} \|^2 + \tfrac{1}{2} \mu \| d \|^2 \;\Big|\; d \in \mathbb{R}^n \} \tag{2.10.31}$$

(ii) $d^{(k)}(\mu) := x^{(k+1)}(\mu) - x^{(k)}$ is oplossing van het beperkt minimaliseringsprobleem

$$\min \{ \| J^{(k)}(x - x^{(k)}) + f^{(k)} \| \;\Big|\; \| x - x^{(k)} \| \leq \| d^{(k)}(\mu) \| \} \tag{2.10.32}$$

d.w.z. $d^{(k)}(\mu)$ minimaliseert de gelineariseerde objectfunctie

$$\Phi^{(k)}(x) = \| J^{(k)}(x - x^{(k)}) + f^{(k)} \|^2 \tag{2.10.33}$$

over de hyperbol

$$\{ x \;\Big|\; \| x - x^{(k)} \| \leq \| [J^{(k)T} J^{(k)} + \mu I]^{-1} J^{(k)T} f^{(k)} \| \} \tag{2.10.34}$$

(iii) $\| d^{(k)}(\mu) \|$ is een monotoon dalende functie van μ en

$$\lim_{\mu \to \infty} \| d^{(k)}(\mu) \| = 0 \tag{2.10.35}$$

(iv) de hoek $\gamma^{(k)}(\mu)$ tussen $d^{(k)}(\mu)$ en de negatieve gradiënt richting (2.10.24) is eveneens een monotoon dalende functie van μ en

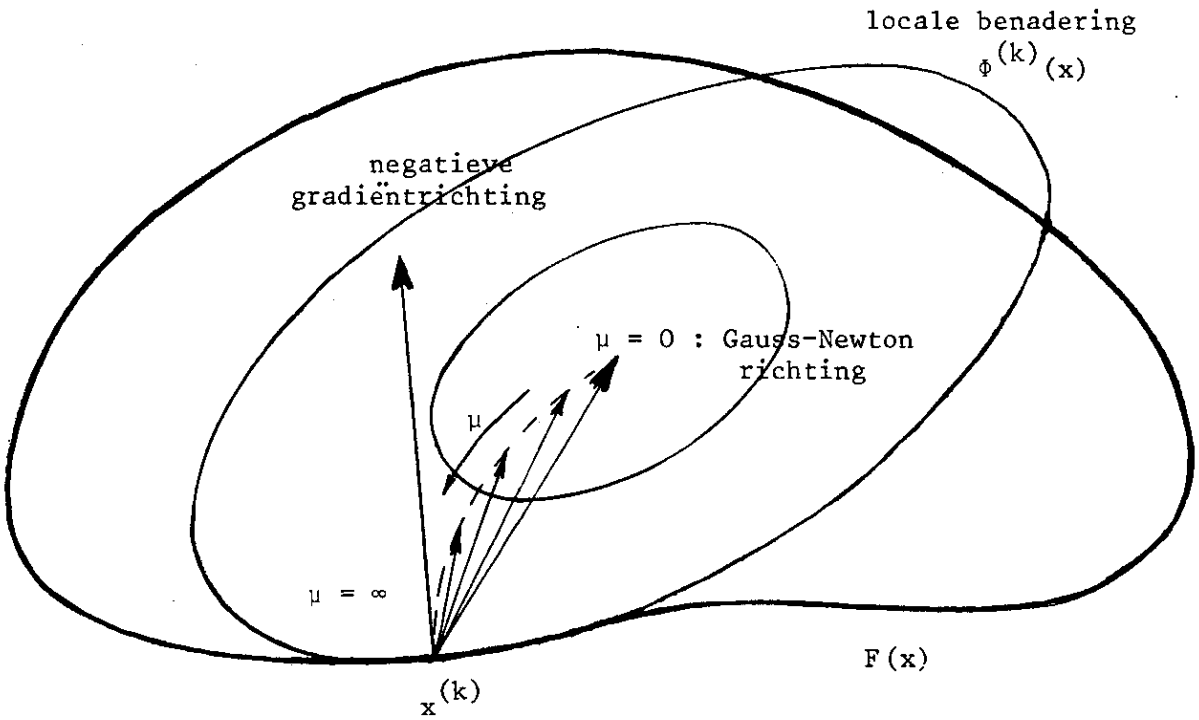
$$\lim_{\mu \to \infty} \gamma^{(k)}(\mu) = 0 \tag{2.10.36}$$

2.10.12. Afhankelijk van de grootte van de parameter $\mu^{(k)}$ gelden de volgende tendenzen voor de uit (2.10.26) bepaalde zoekrichting $d^{(k)}$ van de methode van Marquardt :

$$\mu^{(k)} \to 0 \; : \; d^{(k)} \to \text{Gauss-Newton richting (2.10.25)}$$

$$\mu^{(k)} \to \infty \; : \; d^{(k)} \to \text{negatieve gradiënt (2.10.24) en } |d^{(k)}| \to 0$$

De parameter $\mu^{(k)}$ interpoleert in zekere zin tussen deze twee uiterste correcties. Dit is geïllustreerd in Figuur 2.10.12.



Figuur 2.10.12. Zoekrichtingen voor de methode van Marquardt bij variërende waarden voor de parameter μ.

2.10.13. Als voordelen van de methode van Marquardt gelden dat deze methode ook kan worden toegepast als $J^{(k)T}J^{(k)}$ singulier is en dat in principe geen lijnminimalisering vereist is. De stapgrootte wordt immers geregeld door de parameter $\mu^{(k)}$. Als nadelen staan hier tegenover:

(i) Het kan voorkomen dat toevoeging van $\mu^{(k)}$ geen effect heeft, d.i. geen andere $d^{(k)}$ oplevert. De matrix $J^{(k)T}J^{(k)}$ zal daarom geschaald moeten worden, bijvoorbeeld zodanig dat de diagonaal elementen 1 worden om de toevoeging van $\mu^{(k)}$ zinvol te maken.

(ii) Bij het vergroten van $\mu^{(k)}$ binnen een iteratieslag moet telkens weer een nieuw stelsel vergelijkingen worden opgelost om $d^{(k)}$ te bepalen.

2.10.14. Een manier om dit laatste bezwaar op te heffen werd gesuggereerd door Jones in hetzelfde artikel [2.10.8] als waarin hij zijn, hierna te bespreken modificatie van de methode van Gauss-Newton beschrijft. De suggestie houdt in gebruik te maken van de spectrale decompositie van de positief semi definiete matrix $\bar{B}^{(k)} := J^{(k)T}J^{(k)}$ (vgl. pt. 2.5.16)

$$\bar{B}^{(k)} = J^{(k)T}J^{(k)} = P^{(k)}\Lambda^{(k)}P^{(k)T} = \sum_{j=1}^{n} \lambda_j^{(k)} p_j^{(k)} p_j^{(k)T} \qquad (2.10.37)$$

waar de λ_j en de p_j resp. de reële niet-negatieve eigenwaarden en de bijbehorende orthonormale eigenvectoren voorstellen. Voor willekeurige $\mu > 0$ volgt hier onmiddellijk mee dat

$$[B^{(k)} + \mu I]^{-1} = \sum_{j=1}^{n} \left(\frac{1}{\lambda_j^{(k)} + \mu}\right) p_j^{(k)} p_j^{(k)T} \qquad (2.10.38)$$

waarmee de zoekrichting $d^{(k)}(\mu)$ voor alle $\mu > 0$ wordt gegeven door

$$d^{(k)}(\mu) = -\sum_{j=1}^{n} \left(\frac{1}{\lambda_j^{(k)} + \mu}\right) \left(p_j^{(k)T} J^{(k)T} f^{(k)}\right) p_j^{(k)} \qquad (2.10.39)$$

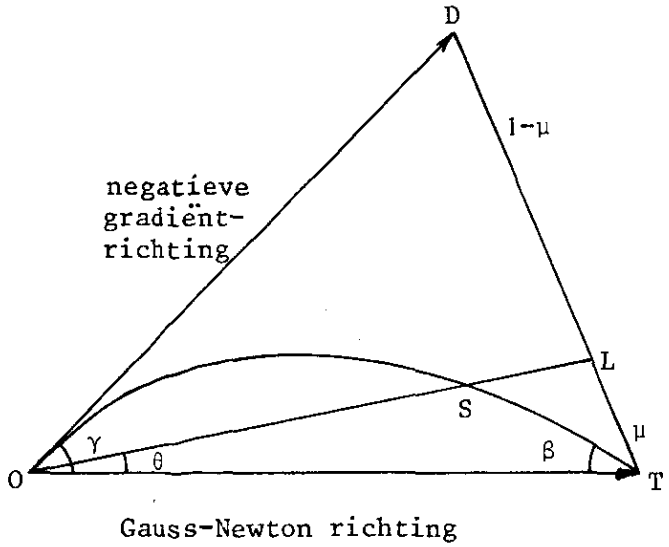
Een verandering van μ betekent in dit geval slechts een verandering van coëfficiënten. In de practijk wordt deze suggestie van Jones, die zeker bijdraagt tot het begrip van de methode van Marquardt, niet vaak toegepast aangezien de diagonalisatie nogal wat extra werk impliceert, terwijl

het geval waarbij $\mu^{(k)}$ binnen een iteratieslag een aantal keren moet worden vergroot zelden voorkomt.

2.10.15. Er bestaan diverse varianten van de methode van Marquardt. In de originele versie wordt voor μ de startwaarde $\mu^{(0)} = 0.01$ en voor ν de waarde $\nu = 10$ genomen. Bovendien wordt μ binnen een iteratieslag slechts zolang met de factor ν vergroot (vgl. (2.10.30)) totdat de hoek tussen de zoekrichting en de negatieve gradiënt kleiner is geworden dan 45°. In dat geval wordt overgestapt op lijnminimalisering langs de laatst bepaalde richting.

Methode van Jones of Spiral methode

2.10.16. Een andere, op dezelfde principes als de methode van Marquardt gebaseerde methode werd geïntroduceerd door Jones [2.10.8] onder de naam Spiral. Deze methode is in wezen niets meer dan een modificatie van de methode van Marquardt. Het voornaamste verschil met de methode van Marquardt bestaat daaruit dat binnen een iteratieslag niet telkens het stelsel vergelijkingen (2.10.26) moet worden opgelost om een nieuwe stap te bepalen, doch dat deze wordt bepaald door interpolatie.
De basisgedachte achter deze Spiral-algorithme is dat in de driehoek gevormd door de Gauss-Newton richting en de richting van de negatieve gradiënt (vgl. Figuur 2.10.16) vanuit het startpunt O altijd een punt gevonden kan worden met een lagere functiewaarde.



Figuur 2.10.16. Bepaling van de zoekrichting en stapgrootte in de methode Jones.

In Figuur 2.10.16 is OT de Gauss-Newton stap. D ligt in de richting van de negatieve gradiënt zodanig dat OD = OT. Aangezien er theoretisch functiewaardeverlaging langs OD moet optreden en de Gauss-Newton stap OT functiewaardeverlaging voorspelt, is het aannemelijk te veronderstellen dat een gedeelte van het gebied OTD bestaat uit punten waarin functiewaardeverlaging optreedt. Uitgaande van deze veronderstelling en van de strategie om bij het bepalen van een nieuw punt de staplengte zo groot mogelijk te maken, wordt allereerst het Gauss-Newton punt T beschouwd. Indien in dit punt geen functiewaardeverlaging optreedt dan is kennelijk de lineaire benadering niet correct voor het punt T. In dat geval worden nieuwe zoekrichtingen bepaald door middel van de punten L die op de lijn TD worden gegenereerd. Deze punten L worden zo gekozen dat L het lijnstuk TD verdeelt in stukken TL en LD die zich verhouden als μ: (1 - μ). De achtereenvolgende waarden van μ worden berekend uit de recurrente betrekking

$$\mu_1 \text{ gegeven } (0 < \mu_1 < 1)$$

$$\mu_{k+1} := \frac{2\mu_k}{1 + \mu_k} . \tag{2.10.40}$$

(Onder de voorwaarde $0 < \mu_1 < 1$ geldt $\lim_{k\to\infty} \mu_k = 1$.)

Voor de coördinaten (r_1,θ) van het punt L gelden de volgende (met de sinusregel te bewijzen) betrekkingen

$$\theta = \arctan \left(\frac{\mu \sin \gamma}{1 - \mu + \mu \cos \gamma}\right) \tag{2.10.41}$$

en

$$r_1 = \frac{r_0 \mu \sin \gamma}{\sin \theta} , \tag{2.10.42}$$

waarin γ de hoek is tussen OT en OD en r_0 de afstand OT (en OD).
De stapgrootte wordt vervolgens geregeld door de lijn OL te snijden met een spiraal door O en T die onder een hoek β vanuit T het gebied OTD binnenloopt en in O aan de lijn OD raakt. De vergelijking van deze spiraal in poolcoördinaten luidt

$$r = r_0(1 - \theta \cos \beta - (1 - \gamma \cos \beta)(\frac{\theta}{\gamma})^2). \qquad (2.10.43)$$

Hierin is r de afstand OS en de hoek β een nog te kiezen parameter (zie Figuur 2.10.16).
De coördinaten S_j van het punt S, uitgedrukt in de coördinaten van T en D met O als oorsprong, worden gegeven door de vergelijkingen

$$S_j = \frac{r}{r_1}\{\mu D_j + (1 - \mu)T_j\}, \; j = 1, \ldots, n. \qquad (2.10.44)$$

In de originele versie van Jones wordt $\beta = \gamma/2$ en $\mu_1 = 0.1$ genomen; er is echter niet aangegeven tot welke grens μ vergroot wordt. Bovendien worden binnen een iteratieslag vier spiralen afgezocht, welke gegenereerd worden door OT telkens te halveren. Indien mogelijk, wordt interpolatie toegepast waarbij de functiewaarde als functie van μ wordt beschouwd. Als daarna nog geen punt met functiewaardeverlaging is gevonden wordt verder gezocht langs de richting van de negatieve gradiënt (voor details zie Jones [2.10. 8]).

Methode van Fletcher en pseudo-inverse

2.10.17. Een derde modificatie van de methode van Gauss-Newton die een oplossing biedt voor de bepaling van de zoekrichting in het geval de matrix $B^{(k)} := 2J^{(k)T}J^{(k)}$ (bijna) singulier is, is de door Fletcher [2.10.7] geintroduceerde generalisatie van de methode van Gauss-Newton die gebruik maakt van het concept van de gegeneraliseerde-inverse of pseudo-inverse van een matrix (vgl.[2.10.4]). Dit concept past geheel in deze problematiek zoals volgt uit een van de mogelijke definities ervan.

Definitie 2.10.17 [2.10.2] : De n x m-matrix A^+ is de pseudo-inverse van de m x n-matrix A als voor alle $b \in \mathbb{R}^m$ geldt dat

$$x = A^+ b \qquad (2.10.45)$$

de minimum norm oplossing is van het lineair kleinste kwadraten probleem

$$\min\{\| Ax - b\|^2 \mid x \in \mathbb{R}^n\} \qquad (2.10.46)$$

2.10.18. Equivalent aan de Definitie 2.10.17 is de bekende definitie van de pseudo-inverse van Penrose.

Definitie 2.10.18 [2.10.4]: De n x m-matrix A^+ is "de" pseudo-inverse van de m x n matrix A indien A^+ voldoet aan de relaties

$$\begin{aligned} A^+AA^+ &= A^+ \qquad (A^+A)^T = A^+A \qquad (n \times n) \\ AA^+A &= A \qquad (AA^+)^T = AA^+ \qquad (m \times m) \end{aligned} \tag{2.10.47}$$

In het geval dat A van volle rang is geldt dat

$$A^+ = (A^TA)^{-1}A^T \tag{2.10.48}$$

Wordt in dat geval de QR-decompositie gegeven door

$$A = QR = [Q_1 \mid Q_2] \begin{bmatrix} R \\ O \end{bmatrix}$$

dan geldt ook

$$A^+ = R^{-1}Q_1^T \tag{2.10.49}$$

en volgt dat

$$A^+A = I_{n\times n} \tag{2.10.50}$$

en

$$AA^+ = Q_1Q_1^T \tag{2.10.51}$$

In het geval dat A niet van volle rang is bestaat er een orthogonale decompositie van de vorm (vgl. [2.10.8])

$$A = HRK^T = [H_1 \mid H_2] \begin{bmatrix} \Delta & O \\ O & O \end{bmatrix} \left[\begin{array}{c} K_1^T \\ \hline K_2^T \end{array}\right] \tag{2.10.52}$$

en geldt dat de minimum norm oplossing van (2.10.46) wordt gegeven door

$$x = K_1 R^{-1} H_1^T b \tag{2.10.53}$$

Hieruit volgt

$$A^+ = K_1 R^{-1} H_1^T \tag{2.10.54}$$

2.10.19. De laatste uitdrukking illustreert de omstandigheid dat de pseudo-inverse van een willekeurige matrix in theorie steeds kan worden bepaald. Hetzelfde geldt voor de practijk. Diverse technieken bestaan om de pseudo-inverse numeriek te berekenen in alle voorkomende gevallen . Voor de details daarvan moet echter worden verwezen naar de relevante literatuur [2.10.4]. Op deze plaats zij slechts een eenvoudige illustratie gegeven van de toepassing en een ad-hoc berekening van een pseudo-inverse.

Voorbeeld 2.10.19: Beschouw het probleem (Fig. 2.10.19b)

$$\min\{(4x_1 + 8x_2 - 5)^2 + (3x_1 + 6x_2)^2 \mid (x_1,x_2)^T \in \mathbb{R}^2\}$$

of equivalent (Fig. 2.10.19a)

$$\min\{\| Ax - b \|^2 \mid x \in \mathbb{R}^2\}$$

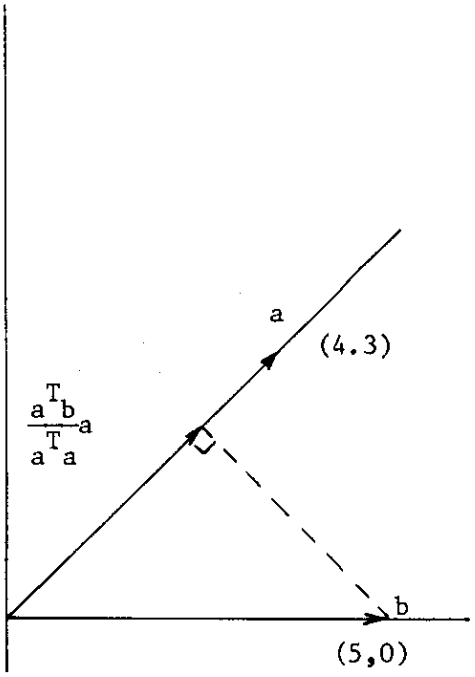
met

$$A = \begin{bmatrix} 4 & 8 \\ 3 & 6 \end{bmatrix} = \begin{bmatrix} a & 2a \\ & \end{bmatrix} \qquad b = \begin{bmatrix} 5 \\ 0 \end{bmatrix}$$

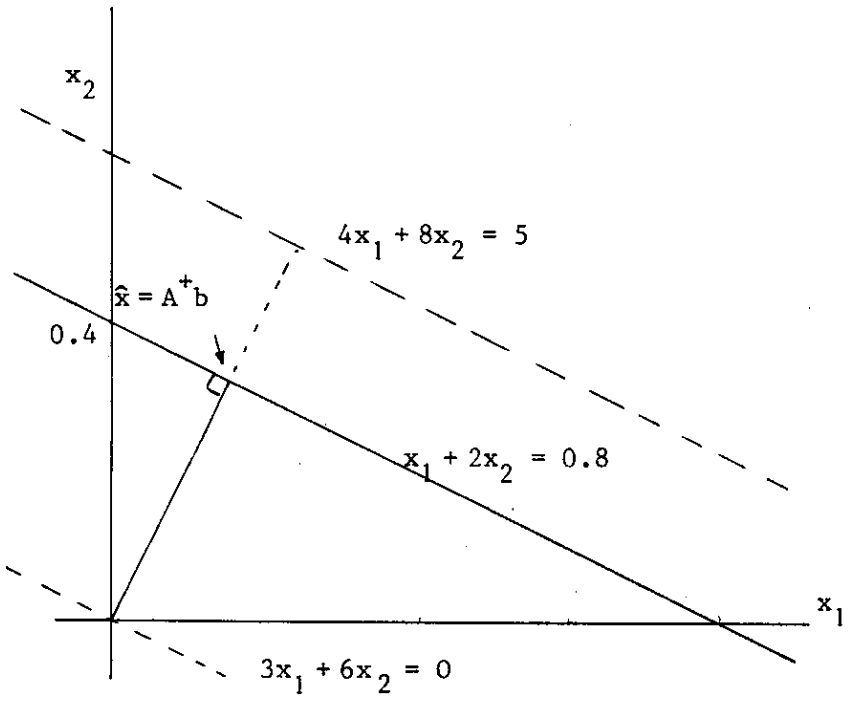
Voor de oplossing $\hat{x}$ van dit probleem moet gelden

$$A^T(A\hat{x} - b) = 0$$

d.w.z. $A\hat{x}$ is de projectie van b op de in dit geval afhankelijke kolommen a en 2a van A (Zie Figuur 2.10.19a). Dit betekent dat geldt



Figuur 2.10.19a : Schets van kleinste kwadratenprobleem in kolomruimte van A



Figuur 2.10.19b : Schets van kleinste kwadratenprobleem in x_1 - x_2-ruimte

$$A\hat{x} = (a)\hat{x}_1 + (2a)\hat{x}_2 = \frac{a^T b}{a^T a} a = \frac{20}{25} \begin{bmatrix} 4 \\ 3 \end{bmatrix}$$

ofwel

$$\hat{x}_1 + 2\hat{x}_2 = \frac{a^T b}{a^T a} = 0.8$$

De vector $\hat{x}$ met minimum norm die hieraan voldoet is de oplossing van het minimaliseringsprobleem

$$\min_{\hat{x}} \{\hat{x}_1^2 + \hat{x}_2^2 \mid \hat{x}_1 + 2\hat{x}_2 = \frac{a^T b}{a^T a}\}$$

waarvoor geldt

$$\hat{x}_1 = \frac{1}{5}\frac{a^T b}{a^T a} = 0.16$$

$$\hat{x}_2 = \frac{2}{5}\frac{a^T b}{a^T a} = 0.32$$

Deze oplossing is juist de vector $\hat{x} = A^+ b$ waaruit volgt dat

$$A^+ = \frac{1}{5\|a\|^2} A^T = \frac{1}{125}\begin{bmatrix} 4 & 3 \\ 8 & 6 \end{bmatrix}$$

Eenvoudig valt te verifiëren dat deze pseudo-inverse inderdaad aan de vergelijkingen van Penrose (2.10.47) voldoet.

Opm.: De niveaulijnen van de functie $F(x) = 16(x_1 + 2x_2 - 1.25)^2 + 9(x_1 + 2x_2)^2$ lopen in Figuur 2.10.19b evenwijdig aan de lijn $x_1 + 2x_2 = 0.8$. De vector $\hat{x} = A^+ b = [0.16, 0.32]^T$ wijst daarom juist in de richting van de (negatieve) gradiënt van deze object functie.

2.10.20. De methode van Fletcher [2.10.7] bestaat nu daaruit dat in stap (iii) van de Gauss-Newton algorithmen de zoekrichting wordt bepaald uit het voorschrift

$$d^{(k)} := - J^{(k)+} f(k) \qquad (2.10.55)$$

In het geval dat de Jacobiaan van volle rang is, is de methode exact hetzelfde als de methode van Gauss-Newton. In dat geval immers gaat (2.10.55) overeenkomstig (2.10.48) over in (2.10.12)

$$d^{(k)} := - [J^{(k)T} J^{(k)}]^{-1} J^{(k)T} f(k)$$

ofwel overeenkomstig (2.10.49) in (2.10.23)

$$d^{(k)} := -R^{(k)-1} Q_1^{(k)T} f(k)$$

In het geval dat $J^{(k)}$ niet van volle rang is, geeft de uitdrukking (2.10.55) een goed (vgl. opmerking in Vb 2.10.19) en duidelijk voorschrift voor een zoekrichting. Nodig in dit laatste geval is wel een procedure voor het bepalen van de pseudo-inverse.

2.10.21. In de practijk blijkt (vgl. [2.10.6]) de methode van Fletcher sterk afhankelijk van het criterium op grond waarvan beslist wordt wanneer een vector als lineair afhankelijk van andere vectoren moet worden beschouwd. Dit criterium speelt een belangrijke rol binnen de procedure om de pseudo-inverse te bepalen. Verandering van dit criterium kan een aanzienlijk verschil veroorzaken in de elementen van J^{+} aangezien deze elementen discontinu zijn bij de overgang van bijna singulier naar exact singulier zijn van de matrix J. Indien een te fijn criterium wordt gebruikt dan zal de methode van Fletcher practisch helemaal gelijk worden aan de methode van Gauss-Newton met als bezwaar dat als $J^{(k)T}J^{(k)}$ bijna singulier is de zoekrichting bijna loodrecht staat op de negatieve gradiënt (vgl. pt. 2.10.8 (2)). In het geval het criterium te grof is dan resulteert een methode die sterke overeenkomst vertoont met gradiënt methode met als gevolg een langzame convergentie van het iteratieproces (eerste orde convergentie).

Kwadraatsomminimalisering zonder afgeleiden

2.10.22. De methoden die tot dusver aan de orde zijn gekomen waren gebaseerd op het gebruik van in analytische vorm gegeven afgeleiden. In een aantal situaties zijn dergelijke analytische uitdrukkingen niet voorhanden en wordt het een probleem hoe deze afgeleiden te benaderen. De meest gebruikelijke manier hiervoor is om numerieke approximatie toe te passen door middel van de uitdrukking

$$J_{tj}(x) := \frac{\partial f_t(x)}{\partial x_j} \approx \frac{f_t(x + \eta e_j) - f_t(x)}{\eta} \quad , \; t = 1,\ldots,m; \; j = 1,\ldots,n \; . \tag{2.10.56}$$

Er zijn echter enkele bezwaren tegen deze approximatie aan te voeren, om welke redenen andere methoden werden ontwikkeld.

1) Het is onduidelijk hoe groot η moet worden gekozen.
2) Om de matrix J uit (2.10.56) te bepalen moeten de functiewaarden $f_t(x)$ $(t = 1,\ldots,m)$ in $(n + 1)$ verschillende punten x berekend worden. Dit impliceert vaak dat het grootste deel van de berekende functiewaarden gebruikt wordt voor het schatten van afgeleiden in plaats van rechtstreeks voor het minimaliseren van F(x).

Secant methode

2.10.23. De secant methode is een generalisatie van de gelijknamige methode voor het oplossen van één vergelijking met één onbekende en is gebaseerd op het volgende idee: Veronderstel dat we een iteratief algorithme hebben waarvan de laatst bepaalde (n + 1) punten zijn $x^{(k-n)}, x^{(k-n+1)}, \ldots, x^{(k)}$ met bijbehorende functiewaarden (vectoren) $f(x^{(k-n)}), f(x^{(k-n+1)}), \ldots, f(x^{(k)})$. We kunnen dan de m functies $f_t(x)$ benaderen door m lineaire functies $\ell_t^{(k)}(x)$ met ieder (n + 1) onbekende coëfficiënten. Voor ieder van deze functies $\ell_t(x)$ kunnen we de onbekende coëfficiënten vinden door oplossing van een stelsel van n + 1 lineaire vergelijkingen

$$\ell_t^{(k)}(x^{(j)}) = f_t(x^{(j)}), \quad j = k-n,\ldots,k \ . \tag{2.10.57}$$

De coëfficiënten van de functies $\ell_t^{(k)}(x)$ vormen juist de benaderingen voor de elementen van de matrix $J^{(k)}$, zodat we vervolgens de methode van Gauss-Newton (pt. 2.10.3) kunnen toepassen. Om deze algorithme te starten, indien alleen het punt $x^{(0)}$ door de gebruiker gegeven is, moeten eerst in n opvolgende punten $x^{(j)}$, $j = 1,\ldots,n$ de functiewaarden berekend worden. Een veel gebruikte procedure voor het kiezen van deze punten is het punt $x^{(j)}$ gelijk te nemen aan $x^{(j-1)}$ plus een kleine relatieve stap langs de j-de coördinaatrichting. Andere procedures zijn mogelijk even goed, mits, en dat geldt algemeen voor de toepassing van de secant methode, dat er voor moet worden gezorgd, dat de vectoren $s^{(j)} := x^{(j+1)} - x^{(j)}$, $j = k-n,\ldots,k-1$ lineair onafhankelijk blijven.

2.10.24. In de practijk wordt de hier geschetste methode, bestaande uit eerst het oplossen van de matrix $J^{(k)}$ uit de stelsels vergelijkingen (2.10.57) gevolgd door het toepassen van de Gauss-Newton algorithme, niet gebruikt. De hoeveelheid werk kan namelijk gereduceerd worden door definitie van de vectoren

$$\begin{aligned} s^{(j)} &:= x^{(j+1)} - x^{(j)} \\ e^{(j)} &:= f(x^{(j+1)}) - f(x^{(j)}) \end{aligned} \qquad j = k-n,\ldots,k-1 \ . \tag{2.10.58}$$

De stelsels vergelijkingen (2.10.57) impliceren nu dat voor willekeurige $z \in \mathbb{R}^n$ geldt

$$\ell^{(k)}(x^{(k)} + \sum_{i=1}^{n} z_i s^{(k-n+i-1)}) = f(x^{(k)}) + \sum_{i=1}^{n} z_i e^{(k-n+i-1)} \quad , \quad (2.10.59)$$

ofwel

$$\ell^{(k)}(x^{(k)} + S^{(k)}z) = f(x^{(k)}) + E^{(k)}z \quad , \quad (2.10.60)$$

waarin $s^{(k-n+i-1)}$ en $e^{(k-n+i-1)}$ de i-de kolom vormen van de matrices $S^{(k)}$ respectievelijk $E^{(k)}$. Uit deze laatste, als een coördinatentransformatie op te vatten uitdrukking volgt dat het lineaire kleinste-kwadratenprobleem

$$\min_z \| \ell^{(k)}(x^{(k)} + S^{(k)}z) \|^2 \quad (2.10.61)$$

dat ten grondslag ligt aan de methode van Gauss-Newton equivalent is met het kleinste kwadratenprobleem

$$\min_z \| f^{(k)} + E^{(k)}z \|^2 \quad (2.10.62)$$

De oplossing $z^{(k)}$ van dit laatste probleem volgt dan uit corresponderende normaalvergelijkingen

$$[E^{(k)T}E^{(k)}]z^{(k)} = -E^{(k)T}f^{(k)} \quad . \quad (2.10.63)$$

De zoekrichting voor de secant methode kan daarmee op zijn beurt worden gevonden uit

$$d^{(k)} := S^{(k)}z^{(k)} \quad (2.10.64)$$

Aangezien de matrices $E^{(k)}$ en $E^{(k+1)}$ slechts in één kolom van elkaar verschillen kan de hoeveelheid werk, nodig om $z^{(k)}$ uit (2.10.63) te bepalen, worden gereduceerd. Een methode hiervoor, welke gebaseerd is op het opbergen en bijwerken van de matrix $[E^{(k)T}E^{(k)}]^{-1}$, is beschreven door Rosen [2.10.16] (vgl. Appendix I van [2.10.6]).

2.10.25. Bij het toepassen van de secant methode moeten we aan twee voorwaarden voldoen, namelijk

1) Om $z^{(k)}$ uit (2.10.63) te kunnen bepalen moet de matrix $E^{(k)}$ van maximum rang n zijn, d.w.z. de vectoren $e^{(j)}$, j = k-n,...,k-1 moeten lineair

onafhankelijk zijn. Een probleem hierbij is dat deze vectoren $e^{(j)}$ (en ook de $s^{(j)}$)in de buurt van het minimum naar 0 convergeren. Daarom moeten voor het behoud van de numerieke onafhankelijkheid de kolommen van de matrix $E^{(k)}$ geschaald worden.

2) Zoals eerder opgemerkt (vgl. pt. 2.10.23) moeten ook de vectoren $s^{(j)}$, $j = k-n,\ldots k-1$ lineair onafhankelijk zijn. We kunnen dit ook als volgt inzien: Daar $s^{(k)} = x^{(k+1)} - x^{(k)} = \alpha^{(k)}d^{(k)} = S^{(k)}z^{(k)}$ is $s^{(k)}$ een lineaire combinatie van de vectoren $s^{(j)}$, $j = k-n,\ldots,k-1$ (de kolommen van $S^{(k)}$). Indien nu deze vectoren $s^{(j)}$ lineair afhankelijk worden blijft deze afhankelijkheid verder gelden. Daar $d^{(k)} := S^{(k)}z^{(k)}$ zoeken we in dit geval verder in een deelruimte van de $\mathbb{R}^n$ en convergeren naar het minimum in deze deelruimte dat in het algemeen niet zal samenvallen met het minimum in de gehele ruimte.

De methoden van Powell en Peckham, welke hierna worden beschreven, zijn gebaseerd op deze secant methode. Zij verschillen ervan doordat extra maatregelen worden getroffen om aan de hierboven genoemde voorwaarden te blijven voldoen.

Methode van Powell (1965) voor kwadraatsomminimalisering

2.10.26. Powell [2.10.14] propageerde in 1965 voor het eerst de secant methode met daarin aangebracht de volgende twee modificaties:

(1) Er wordt lijnminimalisering toegepast, dus $d^{(k)} := S^{(k)}z^{(k)}$ wordt niet gebruikt als de correctievector maar als zoekrichting.

(2) In de $(k+1)$-de iteratieslag wordt niet het paar $(s^{(k-n)}, e^{(k-n)})$ vervangen (d.i. het paar $(s^{(j)}, e^{(j)})$met de laagste waarde van j), doch het paar $(s^{(t)}, e^{(t)})$ waarvoor geldt

$$|z_t^{(k)}\{E^{(k)T}f(x^{(k)})\}_t| = \max_{1\le i\le n} |z_i^{(k)}\{E^{(k)T}f(x^{(k)})\}_i| \ , \qquad (2.10.65)$$

waar $\{E^{(k)T}f(x^{(k)})\}_i$ de i-de component is van de vector $E^{(k)T}f(x^{(k)})$.

Indien de matrix $E^{(k)}$ van maximum rang n is, geldt dat de matrix $[E^{(k)T}E^{(k)}]^{-1}$ positief definiet is. In dat geval geldt, aangezien

$$z^{(k)} := -[E^{(k)T}E^{(k)}]^{-1}E^{(k)T}f(x^{(k)}),$$

door de keuze (2.10.65) dat $z_t^{(k)} \neq 0$, tenzij $x^{(k)} = x^*$, waar $\nabla F(x^*) = 0$
Hierdoor is in te zien dat de kolommen van de matrix $S^{(k)}$ onafhankelijk blijven.

Methode van Peckham

2.10.27. Noch bij de originele secant methode, noch bij de methode van Powell is het mogelijk dat de matrix $E^{(k)}$ niet van maximum rang n is, zodat we dan $z^{(k)}$ niet kunnen bepalen uit (2.10.63). Om dit euvel te overkomen, ontwikkelde Peckham [2.10.13] een algorithme gebaseerd op het idee dat door het gebruik van meer dan (n + 1) punten voor het bepalen van de lineaire functies $\ell_t^{(k)}(x)$ in veel gevallen kan worden voorkomen dat de kolommen van de matrices $E^{(k)}$ en/of $S^{(k)}$ lineair afhankelijk worden. Peckham probeert zodoende voor (r + 1) punten $x^{(k-r)},\ldots,x^{(k)}$ aan de vergelijkingen (2.10.57) te voldoen waarbij $n \leq r \leq 3n$. Indien $r > n$ hebben we echter meer vergelijkingen dan onbekenden om de functies $\ell_t^{(k)}(x)$, $t = 1,\ldots,m$ uit (2.10.57) te bepalen. Daarom worden deze vergelijkingen opgelost in de zin van een gewogen lineair kleinste-kwadratenprobleem, waarbij aan punten die dichter in de buurt van het minimum liggen een groter gewicht wordt toegekend.
Hoewel de methode van Peckham in veel gevallen beter garandeert dat de kolommen van de matrix $S^{(k)}$ de gehele $\mathbb{R}^n$ opspannen, hetgeen noodzakelijk is om de functies $\ell_t^{(k)}(x)$ te kunnen bepalen, gaat ook deze methode mis als opvolgende punten $x^{(k)}$ op een lijn liggen. Peckham is zich bewust van deze moeilijkheid en gebruikt daarom een "pseudo-random number procedure" in het geval dat het stelsel (2.10.57) slecht geconditioneerd is.
Bij toepassing is de methode van Peckham dezelfde als de secant methode waarbij echter $z^{(k)}$ moet worden bepaald als oplossing van een gewogen lineair kleinste-kwadratenprobleem. Aangezien we $z^{(k)}$ nu niet meer kunnen bepalen uit (2.10.63), waarbij gebruik kan worden gemaakt van de methode van Rosen (vgl. pt. 2.10.24), zal de methode van Peckham meer werk per iteratieslag opleveren dan de methode van Powell of de secant methode. Peckham zelf komt in zijn artikel [2.10.13] tot minder functie-evaluaties dan voor de methode van Powell.

Practische suggesties

2.10.28. Als resultaat van een numeriek onderzoek kwam Eilers [2.10.6] tot de vol-

gende practische suggesties t.a.v. te gebruiken methoden voor het minimaliseren van niet-lineaire kwadraatsommen.

(1) De methode van Marquardt is voor algemene toepassing het meest geschikt.
(2) Voor goed-geconditioneerde problemen (de matrix J^TJ regulier) verdient de methode van Gauss-Newton de voorkeur. Hierbij kan in het geval dat $m \approx n$ het best gebruik worden gemaakt van orthogonale transformaties en in het geval dat $m >> n$ het best van de normaalvergelijkingen in combinatie met de Choleski-methode.
(3) De methode van Powell kan voor bepaalde soorten problemen aanzienlijk beter zijn dan de methoden van Marquardt of Gauss-Newton. Deze methode is echter minder algemeen toepasbaar.
(4) De methode van Fletcher is alleen aan te bevelen voor slecht-geconditioneerde problemen (de matrix J^TJ singulier).
(5) De methode van Jones (Spiral) lijkt minder geschikt voor algemene toepassing.
(6) Het nastreven van grote nauwkeurigheid bij de stapgroottebepaling door nauwkeurige lijnminimalisering in iedere iteratieslag is onvoordelig. De extra hoeveelheid werk (meer functieëvaluaties) die nodig is voor deze nauwkeuriger lijnminimalisering weegt niet op tegen de relatief geringe winst door een kleiner aantal benodigde iteraties voor convergentie.
(7) Het toepassen van numeriek bepaalde, dan wel exact gegeven afgeleiden (elementen van de matrix J) maakt weinig verschil voor het verloop van het iteratieproces. Exact gegeven afgeleiden vereisen echter minder computertijd.

Literatuur

2.10.29. Meer details over de in deze paragraaf besproken methoden kunnen worden gevonden in de volgende publicaties.

[2.10.1] : Zie [2.2.7] Kowalik and Osborne (1968)
[2.10.2] : Zie [2.4.7] Luenberger (1969)
[2.10.3] : Zie [2.5.11] Levenberg (1944)
[2.10.4] : Ben-Israel, A. and Greville, Th.M.E. : Generalized inverses, Theory and applications, Wiley, New York (1974)

[2.10.5] : Bus, J.C.P., van Domselaar, B. en Kok, J. : Nonlinear least squares estimation, Mathematisch Centrum, Amsterdam, Report NW 17/75, May 1975.

[2.10.6] : Eilers, G.A.M.: Het minimaliseren van sommen van kwadraten van niet lineaire functies, Technische Hogeschool Eindhoven, Onderafdeling der Wiskunde, Memorandum COSOR 75-08, juni

[2.10.7] : Fletcher, R. : Generalized inverse methods for the best least squares solution of systems of nonlinear equations, Comp J., 10 (1968), pp. 392-399.

[2.10.8] : Jones, A. : A new algorithm for nonlinear parameter estimation using least squares, Comp J., 13 (1970), pp. 301-308.

[2.10.9] : Lawson, C.L. and Hanson, R.J. : Solving least squares problems, Prentice Hall Inc., Englewood Cliffs, N.J. (1974).

[2.10.10] : Lill, S.A. : A survey of methods for minimizing sums of squares of nonlinear functions, University of Liverpool, Rept. 034/1, June 1975.

[2.10.11] : Marquardt, D.W. : An algorithm for least squares estimation of nonlinear parameters, SIAM J., 11 (1963) pp. 431-441.

[2.10.12] : McKeown, J.J. : Specialized versus general purpose algorithms for minimising functions that are sums of squared terms, Math. Progr., 9 (1975) pp. 57-68.

[2.10.13] : Peckham, G. : A new method for minimizing a sum of squares of nonlinear functions without calculating derivatives, Comp. J., 8, (1970) pp. 418-420.

[2.10.14] : Powell, M.J.D. : A method for minimizing a sum-of-squares of nonlinear functions without calculating derivatives, Comp. J. 7 (1965) pp. 303-307.

[2.10.15] : Powell, M.J.D. : Problems related to unconstrained optimimization, Ch. III of [1.1.3] Murray (1972)

[2.10.16] : Rosen, J.B. : The gradient projection method for nonlinear programming, Part I: Linear constraints J. SIAM 8 (1960) pp. 181-217.

3. METHODEN VOOR MINIMALISERING MET NEVENVOORWAARDEN

§ 3.1. Algemeen.

3.1.1. In dit hoofdstuk zullen methoden worden besproken voor het oplossen van minimaliseringsprobleem van het type CMP (vgl.(1.1.2))(of GNLI (zie pt. 3.1.4))

$$\min \{f(x) \mid g_i(x) \geq 0,\ i \in I,\ h_j(x) = 0, j \in E,\ x \in \mathbb{R}^n\} \qquad (3.1.1)$$

waarin $f(x)$, $g_i(x)$ en $h_j(x)$ reëelwaardige functies van n variabelen voorstellen en I en E eindige indexverzamelingen. Een bijzonder geval van het probleem CMP, dat in het navolgende een centrale rol zal vervullen is het minimaliseringsprobleem met uitsluitend gelijkheidsbeperkingen dat bekend is als het Lagrange probleem (of het GNLE-probleem (zie pt. 3.1.4))

$$\min \{f(x) \mid c(x) = 0,\ x \in \mathbb{R}^n,\ c(x) \in \mathbb{R}^m\} \qquad (3.1.2)$$

waarin f opnieuw een reëelwaardige functie voorstelt ($f : \mathbb{R}^n \to \mathbb{R}$) en c een m-vector functie ($c : \mathbb{R}^n \to \mathbb{R}^m$). Bij de ontwikkeling van de methoden wordt er meestal vanuitgegaan dat f een functie is met een convex karakter en de $g_i(x)$ functies met een concaaf karakter zodat het toegelaten gebied

$$S := \{x \in \mathbb{R}^n \mid g_i(x) \geq 0,\ i \in I,\ h_j(x) = 0,\ j \in E\} \qquad (3.1.3)$$

gesloten en eventueel convex is. Met behulp van deze definitie kan het probleem CMP (3.1.1) ook worden geformuleerd als

$$\min \{f(x) \mid x \in S \subset \mathbb{R}^n\} \qquad (3.1.4)$$

3.1.2. Naast de gradiënt $g := \nabla f(x)$ ((2.1.2)) en de Hessiaan $G(x) := \nabla^2 f(x)$ ((2.1.3)) van de objectfunctie spelen bij minimaliseringsproblemen met nevenvoorwaarden ook de gradiënten van de beperkingen

$$\nabla g_i(x) := \left(\frac{dg_i}{dx}(x)\right)^T \qquad \nabla h_j(x) := \left(\frac{dh_j}{dx}(x)\right)^T \qquad (3.1.5)$$

respectievelijk

$$\nabla c_i(x) := \left(\frac{dc_i}{dx}(x)\right)^T \tag{3.1.5}$$

en de tweede-afgeleiden-matrices van de beperkingen

$$G_i(x) := \nabla^2_{xx} g_i(x) \qquad H_j(x) := \nabla^2_{xx} h_j(x)$$

respectievelijk (3.1.6)

$$\Gamma_i(x) := \nabla^2_{xx} c_i(x) = \left[\frac{\partial^2 c_i}{\partial x_j \partial x_k}(x)\right]$$

een rol bij de ontwikkeling van minimaliserings-algorithmen. Voor de gradiënten (3.1.5) wordt dezelfde notatie afspraak gehanteerd als voor de gradiënt $\nabla f(x)$ van de objectfunctie (vgl. pt. 2.1.1) d.w.z. de gradiënt $\nabla c_i(x)$ wordt opgevat als een kolomvector en de afgeleide dc_i/dx als een rijvector *zodat* (als $c_i \in C^2$)

$$dc_i = \frac{dc_i}{dx}dx = \nabla c_i^T dx = \quad <\nabla c_i, dx>$$

In sommige gevallen wordt ook gebruik gemaakt van de n x m-matrix met als kolommen de gradiënten van de beperkingen.Deze wordt genoteerd als

$$[\nabla c_i(x)] := [\nabla c_1(x) \ \ldots \ \nabla c_m(x)] \tag{3.1.7}$$

De gradiënt van een beperking in een punt x waar aan die beperking, als een gelijkheid wordt voldaan wordt gebruikelijk de <u>normaal</u> van die beperking genoemd en wordt genoteerd als

$$n_i := n_i(x) := \nabla c_i(x) \qquad (c_i(x) = 0) \tag{3.1.8}$$

Voor de matrix met als kolommen de normalen van de beperkingen in een punt x waar aan die beperkingen als gelijkheden wordt voldaan wordt, mits deze normalen lineair onafhankelijk zijn, gebruik gemaakt van de notatie

$$N := N(x) := [n_1 \ \ldots\ldots \ n_k] \qquad (c_1(x) = 0 \ \ldots \ c_k(x) = 0) \tag{3.1.9}$$

3.1.3. Een belangrijke functie die een centrale rol speelt in de theorie en de practijk van de *minimalisering onder nevenvoorwaarden* is de Lagrange-functie die voor de problemen (3.1.1) en (3.1.2) wordt gedefinieerd als

$$L(x,\lambda,\mu) = f(x) - \sum_{i\in I} \mu_i g_i(x) - \sum_{j\in E} \lambda_j h_j(x)$$

respectievelijk (3.1.10)

$$L(x,\lambda) = f(x) - c^T(x)\lambda$$

De parameters μ_i , $i \in I$ en λ_j , $j \in E$ worden de Lagrange multiplicatoren genoemd, de vector λ de Lagrange-(multiplicatoren-)vector. De min-tekens in de formuleringen van de Lagrange-functies zijn eigenlijk alleen van belang bij de ongelijkheden $g_i(x) \geq 0$: Als $f(x)$ convex is, $g_i(x)$ concaaf en $\mu_i \geq 0$ dan wordt door de keuze van het min-teken ook de Lagrange-functie een convexe functie van x. Uit het oogpunt van *uniformiteit* wordt het *min*-teken in de Lagrange-functie ook gebruikt voor de gelijkheden. Zowel in de theorie als in de practijk van de optimalisering wordt ook een belangrijke rol gespeeld door de gradiënt m.b.t. x van de Lagrange-functie

$$\nabla_x L(x,\lambda,\mu) = \nabla f(x) - \sum_{i\in I} \mu_i \nabla g_i(x) - \sum_{j\in E} \lambda_j \nabla h_j(x)$$

respectievelijk (3.1.11)

$$\nabla_x L(x,\lambda) = \nabla f(x) - [\nabla c_i(x)]\lambda$$

en de Hessiaan van de Lagrange-functie

$$\nabla^2_{xx} L(x,\lambda,\mu) = G(x) - \sum_{i\in I} \mu_i G_i(x) - \sum_{j\in E} \lambda_j H_j(x)$$

respectievelijk (3.1.12)

$$\nabla^2_{xx} L(x,\lambda) = G(x) - \sum_{i=1}^{m} \lambda_i \Gamma_i(x)$$

De afgeleiden van de Lagrange-functie naar de Lagrange-multiplicatoren resp. de gradiënt van de Lagrange-functie m.b.t. de Lagrange-vector zijn de beperkingen zelf (voorzien van een min-teken)

$$\frac{\partial L(x,\lambda,\mu)}{\partial \mu_i} = -g_i(x) \qquad \frac{\partial L(x,\lambda,\mu)}{\partial \lambda_j} = -h_j(x)$$

respectievelijk (3.1.13)

$$\nabla_\lambda L(x,\lambda) = -c(x)$$

Indeling in typen van beperkingen en speciale problemen

3.1.4. De beperkingen die het toegelaten gebied S (3.1.3) bepalen kunnen behalve in gelijkheids- en ongelijkheids beperkingen ook worden onderverdeeld naar het karakter van de functie die de beperking representeert. Men maakt dan onderscheid tussen a) coördinaatbeperkingen met als karakteristieke vorm

$$x_i \geq a_i \tag{3.1.14}$$

b) lineaire beperkingen met als karakteristieke vorm

$$A_1^T x - b_1 = 0 \qquad A_2^T x - b_2 \geq 0 \tag{3.1.15}$$

en c) algemene of niet-lineaire beperkingen

$$g_i(x) \geq 0 \qquad h_j(x) = 0 \qquad i \in I \ , \ j \in E$$

Coördinaatbeperkingen hebben uiteraard alleen zin als ongelijkheids beperkingen, de lineaire en niet-lineaire beperkingen kunnen zowel gelijkheids- als ongelijkheidsbeperkingen zijn. Coördinaatbeperkingen vormen een deelverzameling van de lineaire beperkingen. Het onderscheid ertussen speelt een wezenlijke rol bij lineaire- en kwadratische programmeringsproblemen. Bij algemene minimaliseringsproblemen voor problemen met nevenvoorwaarden is het onderscheid niet wezenlijk en is vrijwel uitsluitend van practisch belang bij de organisatie van de algorithmen. In de theorie van de algemene algorithmen is het onderscheid vrijwel niet van belang en wordt om die reden hierna dan ook achterwege gelaten.

Afhankelijk van het karakter van de object functie en het karakter van de beperkingen kunnen de minimaliseringsproblemen met nevenvoorwaarden worden onderverdeeld in de volgende probleemtypen (met de standaardnotatie voor later gebruik) :

a) lineaire-programmerings- of LP-problemen

$$\min \{c^T x \mid A^T x - b = 0 \ , \ x \geq 0\} \tag{3.1.16}$$

b) kwadratische-programmerings- of QP-problemen

$$\min \{\tfrac{1}{2} x^T Q x + q^T x \mid A^T x - b = 0 \quad x \geq 0\} \tag{3.1.17}$$

c) minimaliseringsproblemen met kwadratische objectfunctie met lineaire gelijkheidsbeperkingen (=: QLE-problemen)

$$\min \{\tfrac{1}{2} x^T Q x + q^T x \mid A^T x - b = 0\} \tag{3.1.18}$$

d) minimaliseringsproblemen met kwadratische objectfunctie met lineaire gelijkheids- en ongelijkheidsbeperkingen (=: QLI-problemen)

$$\min \{\tfrac{1}{2} x^T Q x + q^T x \mid A_1^T x - b_1 = 0 \quad A_2^T x - b_2 \geq 0\} \tag{3.1.19}$$

e) minimaliseringsproblemen met algemene objectfunctie en lineaire gelijkheidsbeperkingen (=: GLE-problemen)

$$\min \{f(x) \mid A^T x = b\} \tag{3.1.20}$$

f) minimaliseringsproblemen met algemene objectfunctie en lineaire gelijkheids- en ongelijkheidsbeperkingen (=: GLI-problemen)

$$\min \{f(x) \mid A_1^T x - b_1 = 0 \ , \ A_2^T x - b_2 \geq 0\} \tag{3.1.21}$$

g) minimaliseringsproblemen met algemene objectfunctie en algemene gelijkheidsbeperkingen (=: GNLE-problemen of Lagrange-problemen (vgl. (3.1.2)))

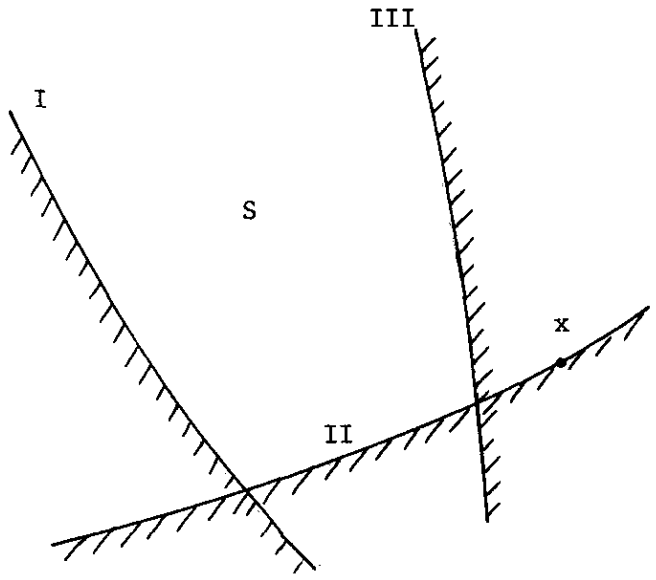
$$\min \{f(x) \mid c(x) = 0\}$$

h) minimaliseringsproblemen met algemene objectfunctie en algemene gelijkheids- en ongelijkheidsbeperkingen (=: GNLI-problemen (vgl. (3.1.1)))

$$\min \{f(x) \mid g(x) \geq 0 \quad , \quad h(x) = 0\}$$

Karakterisering van optimale punten

3.1.5. Bij de voorwaarden voor optimaliteit speelt weer een ander indeling van de beperkingen een rol dan die besproken in het voorafgaande, namelijk een indeling die gebaseerd is op het al of niet voldaan zijn aan de beperkingen in het beschouwde punt. Men onderscheidt daarbij (zie Figuur 3.1.5) overschreden, actieve en passieve beperkingen: Een beperking heet overschreden (eng.:violated) indien in dat punt niet voldaan wordt aan de beperking, een beperking heet actief in een punt x



Figuur 3.1.5. Indeling van de beperkingen

indien in dat punt aan die beperking voldaan is als een gelijkheid en een beperking heet passief in een punt x indien in dat punt voldaan is aan de (ongelijkheids-) beperking als een strikte ongelijkheid. De overeenkomstige indexverzamelingen worden in de terminologie van probleem CMP (3.1.1) respectievelijk gedefinieerd als

a) indexverzameling van overschreden beperkingen

$$I_V(x) := \{ i \in I \mid g_i(x) < 0 \} \cup \{ j \in E \mid h_j(x) \neq 0 \} \qquad (3.1.22a)$$

b) indexverzameling van actieve beperkingen

$$I_A(x) := \{ i \in I \mid g_i(x) = 0 \} \cup \{ j \in E \mid h_j(x) = 0 \} \qquad (3.1.22b)$$

c) indexverzameling van passieve beperkingen

$$I_p(x) := \{i \in I \mid g_i(x) > 0\} \qquad (3.1.22c)$$

3.1.6. Van veel belang voor de theorie van de optimalisering onder nevenvoorwaarden is de observatie dat alleen die beperkingen die actief zijn in het optimale punt een rol spelen bij de optimaliteitsvoorwaarden. Alle andere beperkingen zouden voor wat de optimaliteitsvoorwaarden betreft verwijderd kunnen worden uit de probleemformulering. Een tweede belangrijke observatie die ook voor de practijk van de algorithmen van belang is (vgl. pt. 3.2.3) de constatering dat ieder minimaliseringsprobleem met gelijkheids- en ongelijkheidsbeperkingen gereduceerd kan worden tot een eenvoudiger probleem met uitsluitend gelijkheidsbeperkingen indien van te voren bekend zou zijn welke beperkingen actief zijn in het optimale punt. De voorwaarden waaraan de oplossing van dit eenvoudiger probleem moet voldoen vormen de kern van de optimaliteitsvoorwaarden voor de oplossing van het originele probleem. In verband met deze observaties wordt hieronder eerst ingegaan op de voorwaarden van problemen met uitsluitend gelijkheidsbeperkingen d.w.z. problemen van het type GNLE (3.1.2)

$$\min \{f(x) \mid c(x) = 0, \quad x \in \mathbb{R}^n, c(x) \in \mathbb{R}^m\}$$

Daarna worden de aanpassingen besproken van de optimaliteitsvoorwaarden voor de originele algemene problemen van het type CMP of GNLI (3.1.1).

Linearisatie van de beperkingen in het optimale punt

3.1.7. Essentieel voor de optimaliteitsvoorwaarden voor oplossingen van problemen van het type GNLE (3.1.2) is de vraag of de beperkingen in het optimale punt op eenduidige wijze kunnen worden gelineariseerd. Een voldoende voorwaarde daarvoor is de lineaire onafhankelijkheid van de locale normalen. Men spreekt in dat geval over een regulier punt.

Definitie 3.1.7. : Een punt x waar $c_j(x) = 0$, $j = 1, \ldots, m$ heet een regulier punt m.b.t. de beperkingen indien de normalen $\nabla c_j(x)$ van de actieve beperkingen in het punt x onderling lineair onafhankelijk zijn in het punt x.

In een regulier punt x geldt (vgl. [3.1.1]) dat het raakvlak, d.i. dat de verzameling van de afgeleiden (of richtingsvectoren) $\dot{x} := dx/dt$ van alle differentieerbare krommen $x(t)$ met $x(0) = x$ in het door de beperkingen $c_j(x) = 0$ bepaalde oppervlak gelijk is aan de lineaire deelruimte

$$M(x) := \{z \in \mathbb{R}^n \mid \nabla^T c_j(x) z = 0, j = 1, \ldots, m\} \tag{3.1.23a}$$

of equivalent in termen van de eerder gedefinieerde matrix van normalen (3.1.9) (met weglating van het argument x)

$$N := [n_1 \ \ldots \ n_m] := [\nabla c_1(x) \ \ldots \ \nabla c_m(x)]$$

de deelruimte

$$M(x) := \{z \in \mathbb{R}^n \mid N^T z = 0\} \tag{3.1.23b}$$

Deze speelt een fundamentele rol bij de optimaliteitsvoorwaarden bij minimaliseringsproblemen met nevenvoorwaarden.

3.1.8. Het is gebruikelijk om in deelruimte M(x) (3.1.23) een orthonormale basis $\{z_1, \ldots, z_{n-m}\}$ te kiezen en een matrix Z te definiëren als de n x (n-m) matrix met als kolommen de basisvectoren $z_1, \ldots, z_{n-m}$

$$Z := [z_1 \ \ldots \ z_{n-m}] \tag{3.1.24}$$

Voor deze matrix gelden dan de relaties

$$N^T Z = 0_{m \times (n-m)} \tag{3.1.25}$$

en

$$Z^T Z = I_{(n-m) \times (n-m)} \tag{3.1.26}$$

De kolommen van de samengestelde matrix

$$[N \mid Z] \tag{3.1.27}$$

vormen een in deze theorie veelgebruikte basis voor de gehele $\mathbb{R}^n$. De observatie dat zowel

$$ZZ^T[N \mid Z] = [0 \mid Z]$$

als ook (3.1.28)

$$(I - N(N^TN)^{-1}N^T)\ [N \mid Z] = [0 \mid Z]$$

illustreert, bijvoorbeeld, dat de matrices ZZ^T en $(I - N(N^TN)^{-1}N^T)$ dezelfde projectie matrices zijn die willekeurige vectoren uit $\mathbb{R}^n$ projecteren op de lineaire deelruimte M(x) (opgespannen door de (orthonormale) kolommen van de matrix Z) en dat daarom geldt

$$ZZ^T = (I - N(N^TN)^{-1}N^T) \tag{3.1.29}$$

Deze laatste formule had ook kunnen worden verkregen uit de eenvoudig te verifiëren betrekking

$$\begin{bmatrix} N^T \\ \hline Z^T \end{bmatrix}^{-1} = [N(N^TN)^{-1} \mid Z] \tag{3.1.30}$$

waaruit volgt dat

$$[N(N^TN)^{-1} \mid Z] \begin{bmatrix} N^T \\ \hline Z^T \end{bmatrix} = N(N^TN)^{-1}N^T + ZZ^T = I$$

<u>Eerste orde noodzakelijke voorwaarden</u>

3.1.9. Een noodzakelijke voorwaarde opdat in een punt $\hat{x}$ een minimum optreedt van de functiewaarden f(x) in die punten x die voldoen aan c(x) = 0 is de in de probleemformulering vervatte eis dat geldt

$$c(\hat{x}) = 0 \tag{3.1.31}$$

Daarnaast is het nodig dat er geen kromme x(t) met $x(0) = \hat{x}$ en $c(x(t)) = 0$ voor $-a \leq t \leq a$ met $a > 0$ bestaat waarlangs de object functiewaarde f(x(t)) in waarde afneemt. Met een eerste orde (= lineaire) benadering leidt dit tot de uitspraak.

Stelling 3.1.9. : Een (eerste orde) noodzakelijke voorwaarde opdat een regulier punt $\hat{x}$ dat voldoet aan (3.1.31)

$$c(\hat{x}) = 0$$

een oplossing is van het minimaliseringsprobleem GNLE (3.1.2) is de eis dat voor alle vectoren $z \in M(\hat{x}) := \{z \in \mathbb{R}^n \mid N^T z = 0\}$ geldt $\nabla^T f(\hat{x}) z = 0$, of wel, anders geformuleerd

$$\forall z \in \mathbb{R}^n : N^T z = 0 \Rightarrow \nabla^T f(\hat{x}) z = 0 \qquad (3.1.32)$$

Bewijs : (vgl. [3.1.1]) Als $\hat{x}$ een minimum is van $f(x)$ onder de nevenvoorwaarden $c(x) = 0$ dan geldt voor de afgeleide naar t van de functiewaarden $f(x(t))$ van alle krommen $x(t)$ met $x(0) = \hat{x}$ en $c(x(t)) = 0$ voor $-a \leq t \leq a$ met $a > 0$ dat

$$\frac{df(x(t))}{dt}\Big|_{t=0} = \nabla^T f(\hat{x})\dot{x}(0) = 0 \qquad (3.1.33)$$

Omdat $\hat{x}$ een regulier punt is geldt dat de verzameling van alle afgeleiden of richtingsvectoren $\dot{x}(0)$ van de differentieerbare krommen $x(t)$ gelijk is aan de verzameling $M(\hat{x})$ (3.1.23) zodat (3.1.33) equivalent is met

$$\forall z \in M(\hat{x}) : \nabla^T f(\hat{x}) z = 0 \qquad (3.1.34)$$

Uit de definitie (3.1.23) van $M(\hat{x})$ volgt de equivalentie van (3.1.34) en (3.1.32). □

3.1.10. De noodzakelijke voorwaarde (3.1.32) kan in termen van de matrices N (3.1.9) en Z (3.1.24) ook worden weergegeven door de voorwaarde

$$Z^T \nabla f(\hat{x}) = 0 \qquad (3.1.35)$$

of equivalent i.v.m.(3.1.29) door de voorwaarde

$$(I - N(N^T N)^{-1} N^T) \nabla f(\hat{x}) = 0 \qquad (3.1.36)$$

In woorden impliceert deze laatste uitdrukking dat "de projectie van de gradiënt van de objectfunctie op het raakvlak aan de beperkingen gelijk moet zijn aan nul".

3.1.11. Een andere bekende formulering van de noodzakelijke voorwaarde (3.1.32) voor optimaliteit wordt gegeven in de volgende aan Stelling 3.1.9. equivalente uitspraak.

Stelling 3.1.11. : Een (eerste orde) noodzakelijke voorwaarde opdat een regulier punt $\hat{x}$ dat voldoet aan (3.1.31)

$$c(\hat{x}) = 0$$

de oplossing is van het minimaliseringsprobleem GNLE (3.1.2) is de eis dat er een (Lagrange-) vector $\hat{\lambda} \in \mathbb{R}^m$ bestaat zo dat

$$\nabla f(\hat{x}) = N\hat{\lambda} \qquad (3.1.37)$$

waar N is gedefinieerd door (3.1.9).

Bewijs : Het resultaat volgt onmiddellijk uit een toepassing van de aan Gale toegeschreven (vgl. [3.1.4] p. 33) variant van de Stelling van Farkas die inhoud dat voor een willekeurige n x m-matrix A en een gegeven vector $c \in \mathbb{R}^m$ de bewering

$$\neg\ \exists\ y \in \mathbb{R}^n \qquad : \quad A^T y = 0 \wedge c^T y \neq 0$$

equivalent is met

$$\exists\ \hat{\lambda} \in \mathbb{R}^m \qquad : \quad A\hat{\lambda} = c$$

Met $y \leftarrow z$ en $c \leftarrow \nabla f$ is dit juist wat te bewijzen was. □

3.1.12. In componenten van de vector $\hat{\lambda}$ uitgeschreven kan de eerste orde voorwaarde (3.1.37) ook worden weergegeven door

$$\nabla f(\hat{x}) = \sum_{i=1}^{m} \hat{\lambda}_i n_i = \sum_{i=1}^{m} \hat{\lambda}_i \nabla c_i(\hat{x}) \qquad (3.1.38)$$

of in woorden : "de gradiënt is een lineaire combinatie van de normalen van de beperkingen in het optimale punt". In termen van de eerder besproken, in verband met deze eerste orde noodzakelijke voorwaarde gecreëerde Lagrange-functie (3.1.10)

$$L(x,\lambda) := f(x) - c^T(x)\lambda = f(x) - \sum_{i=1}^{m} \lambda_i c_i(x)$$

is de voorwaarde (3.1.37) gelijk aan

$$\nabla_x L(\hat{x},\hat{\lambda}) = \nabla f(\hat{x}) - N\hat{\lambda} = 0 \qquad (3.1.39)$$

Deze voorwaarde samen met de voorwaarde dat $\hat{x}$ voldoet aan de beperkingen

$$\nabla_\lambda L(\hat{x},\hat{\lambda}) = -c(\hat{x}) = 0 \qquad (3.1.40)$$

bepaalt de verzameling van stationaire punten van het probleem GNLE, van welke verzameling ook de oplossing van het probleem GNLE deel uit maakt.

3.1.13. Vergelijking van de uitdrukkingen (3.1.36) en (3.1.39) leert dat

$$N\hat{\lambda} = N(N^TN)^{-1}N^T\nabla f(\hat{x})$$

waaruit kan worden afgeleid dat (in het beschouwde geval waar N volle rang verondersteld wordt te hebben) geldt

$$\hat{\lambda} = N^+\nabla f(\hat{x}) = (N^TN)^{-1}N^T\nabla f(\hat{x}) \qquad (3.1.41)$$

waar N^+ de pseudo-inverse (vgl. Definitie 2.10.17) van de matrix N in het optimale punt $\hat{x}$ voorstelt. Wordt gebruik gemaakt van een QR-decompositie (vgl. pt. 2.10.18) van de matrix N in het optimale punt

$$N = QR = [Q_1 \mid Q_2] \begin{bmatrix} \Delta \\ 0 \end{bmatrix} \qquad (3.1.42)$$

dan volgt dat voor $\hat{\lambda}$ geldt (vgl. (2.10.49))

$$\hat{\lambda} = R^{-1}Q_1^T \nabla f(\hat{x}) \qquad (3.1.43)$$

Deze laatste uitdrukkingen (3.1.41) en (3.1.43) worden soms gebruikt bij practische toepassingen voor de bepaling van de Lagrange multiplicatoren in het optimum.

Tweede orde noodzakelijke en voldoende voorwaarden

3.1.14. Tweede orde noodzakelijke voorwaarden voor optimaliteit van de oplossing van minimaliseringsproblemen met nevenvoorwaarden kunnen worden afgeleid door beschouwing van tweede orde (Taylor ontwikkelingen van zowel de objectfunctie als de beperkingen in het bekend veronderstelde optimale punt. Toevoeging van lineaire combinaties van de tweede orde ontwikkelingen van de beperkingen aan de tweede orde ontwikkeling van objectfunctie maakt het mogelijk (door de keuze van de optimale Lagrange multiplicatoren (3.1.37) als coëfficiënten) de eerste orde termen uit de ontwikkeling te elimineren met als resultaat, dat de tweede orde ontwikkeling van de aldus gemodificeerde objectfunctie een zuivere kwadratische vorm wordt met als Hessiaan de Hessiaan van de Lagrange-functie. Noodzakelijk voor optimaliteit is het niet-negatief definiet zijn van deze Hessiaan voor alle variaties die in eerste orde aan de beperkingen voldoen (Hoger orde variaties die aan de beperkingen voldoen leiden tot hoger dan tweede orde termen in de gemodificeerde objectfunctie). Uitgewerkt leidt de hier geschetste aanpak tot de uitspraak :

Stelling 3.1.14 : Een (tweede orde) noodzakelijke voorwaarde opdat een regulier punt $\hat{x}$ dat voldoet aan $c(\hat{x}) = 0$ (3.1.31) een oplossing is van het minimaliseringsprobleem GNLE (3.1.2)

$$\min \{f(x) \mid c(x) = 0 \ , \quad x \in \mathbb{R}^n, c(x) \in \mathbb{R}^m\}$$

is dat voor alle vectoren $z \in M(\hat{x}) := \{z \mid N^T z = 0\}$ geldt dat $z^T \nabla^2_{xx} L(\hat{x},\hat{\lambda}) z \geq 0$, of anders geformuleerd,

$$\forall z \in \mathbb{R}^n : N^T z = 0 \Rightarrow z^T \nabla^2_{xx} L(\hat{x},\hat{\lambda}) z \geq 0 \qquad (3.1.44)$$

waar (vgl. (3.1.12))

$$\nabla^2_{xx} L(\hat{x},\hat{\lambda}) = G(\hat{x}) - \sum_{i=1}^{m} \hat{\lambda}_i \Gamma_i(\hat{x})$$

en $\hat{\lambda}$ voldoet aan (3.1.39)

$$\nabla_x L(\hat{x},\hat{\lambda}) = \nabla f(\hat{x}) - N\hat{\lambda} = 0$$

Bewijs : (vgl. [3.1.1]) Als $\hat{x}$ een minimum is van $f(x)$ onder de nevenvoorwaarde $c(x) = 0$ dan geldt (vgl. (3.1.33)) voor de eerste afgeleide naar t van de functiewaarden $f(x(t))$ van alle krommen $x(t)$ met $x(0) = \hat{x}$ en $c(x(t)) = 0$ voor $-a \leq t \leq a$ met $a > 0$ dat

$$\left.\frac{df(x(t))}{dt}\right|_{t=0} = \nabla^T f(\hat{x})\dot{x}(0) = 0$$

en voor de tweede afgeleide

$$\left.\frac{d^2 f(x(t))}{dt^2}\right|_{t=0} = \dot{x}^T(0) G(\hat{x})\dot{x}(0) + \nabla^T f(\hat{x})\ddot{x}(0) \geq 0 \qquad (3.1.45)$$

Analoog geldt dat voor de analoge afgeleiden voor de individuele beperkingen ($i = 1, \ldots, m$) moet gelden

$$\left.\frac{dc_i(x(t))}{dt}\right|_{t=0} = \nabla^T c_i(\hat{x})\dot{x}(0) = 0$$

en

$$\left.\frac{d^2 c_i(x(t))}{dt^2}\right|_{t=0} = \dot{x}^T(0)\Gamma_i(\hat{x})\dot{x}(0) + \nabla^T c_i(\hat{x})\ddot{x}(0) = 0 \qquad (3.1.46)$$

Sommatie van (3.1.45) met een lineaire combinatie van de uitdrukkingen (3.1.46) geeft als voorwaarde dat

$$\dot{x}^T(0)[G(\hat{x}) - \sum_{i=1}^{m} \lambda_i \Gamma_i(\hat{x})]\dot{x}(0) + (\nabla f(\hat{x}) - \sum_{i=1}^{m} \lambda_i \nabla c_i(\hat{x}))^T \ddot{x}(0) \geq 0 \quad (3.1.47)$$

welke voorwaarde als voor λ gekozen wordt de optimale waarde $\hat{\lambda}$ die voldoet aan (3.1.39) overgaat in

$$\dot{x}^T(0)[G(\hat{x}) - \sum_{i=1}^{m} \hat{\lambda}_i \Gamma_i(\hat{x})]\dot{x}(0) = \dot{x}^T(0)\nabla^2_{xx} L(\hat{x},\hat{\lambda})\dot{x}(0) \geq 0 \quad (3.1.48)$$

Omdat $\hat{x}$ een regulier punt is geldt dat de verzameling van alle afgeleiden of richtingsvectoren $\dot{x}(0)$ van de differentieerbare krommen gelijk is aan $M(\hat{x})$ ((3.1.23)). De voorwaarde (3.1.48) is dus juist equivalent met de in de Stelling genoemde voorwaarde (3.1.44). □

3.1.15. In het geval het minimaliseringsprobleem uitsluitend lineaire nevenvoorwaarden kent, zoals in het geval bij de problemen van het type QP, QLE, QLI, GLE en GLI (vgl. pt. 3.1.4) geldt dat

$$\nabla^2_{xx} L(x,\lambda) = G(x) \quad (3.1.49)$$

in welk geval de noodzakelijke voorwaarde (3.1.44) overgaat in

$$\forall z \in \mathbb{R}^n : N^T z = 0 \Rightarrow z^T G(\hat{x}) z \geq 0 \quad (3.1.50)$$

In woorden impliceren de noodzakelijke voorwaarden (3.1.44) dat "de Hessiaan van de Lagrange-functie (= aangepaste objectfunctie) niet-negatief-definiet is voor verstoringen in de gelineariseerde, resp. lineaire beperkingen".

3.1.16. Analoog aan de situatie in het geval van minimalisering zonder nevenvoorwaarden (vgl. pt. 2.1.3) wordt een set voldoende voorwaarden voor de oplossing van het probleem GNLE (3.1.2) gegeven door een combinatie van de gevonden noodzakelijke eerste orde voorwaarden en een sterkere versie van de tweede orde voorwaarde. Dit is weergegeven in de volgende uitspraak:

Stelling 3.1.16. : Voldoende voor het optreden van een oplossing van het probleem GNLE (3.1.2) in een punt $\hat{x}$ is het bestaan van een Lagrange-vector $\hat{\lambda}$ waarvoor geldt (vgl. (3.1.39) en (3.1.40))

$$\nabla_x L(\hat{x},\hat{\lambda}) = \nabla f(\hat{x}) - N\hat{\lambda} = 0$$

$$\nabla_\lambda L(\hat{x},\hat{\lambda}) = -c(\hat{x}) = 0$$

en

$$\forall z \in \mathbb{R}^m : z \neq 0 \wedge N^T z = 0 \Rightarrow z^T \nabla^2_{xx} L(\hat{x},\hat{\lambda}) z > 0 \qquad (3.1.51)$$

Bewijs : Voor een bewijs (uit het ongerijmde) zij verwezen naar [3.1.1]. □

Noodzakelijke en voldoende voorwaarden bij aanwezigheid van ongelijkheden.

3.1.17. In het voorgaande (pt. 3.1.6) werd reeds opgemerkt dat bij de optimaliteits-voorwaarden voor algemene minimaliseringsproblemen van het type GNLI (of CMP : (3.1.1))

$$\min \{f(x) \mid g_i(x) \geq 0,\ i \in I,\ h_j(x) = 0,\ j \in E\}$$

alleen die beperkingen een rol spelen die actief zijn in het optimale punt. Bovendien geldt dat de optimaliteitsvoorwaarden voor het eenvoudiger probleem (van het type GNLE) dat betrekking heeft op het minimaliseren van dezelfde objectfunctie met als nevenvoorwaarden de actieve beperkingen in het optimale punt als gelijkheidsbeperkingen, d.i. het probleem

$$\min \{f(x) \mid g_i(x) = 0,\ i \in I_A(\hat{x}),\ h_j(x) = 0,\ j \in E\} \qquad (3.1.52)$$

een deelverzameling vormen van de optimaliteitsvoorwaarden van het originele probleem van het type GNLI (3.1.1). Dit laatste feit leidt tot de uitspraak:

Stelling 3.1.17. : Noodzakelijke voorwaarden opdat in een regulier punt $\hat{x}$ de oplossing gevonden wordt van het minimaliseringsprobleem GNLI (3.1.1) zijn o.a. de voorwaarden dat

$$g_i(\hat{x}) = 0 \qquad i \in I_A(\hat{x}) \tag{3.1.53}$$

$$h_j(\hat{x}) = 0 \qquad j \in E \tag{3.1.54}$$

$$\nabla f(\hat{x}) - \sum_{i \in I_A(\hat{x})} \hat{\mu}_i \nabla g_i(\hat{x}) - \sum_{j \in E} \hat{\lambda}_j \nabla h_j(\hat{x}) = 0 \tag{3.1.55}$$

en

$$\forall z \in M(\hat{x}) : z^T [G(\hat{x}) - \sum_{i \in I_A(\hat{x})} \hat{\mu}_i G_i(\hat{x}) - \sum_{j \in E} \hat{\lambda}_j H_j(\hat{x})] z \geq 0 \tag{3.1.56}$$

waar

$$M(\hat{x}) := \{z \in \mathbb{R}^n \mid \nabla^T g_i(\hat{x}) z = 0,\ i \in I_A(\hat{x}),\ \nabla^T h_j(\hat{x}) z = 0,\ j \in E\} \tag{3.1.57}$$

Bewijs : Zie voorgaande argumentatie en pt. 3.1.9. en 3.1.15. □

3.1.18. Bij aanwezigheid van ongelijkheden in de probleemformulering is de verzameling van toegelaten richtingen in het reguliere optimale punt $\hat{x}$ groter dan de boven gedefinieerde verzameling $M(\hat{x})$ (3.1.57) en wel inplaats daarvan gelijk aan

$$M_I(\hat{x}) = \{z \in \mathbb{R}^n \mid \nabla^T g_i(\hat{x}) z \geq 0,\ i \in I_A(\hat{x}), \nabla^T h_j(\hat{x}) z = 0,\ j \in E\} \tag{3.1.58}$$

Naast de eis dat er (vgl. pt. 3.1.9)geen $z \in M(\hat{x})$ is zodat $\nabla^T f(\hat{x}) z < 0$ komt nu de eis dat

$$\neg z \in \{M_I(\hat{x}) \setminus M(\hat{x})\} : \nabla^T f(\hat{x}) z < 0 \tag{3.1.59}$$

Deze eis geeft aanleiding tot de volgende extra noodzakelijke voorwaarde.

Stelling 3.1.18. : Een extra noodzakelijke voorwaarde opdat een regulier punt $\hat{x}$ dat voldoet aan de noodzakelijke optimaliteitsvoorwaarden (3.1.53), (3.1.54), (3.1.55) en (3.1.56) voor het eenvoudiger probleem (3.1.52) ook

een regulier punt is voor het minimaliseringsprobleem met ongelijkheden is de voorwaarde dat voor de Lagrange-multiplicatoren $\hat{\mu}_i$ die corresponderen met de actieve ongelijkheidsbeperkingen geldt

$$\hat{\mu}_i \geq 0 \qquad i \in I_A(\hat{x}) \tag{3.1.60}$$

Bewijs : Stel dat $\hat{x}$ een optimaal punt is dat voldoet aan alle voorwaarden in Stelling 3.1.17 en dat voor de i_0-de beperking met $i_0 \in I_A(\hat{x})$ geldt

$$\hat{\mu}_{i_0} < 0$$

Dan geldt voor willekeurige $d \in \mathbb{R}^n$ dat

$$\nabla^T f(\hat{x})d = \Big(\sum_{\substack{i \in I_A(\hat{x}) \\ i \neq i_0}} \hat{\mu}_i \nabla g_i(\hat{x}) + \sum_{j \in E} \hat{\lambda}_j \nabla h_j(\hat{x}) \Big)^T d + \mu_{i_0} \nabla^T g_{i_0}(\hat{x})d$$

Omdat $\hat{x}$ een regulier punt is geldt dat er een richting $\bar{d}$ gevonden kan worden zodat

$$\Big(\sum_{\substack{i \in I_A(\hat{x}) \\ i \neq i_0}} \hat{\mu}_i \nabla g_i(\hat{x}) + \sum_{j \in E} \hat{\lambda}_j \nabla h_j(\hat{x}) \Big)^T \bar{d} = 0$$

en

$$\nabla^T g_{i_0}(\hat{x})\bar{d} > 0$$

Voor deze $\bar{d}$ die een element is van $M_I(\hat{x}) \setminus M(\hat{x})$ geldt dan met $\hat{\mu}_{i_0} < 0$ dat

$$\nabla^T f(\hat{x})\bar{d} = \hat{\mu}_{i_0} \nabla^T g_{i_0}(\hat{x})\bar{d} < 0$$

Dit is in tegenspraak met de veronderstelling dat $\hat{x}$ een optimaal punt is en de aanname dat $\hat{\mu}_{i_0} < 0$ is dus onjuist. Het bewijs is daarmee geleverd. □

3.1.19. Herdefinieert men de matrix N (3.1.9) als de matrix met als (lineair onafhankelijke) kolommen de normalen van alle actieve gelijkheids- en ongelijkheidsbeperkingen in een regulier punt

$$N := N(x) = [\underbrace{\ldots \quad \nabla g_i(x) \quad \ldots}_{i \in I_A(x)} \quad \underbrace{\ldots \quad \nabla h_j(x) \quad \ldots}_{j \in E}] \tag{3.1.61}$$

dan kan de extra noodzakelijke voorwaarde (3.1.60) worden geherformuleerd als de voorwaarde dat i-de componenten van de Lagrange-vector (3.1.41)

$$\hat{\lambda} = N^+ \nabla f(\hat{x}) = (N^T N)^{-1} N^T \nabla f(\hat{x})$$

niet-negatief zijn voor alle indices $i \in I_A(\hat{x})$. Noteert men de rijvectoren van de matrix N^+ beschouwd als kolomvectoren door n_l^+, d.w.z. stelt men dat

$$N^+ = \begin{bmatrix} (n_1^+)^T \\ (n_2^+)^T \\ \vdots \end{bmatrix} \tag{3.1.62}$$

dan kan de extra noodzakelijke voorwaarde worden herschreven als

$$(n^+_i)^T \nabla f(\hat{x}) \geq 0 \qquad i \in I_A(\hat{x}) \tag{3.1.63}$$

Aan deze laatste vorm van de noodzakelijke voorwaarde (3.1.60) kan gemakkelijk een geometrische interpretatie gegeven worden met behulp van de voor de vectoren n_i^+ en n_i geldende relaties

$$\begin{aligned} (n_i^+)^T n_j &= 0 \qquad j \neq i \\ &= 1 \qquad j = i \end{aligned} \tag{3.1.64}$$

3.1.20. De noodzakelijke voorwaarden gegeven in de Stellingen 3.1.17 en 3.1.18

representeren niet de gebruikelijke vorm waarin de eerste orde noodzakelijke voorwaarden voor het optimum van problemen van het type GNLI worden gegeven. Deze gebruikelijke vorm, die bekend staat onder de naam Kuhn-Tucker-condities, gaat uit van een iets minder sterke eis t.a.v. de linearisatie van de beperkingen in het optimale punt. Men spreekt in dat verband over een gekwalificeerd punt (d.i. een punt waar de "constraint qualificatie" (vgl. [3.1.4]) is voldaan) i.p.v. over een regulier punt. Een regulier punt is altijd een gekwalificeerd punt, het omgekeerde is niet altijd het geval. Met dit begrip van gekwalificeerd punt luiden de Kuhn-Tucker condities voor een oplossing van het probleem GNLI (3.1.1).

Stelling 3.1.20 (vgl. [3.1.4]) : Noodzakelijk opdat een gekwalificeerd punt $\hat{x}$ de oplossing is van het probleem GNLI (3.1.1) is dat er Lagrange (multiplicatoren)-vectoren $\hat{\lambda}$ en $\hat{\mu}$ bestaan zodat voldaan wordt aan de voorwaarden

$$\begin{aligned} &(1) \qquad g_i(\hat{x}) \geq 0 \qquad i \in I \\ &\qquad\quad\; h_j(\hat{x}) = 0 \qquad j \in E \end{aligned} \tag{3.1.65}$$

$$(2) \qquad \nabla f(\hat{x}) - \sum_{i\in I} \hat{\mu}_i \nabla g_i(\hat{x}) - \sum_{j\in E} \hat{\lambda}_j \nabla h_j(\hat{x}) = 0 \tag{3.1.66}$$

$$(3) \qquad \hat{\mu}_i \geq 0 \qquad i \in I \tag{3.1.67}$$

$$(4) \qquad \hat{\mu}_i g_i(\hat{x}) = 0 \tag{3.1.68}$$

Bewijs : De voorwaarden zijn juist dezelfde als de voorwaarden in Stellingen 3.1.17 en 3.1.18 indien men voor de waarden $\hat{\mu}_i$ voor indices $i \in I_p(\hat{x})$ kiest

$$\hat{\mu}_i := 0 \qquad i \in I_p(\hat{x})$$

welke keuze in overeenstemming is met de 4^e voorwaarde (3.1.68). □

3.1.21. Bij de voldoende voorwaarden voor een oplossing van het probleem GNLI treedt in vergelijking met de voldoende voorwaarden voor een oplossing van het probleem GNLE een complicatie op indien een of meerdere van de Lagrange-multiplicatoren $\hat{\mu}_i$ die corresponderen met actieve ongelijk-

heids beperkingen gelijk zijn aan nul. In een dergelijke situatie doet de betreffende beperking bij de eerste orde voorwaarden in het geheel niet mee en komt ook niet terecht in de Lagrange-functie (men spreekt in dit geval over gedegenereerde ongelijkheidsbeperkingen). Bij de tweede orde beperkingen dient wel rekening gehouden te worden met de betreffende beperking. Hiervoor kan worden gezorgd door weglating van de betreffende beperking bij de definitie van de verzameling $M(\hat{x})$ (3.1.57). Dit is weergegeven in de met Stelling 3.1.20 corresponderende uitspraak.

Stelling 3.1.21 (vgl. [3.1.1]): Voldoende voor het optreden van een oplossing van het probleem GNLI (3.1.1) in een punt $\hat{x}$ is het bestaan van Lagrange-multiplicatoren $\hat{\mu}_i$, $i \in I$ en $\hat{\lambda}_j$, $j \in E$ zodat geldt

$$(1) \qquad g_i(\hat{x}) \geq 0 \qquad i \in I$$

$$h_j(\hat{x}) = 0 \qquad j \in E$$

$$(2) \quad \nabla f(\hat{x}) - \sum_{i\in I} \hat{\mu}_i \nabla g_i(\hat{x}) - \sum_{j\in E} \hat{\lambda}_j \nabla h_j(\hat{x}) = \nabla_x L(\hat{x},\hat{\lambda},\hat{\mu}) = 0$$

$$(3) \qquad \hat{\mu}_i \geq 0$$

$$(4) \qquad \hat{\mu}_i g_i(\hat{x}) = 0$$

en

$$(5) \quad \forall z \in M_{II}(\hat{x}) : z^T[G(\hat{x}) - \sum_{i\in I} \hat{\mu}_i G_i(\hat{x}) - \sum_{j\in E} \hat{\lambda}_j H_j(\hat{x})]\, z > 0 \qquad (3.1.69)$$

waar

$$M_{II}(\hat{x}) := \{z \in \mathbb{R}^n \mid \nabla^T g_i(\hat{x}) z = 0 \ , \ i \in J \ , \ \nabla^T h_j(\hat{x}) z = 0 \ j \in E\} \qquad (3.1.70)$$

en

$$J := \{i \in I_A(\hat{x}) \mid \hat{\mu}_i > 0\} \qquad (3.1.71)$$

Bewijs: Voor een bewijs zij verwezen naar [3.1.1]. □

Methoden

3.1.22. Voor de numerieke oplossing van niet-lineaire minimaliseringsproblemen met nevenvoorwaarden bestaan een groot aantal methoden, waarvan de meeste kunnen worden gerekend tot een van de volgende drie klassen van methoden:

(a) primale methoden
(b) boetefunctie methoden
(c) duale methoden

Deze drie klassen betreffen iteratieve methoden die in iedere stap een betere benadering van de oplossing genereren. Bij de primale methoden wordt deze betere benadering gegenereerd door een een- dimensionaal zoekproces in $\mathbb{R}^n$ zoals bij de in het voorgaande hoofdstuk besproken methoden voor onbeperkte minimalisering, met dien verstande dat de zoekrichtingen en het zoekproces zijn aangepast aan de beperkingen. Bij de boetefunctiemethoden wordt de betere benadering gegenereerd als de oplossing van een in iedere iteratiestap ander onbeperkt minimaliseringsprobleem met als object-functie de originele objectfunctie aangevuld met een aantal boetefunctietermen. Deze laatsten nemen toe in waarde naarmate de overschrijdingen van de beperkingen groter worden en naarmate de in iedere stap verhoogde gewichtsfactoren groeien. Bij de duale methoden worden in iedere stap betere benaderingen gegenereerd van zowel de oplossing zelf als van de Lagrange-multiplicatoren in het optimum. Dit geschiedt juist als bij de boetefunctiemethoden door de oplossing van een in iedere stap verschillend onbeperkt minimaliseringsprobleem. Het verschil tussen de boetefunctie methoden en de duale methoden ligt voornamelijk daarin dat de exacte oplossing van het originele probleem bij de duale methoden gevonden kan worden zonder het gebruikelijke limietproces (gewichtsfactoren oneindig groot) van de boetefunctie methoden. In het resterende deel van dit hoofdstuk zal aan ieder van deze drie klassen van methoden speciale aandacht worden besteed.

3.1.23. Naast de genoemde methoden voor algemene minimaliseringsproblemen met nevenvoorwaarden zijn er nog een aantal zeer bekende methoden voor speciale problemen met nevenvoorwaarden. Tot die categorie behoren o.a. :

(a) kwadratische-programmeringsmethoden (zie [3.1.7])
(b) geometrische - programmeringsmethoden (zie [3.1.6])
(c) lineaire-approximatiemethoden (zie [3.1.5])
(d) "cutting plane" methoden (zie [3.1.1] en [3.1.3])
(e) transformatie methoden

De eerste vier van deze categorieën van methoden zijn voor een belangrijk deel gebaseerd op (de ideeën achter) lineaire programmeringsmethoden en passen als zodanig niet helemaal in het kader van deze syllabus. Voor verdere informatie wordt de lezer daarom verwezen naar de aangegeven literatuur. Tot de vijfde categorie van tranformatie-methoden worden die methoden gerekend waarbij door variabelen- of coördinaten transformaties beperkingen worden geëlimineerd. Deze methoden zijn alleen mogelijk bij beperkingen met een simpele (b.v. lineaire) structuur zoals bijvoorbeeld bij coördinaat-beperkingen van het type

$$x_i \geq a \tag{3.1.72}$$

en

$$b \leq x_i \leq c \tag{3.1.73}$$

In het eerste geval (eenzijdige ongelijkheidsbeperkingen) kan de beperking (3.1.72), bijvoorbeeld, worden geëlimineerd door vervanging van de beperkte variabele x_i, door de niet beperkte variabele y_i gedefinieerd door

$$x_i = : y_i^2 + a \tag{3.1.74}$$

Analoog kan in het tweede geval (tweezijdige ongelijkheidsbeperkingen) x_i worden vervangen door de onbeperkte variabele z_i gedefinieerd door

$$x_i = :(\frac{b+c}{2}) + (\frac{b-c}{2}) \cos z_i \tag{3.1.75}$$

Uiteraard zijn op deze manier diverse transformaties mogelijk. Het voornaamste op dit punt is,deze in de practijk dikwijls bruikbare mogelijkheid voor de eliminatie van een of meerdere beperkingen te signaleren.

Referenties

3.1.24 Meer informatie over de in deze paragraaf besproken onderwerpen kan worden gevonden in de volgende publicaties.

[3.1.1] : Zie [1.1.1] Luenberger (1973)

[3.1.2] : Zie [1.1.4] Gill & Murray (1974)

[3.1.3] : Zie [2.1.3] Zangwill (1969)

[3.1.4] : Zie [2.4.11] Mangasarian (1969)

[3.1.5] : Zie [2.9.8] Himmelblau (1972)

[3.1.6] : Avriel, M., Rijckaert, M.J. and Wilde, D.J. (Eds) : "Optimatization and design", Proc. Int. Summer School on the Impact of Optimization Theory on Technological Design, Leuven, July 26 - Aug. 6, 1971, Prentice Hall, Inc. London, 1973.

[3.1.7] : Van de Panne, C., Methods for linear and quadratic programming, North Holland Publ. Cy, Amsterdam, 1975.

§ 3.2. Primale methoden I: Eerste orde methoden en lineaire beperkingen

3.2.1. Tot de primale methoden voor het oplossen van minimaliseringsproblemen met nevenvoorwaarden worden al die iteratieve methoden gerekend die uitgaande van een benadering voor de oplossing $x^{(k)}$ met behulp van een één-dimensionaal zoekproces een nieuwe benadering $x^{(k+1)}$ bepalen die ligt in het toegelaten gebied S

$$x^{(k+1)} \in S$$

en die tegelijk voldoet aan

$$f(x^{(k+1)}) < f(x^{(k)})$$

De primale methoden zijn in de meeste gevallen directe aanpassingen van de in het voorgaande hoofdstuk besproken methoden voor onbeperkte minimalisering voor het geval dat er nevenvoorwaarden zijn. Men onderscheidt in dat verband dan ook twee categorieën van primale methoden (vgl. pt. 2.1.4)

a) aangepaste "direct-search"-methoden
b) aangepaste descent-methoden

De aangepaste "direct-search"-methoden zijn meestal heuristisch van opzet en dragen een ad-hoc karakter. Zij maken vaak in onderdelen en lokaal gebruik van dezelfde soort technieken als de algemenere aangepaste-descent- en boetefunctie-methoden en lenen zich slecht voor een korte algemene bespreking . Zij worden daarom in deze syllabus niet verder behandeld. Geïnteresseerden worden verwezen naar de literatuur(b.v. [3.2.18]). De aangepaste-descent-methoden maken in principe gebruik van dezelfde algorithme als de descent-methoden voor onbeperkte minimalisering (vgl. pt. 2.1.5)

$$x^{(k+1)} := x^{(k)} + \alpha^{(k)} d^{(d)} \qquad (3.2.1)$$

met dat verschil dat zowel de zoekrichting $d^{(k)}$ als de stapgrootte $\alpha^{(k)}$ worden aangepast aan de beperkingen.

3.2.2. De aangepaste descent-methoden wijken af van de descent-methoden voor onbeperkte minimalisering op twee essentiële punten:

a) voordat met zoeken begonnen wordt moet eerst een toegelaten start-punt worden bepaald
b) tijdens het zoekproces moet er voor worden gezorgd dat het volgende punt weer toegelaten is.

Het verzorgen van deze punten in practische algorithmen verschilt naar gelang de nevenvoorwaarden uitsluitend lineair zijn dan wel of er ook niet-lineaire nevenvoorwaarden zijn: In het geval van uitsluitend lineaire nevenvoorwaarden kan voor het vinden van een eerste toegelaten punt gebruik worden gemaakt van de z.g. Phase-I-methode uit de lineaire programmering (zie [3.2.1] en [3.2.12]) naast de hieronder te bespreken methoden voor het vinden van toegelaten punten in het geval van algemene beperkingen. Bij het genereren en het handhaven van het toegelaten zijn van opvolgende punten speelt het lineair zijn van de beperkingen nog een grotere rol. In het geval van uitsluitend lineaire beperkingen is het namelijk voldoende om de zoekrichting en de stapgrootte aan te passen bij toepassing van de algorithme (3.2.1). In het geval van niet-lineaire beperkingen treden complicaties op die vragen om extra maatregelen. Deze laatsten zullen worden besproken in paragraaf 3.3.

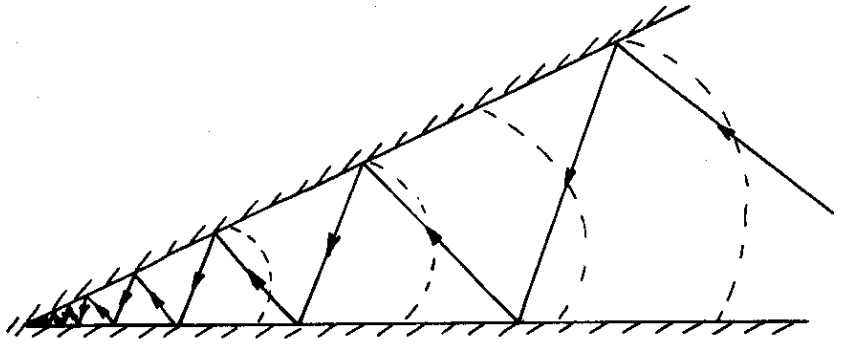
3.2.3. Bij het aanpassen van de zoekrichting en de stapgrootte (in (3.2.1)) voor het toegelaten houden van opvolgende iteratiepunten maakt men gewoonlijk gebruik van een strategie, die op het eerste gezicht niet voor de hand ligt, en die bekend staat als de actieve-set-strategie. Deze strategie houdt in dat bij het bepalen van een toegelaten zoekrichting alle actieve beperkingen in het uitgangspunt als gelijkheidsbeperkingen worden geïnterpreteerd. De verzameling van actieve beperkingen wordt uitgebreid zodra in het zoekproces een eerdere passieve beperking actief wordt en, andersom, met één (en niet meer dan een) beperking verminderd in het geval dat het minimum is gevonden over de tot dusver actief veronderstelde beperkingen. De (ongelijkheids-) beperking die de verzameling van actieve beperkingen in dat geval verlaat wordt bepaald aan de hand van het criterium of de corresponderende Lagrange-multiplicator negatief is (vgl. (3.1.63)) d.i. als

$$\mu_i := (n_i^+)^T \nabla f < 0 \qquad (3.2.2)$$

In dat geval kan een zoekrichting d met een component in de richting van betreffende vector n_i^+ tot een verlaging van de objectfunctiewaarde leiden met gelijktijdige handhaving van het actief blijven van de overige actieve beperkingen. Immers (vgl. (3.1.64))

$$n_j^T n_i^+ = 0 \qquad i \neq j \ .$$

De reden voor het hanteren van de actieve-set-strategie is gelegen in het tegengaan van het bij vroegere onderzoekingen geobserveerde "zig-zagging"- of "jamming" verschijnsel (zie [3.2.4], [3.2.16]), dat optreedt indien tijdens het zoekproces telkens op en neer wordt gesprongen tussen twee of



Figuur 3.2.3.: Het zig-zagging-verschijnsel

meer beperkingen. Niet alleen wordt hierdoor de convergentiesnelheid aangetast, er zijn voorbeelden bekend ([3.2.16] p. 19) waarbij convergentie optreedt naar een punt dat geen oplossing is. De actieve-set-strategie maakt het onmogelijk dat het zoekproces (in dezelfde configuratie) terugkeert naar een eerder verlaten beperking.

Standaard algorithme primale methoden

3.2.4. Met de incorporatie van de voorzieningen nodig voor het genereren van eerste en volgende toegelaten punten en bij gebruik van de actieve-set-strategie krijgt de standaard-algorithme van de primale methoden de volgende vorm (vgl. pt. 1.1.6 en 2.1.5):

(0) kies een startpunt $\bar{x}^{(0)}$, zet k:= 0

(i) met $\bar{x}^{(k)}$ als uitgangspunt bepaal een toegelaten punt $x^{(k)}$ (dit wordt "restauratie" genoemd), bepaal welke beperkingen actief zijn in $x^{(k)}$ en evalueer de corresponderende normalen $n_i := \nabla c_i(x^{(k)})$

(ii) bepaal de functiewaarde $f(x^{(k)})$ en de gradiënt $\nabla f(x^{(k)})$ in $x^{(k)}$

(iii) bepaal de met de actieve beperkingen in $x^{(k)}$ corresponderende matrix van normalen $N^{(k)}$ en de restrictie van de gradiënt tot het daarmee corresponderende raakvlak aan de actieve beperkingen, d.i. òf de geprojecteerde gradiënt (vgl. pt. 3.2.10)

$$\bar{P}^{(k)}\nabla f^{(k)} := (I - N^{(k)}(N^{(k)T}N^{(k)})^{-1}N^{(k)T})\nabla f(x^{(k)})$$

òf de gereduceerde gradiënt (vgl. pt. 3.2.13) (die met weglating van de index (k) als in (3.2.41) gegeven wordt door)

$$\bar{R}\nabla f := \begin{bmatrix} B^{-1}DD^{T}B^{-T} & -B^{-1}D \\ -D^{T}B^{-T} & I \end{bmatrix} \begin{bmatrix} \nabla_{x_B} f \\ \nabla_{x_D} f \end{bmatrix}$$

(iv) ga na of het punt $x^{(k)}$ het optimale punt is in de huidige actieve set; zo ja, dan evalueer het teken van de Lagrange-multiplicatoren corresponderend met de ongelijkheidsbeperkingen

$$\mu_i := (n_i^{+})^{(k)T}\nabla f(x^{(k)}) \tag{3.2.3}$$

als alle μ_i positief (of gelijk aan nul) dan klaar; als een of meerdere μ_i negatief dan ga terug naar stap (iii); als het punt $x^{(k)}$ niet optimaal is in de huidige actieve set, dan ga door:

(v) bepaal een nieuwe zoekrichting $d^{(k)}$

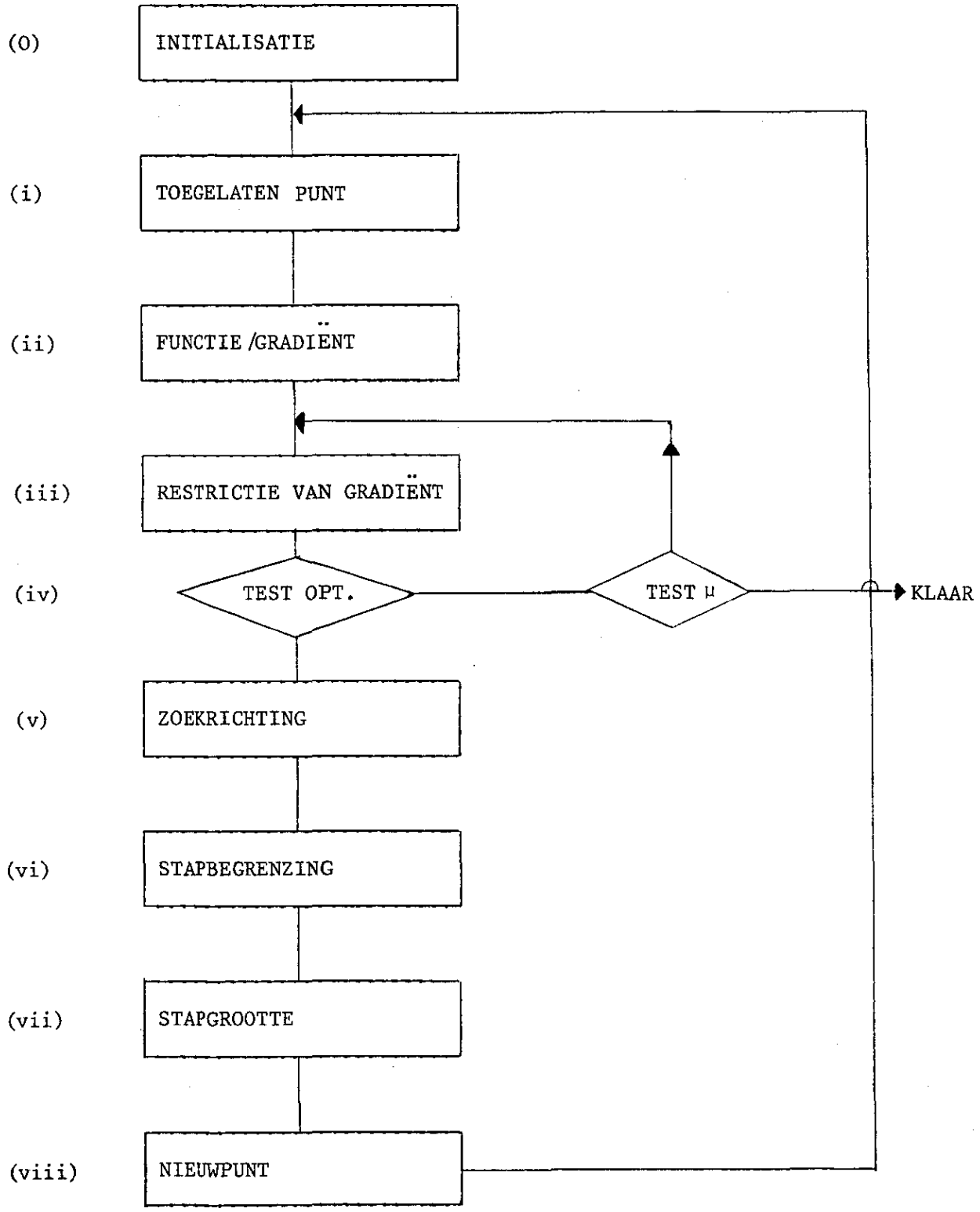
(vi) bepaal een maximale stapgrootte $\bar{\alpha}^{(k)}$

(vii) bepaal een stapgrootte $\alpha^{(k)}$ met $0 < \alpha^{(k)} \leq \bar{\alpha}^{(k)}$ zodat

$$f(x^{(k)} + \alpha^{(k)}d^{(k)}) < f(x^{(k)}) \tag{3.2.4}$$

(viii) zet $\bar{x}^{(k+1)} := x^{(k)} + \alpha^{(k)} d^{(k)}$, k:= k + 1 en ga terug naar stap (i).

Een flowdiagram van deze algorithme geeft het in Figuur 3.2.4 geschetste beeld. In het navolgende zal in het bijzonder op de stappen (i), (v) en (vi) nader worden ingegaan



Figuur 3.2.4.: Flowdiagram van de standaard algorithme voor primale methoden.

Bepaling van een toegelaten startoplossing (stap (i))

3.2.5. Zoals reeds opgemerkt in pt. 3.2.2. kan in het geval van uitsluitend lineaire beperkingen in de probleemformulering gebruik worden gemaakt van de uit de lineaire programmering bekende Phase-I-methode (zie b.v.[3.2.1]). In het geval van een probleemformulering van de vorm GLI (3.1.21)

$$\min\{f(x) \mid A_1^T x - b_1 \geq 0 \ , \ A_2^T x - b_2 = 0\}$$

bestaat de Phase-I-methode daaruit dat de beperkingen met behulp van "slack" variabelen $y_i \geq 0$ en met de definitie van niet-negatieve variabelen x^+ en x^- volgens

$$x =: x^+ - x^- \tag{3.2.5}$$

eerst wordt herschreven in standaard LP-vorm (met $b_1 \geq 0$ en $b_2 \geq 0$)

$$\min\left\{ f(x) \;\middle|\; \begin{array}{l} A_1^T x^+ - A_1^T x^- + y = b_1 \ , \ A_2^T x^+ - A_2^T x^- = b_2 \\ x^+ \geq 0 \ , \ x^- \geq 0 \ , \ y \geq 0 \end{array} \right\} \tag{3.2.6}$$

en dat vervolgens een oplossing wordt gezocht van de som van de in te voeren artificiële variabelen

$$\min\left\{ \bar{c}_1^T z_1 + \bar{c}_2^T z_2 \;\middle|\; \begin{array}{l} A_1^T x^+ - A_1^T x^- + y_1 + z_1 = b_1 \quad A_2^T x^+ - A_2^T x^- + z_2 = b_2 \\ x^+ \geq 0 \quad x^- \geq 0 \ , \ y_1 \geq 0 \ , \ z_1 \geq 0 \ , \ z_2 \geq 0 \end{array} \right\}$$

(3.2.7)

waar voor $i = 1,2$ $\bar{c}_i = (1,1,\ldots,1)$. Dit laatste probleem (3.2.7) is een zuiver LP probleem dat met de simplexmethode kan worden opgelost.

3.2.6. Een tweede algemene methode voor het bepalen van een eerste toegelaten oplossing, welke ook kan worden gebruikt bij problemen met niet-lineaire beperkingen, d.i. problemen van het type GNLI (3.1.1)

$$\min\{ f(x) \mid g_i(x) \geq 0 \ , \ i \in I \ , \ h_j(x) = 0, \ j \in E\}$$

is de methode die gebaseerd is op het minimaliseren van een aan de beperkingen gerelateerde boetefunctie. Vooruitlopend op de behandeling van de boetefunctie-

methoden in paragraaf 3.4. kan worden opgemerkt dat een toegelaten oplossing van het probleem GNLI bepaald kan worden als oplossing van het onbeperkte minimaliseringsprobleem

$$\min \{ \sum_{i\in I} k_i \{ \min [0, g_i(x)] \}^2 + \sum_{j\in E} k_j h_j^2(x) \mid x \in \mathbb{R}^n \} \qquad (3.2.8)$$

in welke uitdrukking k_i en k_j positieve gewichtsfactoren voorstellen. Op zichzelf is deze boetefunctie aanpak voor het bepalen van een eerste toegelaten punt zeer betrouwbaar en algemeen toepasbaar. Een nadeel van de methode in algorithmen voor primale methoden is dat de aanpak voor het bepalen van het eerste toegelaten punt dan meestal wezenlijk verschillend is van de aanpak voor het bepalen van volgende toegelaten punten.

3.2.7. Naast de in voorgaande punten genoemde, algemene methoden bestaan een groot aantal heuristische methoden voor het bepalen van het eerste toegelaten punt. Een voorbeeld daarvan is de methode beschreven door Dirkx [3.2.13] . Deze methode is in essentie hetzelfde als de in het voorgaande punt besproken boetefunctiemethode: De methode bestaat daaruit dat in een aantal opvolgende stappen telkens de minimum-norm kleinste-kwadraten oplossing van het stelsel lineaire vergelijkingen gegenereerd door linearisatie van de lokale actieve en overschreden beperkingen in het laatst gevonden punt wordt bepaald. Wordt dit stelsel

$$\begin{aligned} &g_i(\bar{x}^{(k)}) + \nabla^T g_i(\bar{x}^{(k)})(x - \bar{x}^{(k)}) = 0 \quad i \in I_V(\bar{x}^{(k)}) \cup I_A(\bar{x}^{(k)}) \\ &h_j(\bar{x}^{(k)}) + \nabla^T h_j(\bar{x}^{(k)})(x - \bar{x}^{(k)}) = 0 \quad j \in E \end{aligned} \qquad (3.2.9)$$

herschreven in de algemene vorm

$$A^T \Delta x - b = 0 \qquad (3.2.10)$$

dan kan de minimum-norm kleinste-kwadraten oplossing met behulp van de pseudo-inverse $(A^T)^+$ van de matrix A^T (vgl. pt. 2.10.17) worden weergegeven als

$$\Delta\hat{x} = (A^T)^+ b \qquad (3.2.11)$$

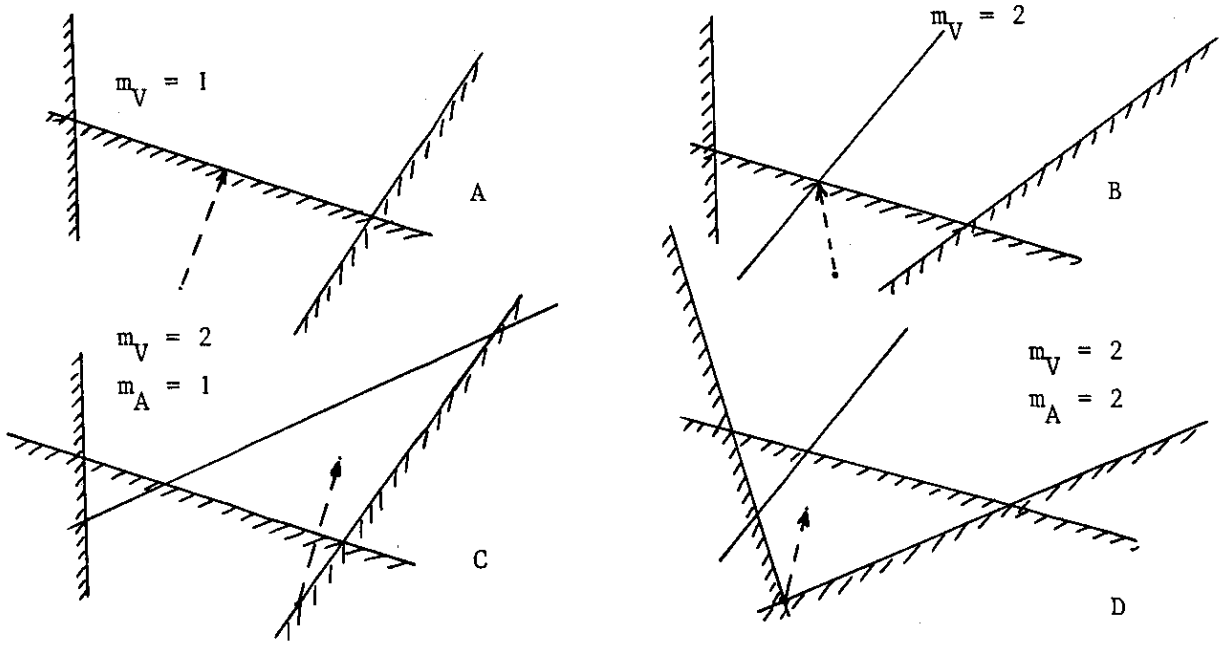
Als A^T een (m x n)-matrix is en verondersteld wordt dat A^T maximale rang bezit dan volgt bij uitwerking van de pseudo-inverse in het geval dat $m < n$ (onder-bepaald stelsel) dat

$$\Delta\hat{x} = (A^T)^+ b = A(A^T A)^{-1} b \tag{3.2.12}$$

en in het geval $m \geq n$ (overbepaald stelsel) dat

$$\Delta\hat{x} = (A^T)^+ b = (A^T A)^{-1} Ab \tag{3.2.13}$$

In het eerste geval is de correctie een lineaire combinatie van de gradiënten, in het tweede geval is de correctie zodanig dat slechts in kleinste-kwadraten zin aan de beperkingen wordt voldaan. Ten grondslag aan het iteratieve restauratieproces ligt de veronderstelling dat er na herhaalde bepaling van de minimum-norm kleinste-kwadraten oplossing een regulier of onderbepaald stelsel overblijft. De werking van de restauratie-procedure is geïllustreerd in Figuur 3.2.7. waarin voor het geval van $\mathbb{R}^2$ enkele verschillende mogelijkheden met betrekking tot overschreden (m_V) en actieve (m_A) beperkingen zijn weergegeven.



Figuur 3.2.7. : Voorbeelden van de werking van de restauratie-procedure van Dirkx [3.2.13]

3.2.8. Naast de in het voorgaande punt beschreven methoden voor het bepalen van een eerste toegelaten punt zijn nog talloze andere varianten te bedenken. In het geval van lineaire beperkingen, bijvoorbeeld, kan bij toepassing van het idee van de Phase-I-methode het minimaliseren van de artificiële objectfunctie worden uitgevoerd met behulp van een primale methode in plaats van de simplex methode (zie [3.2.19]). Ook kan gebruik gemaakt worden van een ad-hoc methode waarbij telkens de meest nabij gelegen beperking van overschreden actief wordt gemaakt (zie [3.2.11]). In principe zijn er in dit verband mogelijkheden (tot improvisatie) te over, reden waarschijnlijk waarom er in de literatuur nauwelijks aandacht aan dit punt wordt besteed. Dit laatste is tamelijk ongelukkig omdat het vinden van een eerste toegelaten punt in de practijk vaak het "goede begin" is dat gelijk is aan het "halve werk".

Bepaling van een nieuwe zoekrichting (stap (v))

3.2.9. Gegeven een toegelaten punt en gegeven de in pt. 3.2.3 besproken actieve-set-strategie, volgens welke alle actieve beperkingen verondersteld worden actief te blijven, is het mogelijk zoekrichtingen te genereren op basis van dezelfde ideeën als bij onbeperkte minimaliseringsproblemen met dien verstande dat rekening moet worden gehouden met de gelineariseerde (of lineaire) beperkingen in het toegelaten punt: Uitgangspunt voor de generatie van zoekrichtingen voor de meeste primale methoden is het van het algemene probleem GNLI (3.1.1)

$$\min \{ f(x) \mid g_i(x) \geq 0 \ , \ i \in I \ , \ h_j(x) = 0, j \in E \}$$

afgeleide probleem van het type GLI (3.1.21)

$$\min \left\{ f(x) \;\middle|\; \begin{array}{l} \nabla^T g_i(x^{(k)}) \ (x - x^{(k)}) = 0 \ , \ i \in I_A(x^{(k)}) \\ \nabla^T h_j(x^{(k)}) \ (x - x^{(k)}) = 0, \quad j \in E \quad \text{en} \\ \nabla^T g_i(x^{(k)}) \ (x - x^{(k)}) \geq -g_i(x^{(k)}), i \in I_p(x^{(k)}) \end{array} \right\} \qquad (3.2.14)$$

dat er na herschrijven uit ziet als

$$\min \{f(x) \mid N^{(k)T}(x - x^{(k)}) = 0 \ , \ \tilde{A}^{(k)T}(x - x^{(k)}) \geq - \tilde{c}^{(k)}\} \quad (3.2.15)$$

Hierin stellen $N^{(k)}$ en $\tilde{A}^{(k)}$ respectievelijk de matrix van normalen van actieve beperkingen en de matrix van de gradiënten van de passieve beperkingen voor en is $-\tilde{c}^{(k)}$ de corresponderende vector van de waarden $g_i(x^{(k)})$ (voorzien van een min-teken) van de passieve beperkingen in het punt $x^{(k)}$.

a) Geprojecteerde gradiënt methode

3.2.10. Een eerste orde zoekrichting voor het probleem (3.2.15) kan juist als in het geval van onbeperkte minimalisering (pt. 2.4.2) worden gevonden als de oplossing van het probleem van de bepaling van het minimum van de locaal gelineariseerde objectfunctie binnen een (hyper-) bol met een (kleine) straal δ, d.i. het probleem

$$\min \{f(x^{(k)}) + \nabla^T f(x^{(k)}) \ (x - x^{(k)}) \mid N^{(k)T}(x - x^{(k)}) = 0 \ ,$$

$$\| x - x^{(k)} \|^2 \leq \delta^2\} \quad (3.2.16)$$

Met de Kuhn-Tucker voorwaarden (Stelling 3.1.20) voor dit probleem

$$\nabla_x L(x,\lambda,\mu) = \nabla f(x^{(k)}) - N^{(k)}\lambda + 2\mu(x - x^{(k)}) = 0$$

$$N^{(k)T}(x - x^{(k)}) = 0$$

$$\delta^2 - (x - x^{(k)})^T \ (x - x^{(k)}) \geq 0 \quad (3.2.17)$$

$$\mu \geq 0$$

$$\mu(\delta^2 - (x - x^{(k)})^T \ (x - x^{(k)})) = 0$$

volgt voor die oplossing (vgl. (2.4.4))

$$(\hat{x} - x^{(k)}) = - \delta \frac{\bar{P}^{(k)} \nabla f(x^{(k)})}{\| \bar{P}^{(k)} \nabla f(x^{(k)}) \|} \quad (3.2.18)$$

waar

$$\bar{P}^{(k)}\nabla f(x^{(k)}) := (I - N^{(k)}(N^{(k)T}N^{(k)})^{-1}N^{(k)T})\nabla f(x^{(k)}) \qquad (3.2.19)$$

Deze oplossing is dus juist het punt op de (hyper) bol in de richting van de geprojecteerde gradiënt $\bar{P}^{(k)}\nabla f(x^{(k)})$, d.i. de projectie van de gradiënt op het orthogonale complement van de ruimte opgespannen door de normalen van de actieve beperkingen (vgl. pt. 3.1.10). Als zoekrichting volgt daaruit

$$\begin{aligned} d^{(k)} &:= -(I - N^{(k)}(N^{(k)T}N^{(k)})^{-1}N^{(k)T})\nabla f(x^{(k)}) \\ &:= -\bar{P}^{(k)}\nabla f(x^{(k)}) \end{aligned} \qquad (3.2.20)$$

De primale methode die gebaseerd is op het gebruik van deze zoekrichting staat bekend als de geprojecteerde-gradiënt-methode. Deze methode werd in 1960 voor het eerst gepropageerd door Rosen ([3.2.7]) die tevens een algorithme ontwikkelde waarmee in het geval van uitsluitend lineaire beperkingen bij verandering van de verzameling van actieve beperkingen met weinig rekenwerk een nieuwe projectiematrix $\bar{P}^{(k+1)}$ kan worden berekend. Voor een beschrijving van deze algorithme kan worden verwezen naar [3.2.2] of [3.2.5].

3.2.11. De geprojecteerde gradiënt (3.2.20) kan ook langs een andere, vooral voor latere ontwikkelingen belangrijke weg worden afgeleid als zoekrichting. Uitgangspunt daarbij is de restrictie van de objectfunctie tot het raakvlak aan de actieve beperkingen. In het geval $N^{(k)}$ de matrix van actieve normalen in $x^{(k)}$ voorstelt en $Z^{(k)}$ de daarmee corresponderende, in pt. 3.1.8 gedefinieerde matrix van orthonormale basisvectoren, waarvoor (vgl. (3.1.26) en (3.1.25))

$$Z^{(k)T}Z^{(k)} = I_{(n-m)\times(n-m)} \quad \text{en} \quad N^{(k)T}Z^{(k)} = 0_{m\times(n-m)} \qquad (3.2.21)$$

dan is dit raakvlak juist de lineaire variëteit of nevenruimte

$$T(x^{(k)}) := \{x \in \mathbb{R}^n \mid x = x^{(k)} + Z^{(k)}w,\ w \in \mathbb{R}^{n-m}\} \qquad (3.2.22)$$

De restrictie van de objectfunctie tot dit raakvlak is een functie van de vector $w \in \mathbb{R}^{n-m}$

$$\varphi^{(k)}(w) := f(x^{(k)} + Z^{(k)}w) \tag{3.2.23}$$

en minimalisering van de functie f(x) onder de restrictie $N^{(k)T}(x - x^{(k)}) = 0$ is equivalent met de onbeperkte minimalisering van de functie $\varphi^{(k)}(w)$. Toepassing van de gradiënt methode (vgl. pt. 2.4.1) voor dit probleem geeft als zoekrichting in de w-ruimte de negatieve gradiënt

$$d_w^{(k)} := -\nabla_w \varphi^{(k)}(0) := -Z^{(k)T}\nabla f(x^{(k)}) \tag{3.2.24}$$

en corresponderend daarmee in de originele (x-) ruimte

$$d^{(k)} := Z^{(k)} d_w^{(k)} := -Z^{(k)}Z^{(k)T}\nabla f(x^{(k)}) \tag{3.2.25}$$

Zoals in pt. 3.1.8 aangetoond geldt dat (vgl. (3.1.29))

$$Z^{(k)}Z^{(k)T} = (I - N^{(k)}(N^{(k)T}N^{(k)})^{-1}N^{(k)T})$$

waarmee volgt dat de laatst gevonden zoekrichting (3.2.25) juist gelijk is aan de negatieve geprojecteerde gradiënt (3.2.20).

3.2.12. Een veel toegepaste mogelijkheid om in practische situaties de matrix $Z^{(k)}$ te bepalen is de toepassing van QR-decompositie van de matrix $N^{(k)}$: Als $Q^{(k)}$ de orthonormale n x n-matrix is die kan worden gepartitioneerd in een n x m-matrix $Q_1^{(k)}$ en een n x (n-m)-matrix $Q_2^{(k)}$ en $R^{(k)}$ een m x m-bovendriehoeksmatrix is zo dat geldt dat (vgl. pt. 2.10.6)

$$N^{(k)} := Q^{(k)}R^{(k)} = \left[Q_1^{(k)} \,\vdots\, Q_2^{(k)}\right] \begin{bmatrix} R^{(k)} \\ \hline 0 \end{bmatrix} \tag{3.2.26}$$

dan kan voor de matrix $Z^{(k)}$ gekozen worden

$$Z^{(k)} = Q_2^{(k)} \tag{3.2.27}$$

Er geldt immers

$$N^{(k)T}Q_2^{(k)} = R^{(k)T}Q_1^{(k)T}Q_2^{(k)} = O_{m \times (n-m)}$$

en (3.2.28)

$$Q_2^{(k)T}Q_2^{(k)} = I_{(n-m) \times (n-m)}$$

De laatst gevonden zoekrichting (3.2.25) kan daarmee worden bepaald als

$$d^{(k)} := -Q_2^{(k)}Q_2^{(k)T}\nabla f(x^{(k)}) \tag{3.2.29}$$

Met de uit de orthogonaliteit volgende relatie

$$I_{n\times n} = Q^{(k)}Q^{(k)T} = Q_1^{(k)}Q_1^{(k)T} + Q_2^{(k)}Q_2^{(k)T} \tag{3.2.30}$$

volgt dat deze zoekrichting (3.2.29) ook bepaald had kunnen worden uit

$$d^{(k)} := -(I - Q_1^{(k)}Q_1^{(k)T})\nabla f(x^{(k)}) \tag{3.2.31}$$

welke uitdrukking ook volgt bij substitutie van de QR-decompositie in (3.2.20). De equivalentie van de in pt. 3.2.10 en in pt. 3.2.11 gevonden zoekrichtingen is daarmee opnieuw geverifieerd. Ter afsluiting van dit punt zij nog opgemerkt dat de in (3.2.26) bedoelde QR-decompositie in de practijk eenvoudig kan worden gerealiseerd met behulp van Householder-transformaties. Voor de details daarvan moet worden verwezen naar de literatuur.(bv. [3.2.6])

b) Gereduceerde gradiënt methode

3.2.13. Een andere eerste orde primale methode die naast de geprojecteerde gradiënt methode grote bekendheid geniet is de gereduceerde-gradiënt-methode. Deze methode die voor het eerst gepropageerd werd door Wolfe [3.2.20] in 1962 en die vooral verder werd ontwikkeld door Abadie en zijn medewerkers ([3.2.8], [3.2.9] [3.2.10]) is gebaseerd op dezelfde ideeën als waarop de simplex-methode voor LP problemen berust. Uitgangspunt voor de afleiding van de methode is het door locale lineari-

satie van de beperkingen van het algemene probleem GNLI (3.1.1) afgeleide en herschreven (zie pt. 3.2. 9) probleem (3.2.15) van het type GLI

$$\min \{f(x) \mid N^{(k)T}(x - x^{(k)}) = 0,\ \tilde{A}^{(k)T}(x - x^{(k)}) \geq -\tilde{c}^{(k)}\}$$

Voor het bepalen van een zoekrichting wordt dit probleem, juist als in het voorgaande dan nog verder vereenvoudigd tot

$$\min \{f(x) \mid N^{(k)T}x = N^{(k)T}x^{(k)} =: b^{(k)}\} \tag{3.2.32}$$

Het basisidee dat aan de gereduceerde-gradiënt methode ten grondslag ligt is juist als bij de simplex methode de overweging dat de lineaire gelijkheidsbeperkingen kunnen worden gebruikt voor het elimineren of, equivalent, het afhankelijk maken van evenzoveel (=m) variabelen als er lineaire vergelijkingen zijn. Om dit te formaliseren wordt de matrix $N^{(k)T}$ gesplitst in een reguliere (m x m)-basis matrix $B^{(k)}$ en een m x (n-m)-niet-basis matrix $D^{(k)}$ en worden daarmee corresponderend de componenten van de vector x gesplitst in basis variabelen x_B en niet-basisvariabelen x_D zodat de gelijkheidsbeperkingen in (3.2.32)

$$N^{(k)T}\, x = b^{(k)}$$

overgaan in

$$B^{(k)}x_B + D^{(k)}x_D = b^{(k)} \tag{3.2.33}$$

Met de veronderstelling dat $B^{(k)}$ niet-singulier is kan x_B direct uit deze vergelijking worden opgelost als

$$x_B = (B^{(k)})^{-1}b^{(k)} - (B^{(k)})^{-1}D^{(k)}x_D \tag{3.2.34}$$

Dit resultaat geeft de mogelijkheid de afhankelijke m-componenten-vector x_B te elimineren waardoor het uitgangsprobleem (3.2.32) met weglating van de index (k)

$$\min \{f(x) \mid Bx_B + Dx_D = b\} \tag{3.2.35}$$

overgaat in het onbeperkte minimaliseringsprobleem met als variabelen de onafhankelijke componenten van x_D

$$\min \{\varphi(x_D) \mid \varphi(x_D) = f(B^{-1}b - B^{-1}Dx_D\ ,\ x_D)\} \tag{3.2.36}$$

De gradiënt van deze functie m.b.t. de (n-m)-vector x_D

$$\nabla_{x_D}\varphi = -D^T B^{-T}\nabla_{x_B} f + \nabla_{x_D} f \tag{3.2.37}$$

wordt om voor de hand liggende reden aangeduid als de <u>gereduceerde gradiënt</u>. Wordt in de onafhankelijke-variabelen- of x_D-ruimte een stap gezet in de richting van de negatieve gradiënt

$$\Delta x_D := -\alpha\nabla_{x_D}\varphi = -\alpha(-D^T B^{-T}\nabla_{x_B} f + \nabla_{x_D} f) \tag{3.2.38}$$

dan correspondeert daarmee een stap in de richting van de afhankelijk variabelen gelijk aan (vgl. (3.2.34))

$$\begin{aligned}\Delta x_B &:= -B^{-1}D\Delta x_D = -B^{-1}D(-\alpha\nabla_{x_D}\varphi)\\ &:= -\alpha(B^{-1}DD^T B^{-T}\nabla_{x_B} f - B^{-1}D\nabla_{x_D} f)\end{aligned} \tag{3.2.39}$$

Deze twee laatste uitdrukkingen leveren als principiële zoekrichting voor de gereduceerde-gradiënt methode

$$\begin{aligned}\bar{d}_B &:= -B^{-1}DD^T B^{-T}\nabla_{x_B} f + B^{-1}D\nabla_{x_D} f\\ \bar{d}_D &:= \quad D^T B^{-T}\nabla_{x_B} f - \quad \nabla_{x_D} f\end{aligned} \tag{3.2.40}$$

of in matrix-vector notatie

$$\bar{d} := -\begin{bmatrix} B^{-1}DD^T B^{-T} & -B^{-1}D \\ -D^T B^{-T} & I \end{bmatrix}\begin{bmatrix}\nabla_{x_B} f \\ \nabla_{x_D} f\end{bmatrix} =: -\bar{R}\nabla f \tag{3.2.41}$$

Met deze laatste uitdrukking kan de gereduceerde-gradiënt-methode in principe op dezelfde wijze worden uitgewerkt als de geprojecteerde-gradiënt methode besproken in het voorgaande.

c) Coördinaatbeperkingen en de gereduceerde-gradiënt methode

3.2.14. De hierboven weergegeven beschrijving van de gereduceerde-gradiënt-methode verschilt van de gebruikelijke beschrijving van de methode in de literatuur (vgl. [3.2.1], [3.2.2]). Dit verschil is het gevolg van de gebruikelijke verschillende behandeling van coördinaatbeperkingen (vgl. (3.1.14)) bij de geprojecteerde-gradiënt methode en bij de gereduceerde-gradiënt methode. Bij de geprojecteerde-gradiënt methode worden coördinaatbeperkingen opgevat als gewone beperkingen en worden alleen dan in de matrix $N^{(k)T}$ (in (3.2.15)) opgenomen indien zij actief zijn. Bij de gereduceerde-gradiënt methode spelen de coördinaat-beperkingen een analoge rol als de coördinaatbeperkingen bij de simplex-methode voor lineaire programmering (vgl. [3.2.12]). In de literatuur is het zelfs gebruikelijk te veronderstellen dat alle variabelen in ieder geval ondergrenzen en zo mogelijk bovengrenzen bezitten: Als uitgangspunt voor de afleiding van de gereduceerde-gradiënt methode wordt in plaats van (3.2.35)

$$\min \{f(x) \mid N^T x = b\}$$

meestal het in "standaard LP-formaat" gestelde probleem genomen

$$\min \{ f(x) \mid N^T x = b \ , \ x \geq 0\} \qquad (3.2.42)$$

Deze probleemformulering kan zoals in pt. 3.2.5. al besproken altijd worden afgeleid door invoering van niet-negatieve "slack"-variabelen in de ongelijkheden en vervanging van niet-teken-gebonden variabelen door combinaties van niet-negatieve variabelen (3.2.5). Door deze beide maatregelen is zowel het totale aantal variabelen als het aantal actieve beperkingen (en daarmee beide dimensies van de matrix N^T) groter dan in de eerdere probleemformulering (3.2.35).

3.2.15. De coördinaatbeperkingen worden bij toepassing van de gereduceerde-gradiënt methode gesplitst in actieve en passieve beperkingen. Alleen die variabelen die corresponderen met passieve coördinaatbeperkingen kunnen basis variabelen zijn (Bij de simplex methode zijn alleen de coördinaatbeperkingen die corresponderen met de basisvariabelen passief en alle overige actief). De voorkeur bij de selectie van basisvariabelen gaat uit naar die variabelen die zover mogelijk van hun begrenzingen liggen. De niet-basis variabelen kunnen (anders dan bij de simplex methode) hun grenswaarden al dan niet aangenomen hebben. Voor die variabelen die op hun grens liggen wordt de zoekrichting (3.2.41) van de gereduceerde-gradiënt methode aangepast en wel zo dat de betreffende component gelijk aan 0 wordt gesteld indien de betrokken coördinaat-beperking overschreden zou worden, in formule vorm

$$\begin{aligned} d_{D,i} &:= \bar{d}_{D,i} \quad \text{als } \bar{d}_{D,i} \geq 0 \quad \vee \quad x_{D,i} > 0 \\ &:= 0 \quad \text{" } \bar{d}_{D,i} < 0 \quad \wedge \quad x_{D,i} = 0 \end{aligned} \tag{3.2.43}$$

De componenten van de zoekrichting die corresponderen met de basis variabelen worden hieraan aangepast volgens (3.2.39) waarin de aangepaste gereduceerde gradiënt de gewone gereduceerde gradiënt vervangt

$$d_B := -B^{-1}Dd_D \tag{3.2.44}$$

Opgemerkt kan worden dat toepassing van de hier geschetste procedure impliceert dat de actieve-set-strategie (pt. 3.2.3) niet wordt toegepast.

3.2.16. Wordt gebruik gemaakt van de zoekrichtingen gedefinieerd door (3.2.43) en (3.2.44) dan stopt het iteratieve proces wanneer

$$d_D := 0 \tag{3.2.45}$$

hetgeen het geval is als voor alle componenten geldt

$$\bar{d}_{D,i} := (-\nabla_{x_D} f + D^T B^{-T} \nabla_{x_B} f)_i \leq 0 \tag{3.2.46}$$

Deze voorwaarde is equivalent aan de voorwaarde bij de simplex methode dat (in geval van maximalisering van de objectfunctie) de z.g. d-rij

positief is. (vgl.[3.2.12] p. 84) Definieert men een vector $\lambda \in \mathbb{R}^m$ door

$$\lambda := B^{-T}\nabla_{x_B} f \tag{3.2.47}$$

dan kan (3.2.46) herschreven worden als

$$\nabla_{x_D} f - D^T\lambda \geq 0 \tag{3.2.48}$$

terwijl uit de definitie (3.2.47) volgt dat

$$\nabla_{x_B} f - B^T\lambda = 0 \tag{3.2.49}$$

Combinatie van deze twee uitdrukkingen levert de vector ongelijkheid

$$\nabla_x f - N\lambda \geq 0 \tag{3.2.50}$$

Definieert men een vector $\mu \in \mathbb{R}^n$ zo dat

$$\nabla_x f - N\lambda - \mu = 0 \tag{3.2.51}$$

dan volgt dat de stopconditie voor de gereduceerde gradiënt algorithme (3.2.50) equivalent is aan de eis dat

$$\mu \geq 0 \tag{3.2.52}$$

Gegeven de omstandigheid dat als Lagrangefunctie van het uitgangsprobleem (3.2.42) kan worden geformuleerd

$$L(x,\lambda,\mu) = f(x) - \lambda^T(N^Tx - b) - \mu^Tx \tag{3.2.53}$$

dan blijkt de gevonden optimaliteits voorwaarde voor de gereduceerde-gradiënt methode juist een van de in pt. 3.1.20 besproken noodzakelijke voorwaarden voor een optimaal punt.

d) Vergelijking van de geprojecteerde- en de gereduceerde-gradiënt methode

3.2.17. De principiële zoekrichting $\bar{d}$ (3.2.41) die ten grondslag ligt aan de gereduceerde gradiënt-methode en de zoekrichting d (3.2.20) van de geprojec-

teerde-gradiënt methode hebben met elkaar gemeen, dat zij beide zoekrichtingen zijn die liggen in het raakvlak aan de actieve beperkingen of equivalent het orthogonale complement van de deelruimte opgespannen door de normalen van de actieve beperkingen. Het verschil tussen beide is gelegen in het gebruik van een andere basis voor dat complement: Bij de gereduceerde-gradiënt methode wordt in plaats van de kolommen van de matrix $Z^{(k)}$ (3.2.21) gebruik gemaakt van de kolommen van de matrix

$$M^{(k)} := \begin{bmatrix} -(B^{(k)})^{-1}D^{(k)} \\ \\ I \end{bmatrix} \tag{3.2.54}$$

waarvoor geldt (bij weglating van de indices (k))

$$N^T M = [B \mid D] \begin{bmatrix} -B^{-1}D \\ I \end{bmatrix} = 0_{m \times (n-m)} \tag{3.2.55}$$

In plaats van door (3.2.22) kan het raakvlak aan de actieve beperkingen worden weergegeven door

$$T(x^{(k)}) := \{x \in \mathbb{R}^n \mid x = x^{(k)} + M^{(k)}v \ , \ v \in \mathbb{R}^{n-m}\} \tag{3.2.56}$$

en de restrictie van de objectfunctie tot dit raakvlak door (vgl. 3.2.23))

$$\varphi^{(k)}(v) := f(x^{(k)} + M^{(k)}v) \tag{3.2.57}$$

Toepassing van de gradiënt-methode in de v-ruimte geeft analoog aan (3.2.24) als zoekrichting in de v-ruimte

$$d_v^{(k)} := -\nabla_v \varphi^{(k)}(0) := -M^{(k)T} \nabla_x f(x^{(k)}) \tag{3.2.58}$$

en corresponderend daarmee als zoekrichting in de originele (x-)ruimte

$$\bar{d}^{(k)} := -M^{(k)} M^{(k)T} \nabla_x f(x^{(k)}) \tag{3.2.59}$$

ofwel (met weglating van de indices (k))

$$\bar{d} := -\left[\begin{array}{c} -B^{-1}D \\ \hline I \end{array}\right] [-D^T B^{-T} \;\vdots\; I]\, \nabla f$$

$$:= -\begin{bmatrix} B^{-1}DD^T B^{-T} & -B^{-1}D \\ -D^T B^{-T} & I \end{bmatrix} \nabla f \qquad (3.2.60)$$

De werkelijke zoekrichting (vgl. (3.2.43) en (3.2.44)) van de gereduceerde-gradiënt methode in het geval coördinaatbeperkingen kan in deze terminologie worden weergegeven als

$$d := \left[\begin{array}{c} -B^{-1}D \\ \hline I \end{array}\right] \begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_{n-m} \end{bmatrix} [-D^T B^{-T} \quad I]\, \nabla f$$

$$:= \begin{bmatrix} -B^{-1}D \\ \\ I \end{bmatrix} \begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_{n-m} \end{bmatrix} \bar{d}_D \qquad (3.2.61)$$

waar

$$\sigma_i := 0 \text{ als } x_{D,i} = 0 \wedge \bar{d}_{D,i} < 0$$

$$:= 1 \text{ anders} \qquad (3.2.62)$$

3.2.18. Andersom, vergeleken met de afleiding in pt. 3.2.17, kan de zoekrichting van de geprojecteerde-gradiënt methode worden afgeleid op dezelfde wijze als de principiële zoekrichting van de gereduceerde-gradiënt methode. Uitgangspunt daarbij is (vgl. [3.2.17]) de algemene aan de locale actieve beperkingen aangepaste (met weglating van de indices (k)) coördinaten transformatie

$$y_1 := N^T x - b$$

$$y_2 := E^T x - c \qquad (3.2.63)$$

ofwel in matrix-vectornotatie

$$\begin{bmatrix} y_1 \\ \\ y_2 \end{bmatrix} := \begin{bmatrix} N^T \\ \\ E^T \end{bmatrix} \begin{bmatrix} \\ x \\ \\ \end{bmatrix} - \begin{bmatrix} b \\ \\ c \end{bmatrix} \tag{3.2.64}$$

waarin E^T en c respectievelijk een nog te kiezen ((n-m) x n)-matrix en een (n-m)-vector zijn en waarvoor de inverse transformatie luidt

$$x = \begin{bmatrix} N^T \\ \\ E^T \end{bmatrix}^{-1} \left\{ \begin{bmatrix} y_1 \\ \\ y_2 \end{bmatrix} + \begin{bmatrix} b \\ \\ c \end{bmatrix} \right\} \tag{3.2.65}$$

De objectfunctie kan in termen van de nieuwe coördinaten worden herschreven als

$$\varphi(y_1,y_2) := f\left(\begin{bmatrix} N^T \\ \\ E^T \end{bmatrix}^{-1} \left\{ \begin{bmatrix} y_1 \\ \\ y_2 \end{bmatrix} + \begin{bmatrix} b \\ \\ c \end{bmatrix} \right\} \right) \tag{3.2.66}$$

en het minimaliseringsprobleem (3.2.32) als

$$\min \{ \varphi(y_1,y_2) \mid y_1 = 0 \ , \ y_2 \in \mathbb{R}^{n-m} \} \tag{3.2.67}$$

Een direct voor de hand liggende, aan de gradiënt methode verwante zoekrichting voor dit minimaliseringsprobleem is

$$\begin{aligned} d_{y_1} &:= 0 \\ \\ d_{y_2} &:= - \nabla_{y_2} \varphi(0,y_2) \end{aligned} \tag{3.2.68}$$

waar

$$\nabla_{y_2}{}^T \varphi(0,y_2) := \nabla_x{}^T f(x) \begin{bmatrix} N^T \\ \\ E^T \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ --- \\ I \end{bmatrix}_{n \times (n-m)} \tag{3.2.69}$$

Vertaald in de originele x-coördinaten resulteert dit in de volgende zoekrichting van deze algemene gereduceerde-gradiënt methode

$$d_x := - \begin{bmatrix} N^T \\ -- \\ E^T \end{bmatrix}^{-1} \begin{bmatrix} O \\ --- \\ I \end{bmatrix} [O \,\vdots\, I] \left[\begin{array}{c:c} N & E \end{array}\right]^{-1} \nabla_x f(x) \qquad (3.2.70)$$

a) Kiest men de matrix E^T zo dat

$$\begin{bmatrix} N^T \\ E^T \end{bmatrix} := \left[\begin{array}{cc} B & D \\ \hline O & I \end{array}\right] \qquad (3.2.71)$$

hetgeen impliceert als c = O

$$y_1 := N^T x - b$$

$$y_2 := x_D \qquad (3.2.72)$$

dan volgt dat

$$\begin{bmatrix} N^T \\ E^T \end{bmatrix}^{-1} = \left[\begin{array}{c:c} B^{-1} & -B^{-1}D \\ \hline O & I \end{array}\right] \qquad (3.2.73)$$

waarmee

$$\begin{bmatrix} N^T \\ E^T \end{bmatrix}^{-1} \begin{bmatrix} O \\ --- \\ I \end{bmatrix} = \begin{bmatrix} -B^{-1}D \\ ----- \\ I \end{bmatrix} = M \qquad (3.2.74)$$

en

$$d_x := -MM^T \nabla f = \begin{bmatrix} -B^{-1}D \\ ----- \\ I \end{bmatrix} [-D^T B^{-T} \,\vdots\, I] \nabla f \qquad (3.2.75)$$

Het resultaat is dan juist de in pt. 3.2.17 afgeleide principiële zoekrichting voor de gereduceerde-gradiënt methode.

b) Kiest men op analoge wijze de matrix E^T zo dat

$$\begin{bmatrix} N^T \\ E^T \end{bmatrix} := \begin{bmatrix} N^T \\ Z^T \end{bmatrix} \qquad (3.2.76)$$

hetgeen impliceert dat als c = 0

$$y_1 := N^T x - b$$
$$y_2 := Z^T x \qquad (3.2.77)$$

dan volgt

$$\begin{bmatrix} N^T \\ E^T \end{bmatrix}^{-1} = \left[\begin{array}{c:c} N(N^TN)^{-1} & Z \end{array}\right] \qquad (3.2.78)$$

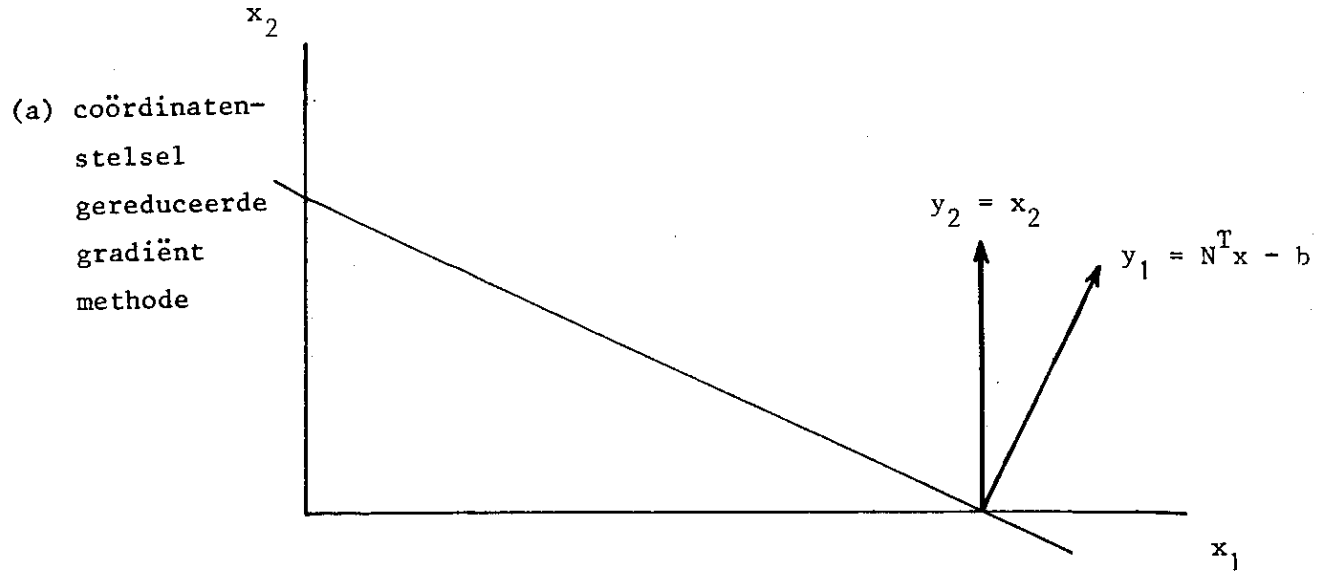
waarmee

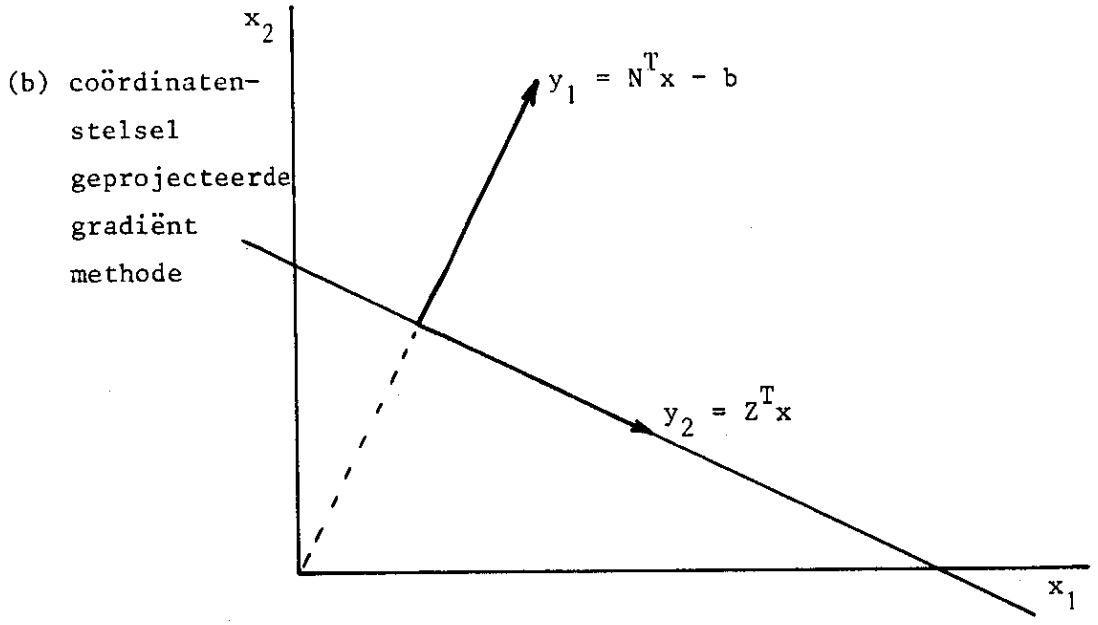
$$\begin{bmatrix} N^T \\ E^T \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ \hline I \end{bmatrix} = \begin{bmatrix} Z \\ \end{bmatrix} = Z \qquad (3.2.79)$$

en

$$d_x := -ZZ^T \nabla f \qquad (3.2.80)$$

Dit resultaat is juist de in pt. 3.2.11 afgeleide zoekrichting voor de geprojecteerde-gradiënt methode. Vergelijking van beide afleidingen en in het bijzonder beide keuzen (3.2.71) en (3.2.76) voor de matrix E^T illustreert opnieuw dat het verschil tussen beide methoden uitsluitend gelegen is in de keuze van het coördinatenstelsel waarin gewerkt wordt.
Een schets die dit verschil vooor het eenvoudige geval dat n = 2 en m = 1 illustreert is weergegeven in Figuur 3.2.18.





Figuur 3.2.18: Vergelijking van de gebruikte coördinatenstelsels bij de gereduceerde en geprojecteerde gradiënt methode.

3.2.19. De algemene formulering van de afleiding van de zoekrichting van de gereduceerde gradiënt-methode in pt. 3.2.18 geeft nog aanleiding tot een tweetal andere interessante resultaten. In de eerste plaats betreft dit de formulering van de stap nodig om terug te keren op de lineaire beperking indien men zich daar niet langer op bevindt. In termen van de vector y is deze stap indien voor het uitgangspunt geldt

$$N^T x - b = e \tag{3.2.81}$$

gelijk aan (vgl. (3.2.63))

$$y_1 := -e \tag{3.2.82}$$

In het geval van de gereduceerde-gradiënt methode is de corresponderende stap in de originele coördinaten gelijk aan (vgl. (3.2.65) en (3.2.73))

$$\Delta x := -\left[\begin{array}{c} N^T \\ \hline E^T \end{array}\right]^{-1} \left[\begin{array}{c} I \\ \hline 0 \end{array}\right] e = -\left[\begin{array}{c} B^{-1}e \\ \hline 0 \end{array}\right] \tag{3.2.83}$$

ofwel

$$\begin{aligned} \Delta x_B &:= -B^{-1}e \\ \Delta x_D &:= 0 \end{aligned} \tag{3.2.84}$$

Dit resultaat illustreert het gebruik bij toepassing van de gereduceerde-gradiënt methode om alleen door middel van aanpassingen van de basis variabelen terug te keren naar de beperkingen (vgl. [3.2.9]).
In het geval van de geprojecteerde-gradiënt methode is de met (3.2.82) corresponderende stap in de originele coördinaten gelijk aan

$$\Delta x := -\left[\begin{array}{c} N^T \\ \hline E^T \end{array}\right]^{-1} \left[\begin{array}{c} I \\ \hline 0 \end{array}\right] e = -N(N^TN)^{-1}e \tag{3.2.85}$$

Deze stap is zoals eenvoudig kan worden nagegaan juist de minimum-norm oplossing van het probleem (vgl. pt. 3.2.7)

$$N^T\Delta x = -e \tag{3.2.86}$$

3.2.20. Een tweede interessant resultaat betreft de uitdrukking voor de Lagrange-vector $\hat{\lambda} \in \mathbb{R}^m$ in het optimale punt op de lokale gelineariseerde beperkingen. Volgens Stelling 3.1.11 geldt daarvoor

$$\nabla_x f(\hat{x}) = N\hat{\lambda} = [N \mid E] \left[\begin{array}{c} \hat{\lambda} \\ \hline 0 \end{array}\right] \tag{3.2.87}$$

De coördinaten transformatie (3.2.64) impliceert dat ook geldt (vgl. (3.2.67))

$$\nabla_x f(\hat{x}) := [N \mid E]\nabla_y \varphi(0,\hat{y}_2) \tag{3.2.88}$$

en vergelijking leert dan onmiddellijk dat

$$\hat{\lambda} = \nabla_{y_1} \varphi(0,\hat{y}_2) \tag{3.2.89}$$

waar,analoog aan (3.2.69)

$$\nabla_{y_1}{}^T\varphi(y_1,y_2) := \nabla_x{}^T f(x) \begin{bmatrix} N^T \\ E^T \end{bmatrix}^{-1} \begin{bmatrix} I \\ --- \\ O \end{bmatrix}_{n\times m} \tag{3.2.90}$$

In het geval van de gereduceerde-gradiënt methode volgt dat

$$\begin{aligned} \hat{\lambda} &= [I \mid O]\,[N \mid E]^{-1}\,\nabla_x f(\hat{x}) \\ &= [I \mid O] \left[\begin{array}{c|c} B^{-T} & O \\ \hline -D^T B^{-T} & I \end{array}\right] \nabla_x f(\hat{x}) \\ &= B^{-T}\nabla_{x_B} f(\hat{x}) \end{aligned} \tag{3.2.91}$$

welk resultaat in overeenstemming is met (3.2.47) en met de uit de lineaire programmering bekende uitdrukking voor de duale variabele

$$\mu^T := c_B^T B^{-1} \tag{3.2.92}$$

In het geval van de geprojecteerde-gradiënt methode geldt analoog (vgl. (3.2.78))

$$\begin{aligned} \hat{\lambda} &= [I \mid O]\,[N \mid E]^{-1}\nabla_x f(\hat{x}) \\ &= [I \mid O] \begin{bmatrix} (N^TN)^{-1}N^T \\ Z^T \end{bmatrix} \nabla_x f(\hat{x}) \\ &= (N^TN)^{-1}N^T\nabla_x f(\hat{x}) \end{aligned} \tag{3.2.93}$$

welk resultaat op zijn beurt in overeenstemming is met de in pt. 3.1.13 afgeleide uitdrukking voor de Lagrange-multiplicatorenvector (3.1.41)

$$\hat{\lambda} = N^+\nabla_x f(\hat{x})$$

Combinatie van beide uitdrukkingen leidt tot de interessante observatie dat in het optimale punt geldt

$$B^{-T}\nabla_{x_B} f(\hat{x}) = (N^T N)^{-1} N^T \nabla_x f(\hat{x}) \qquad (3.2.94)$$

Dit resultaat werpt een interessant licht op de relatie tussen de simplex methode voor lineaire programmeringsproblemen enerzijds en de primale methoden voor niet-lineaire programmeringsproblemen anderzijds.

Bepaling van de stapbegrenzing (stap(vi))

3.2.21. Van de drie te bespreken stappen (i), (v) en (vi) en van de standaard algorithme voor primale methoden (vgl. pt. 3.2.4) is de laatste, stap (vi), die betrekking heeft op de bepaling van de stapbegrenzing nadat de zoekrichting d bepaald is de eenvoudigste, speciaal in het geval van lineaire beperkingen. Uitgaande van het probleem (3.2.15)

$$\min \{f(x) \mid N^{(k)T}(x - x^{(k)}) = 0 \ , \ \tilde{A}^{(k)T}(x - x^{(k)}) \geq -\tilde{c}^{(k)}\}$$

bestaat het bepalen van de stapgroottebegrenzing uit het bepalen van de maximale waarde $\bar{\alpha}$ waarvoor nog juist voldaan wordt aan de vector ongelijkheid

$$\tilde{A}^{(k)T}(x^{(k)} + \bar{\alpha} d^{(k)} - x^{(k)}) = \bar{\alpha}\tilde{A}^{(k)T} d^{(k)} \geq -\tilde{c}^{(k)}$$

of componentsgewijs

$$\bar{\alpha}\tilde{a}_i^{(k)T} d^{(k)} \geq -\tilde{c}_i^{(k)} \qquad i \in I_p(x^{(k)}) \qquad (3.2.95)$$

In het geval dat voor zekere $i \in I_p(x^{(k)})$ geldt dat $\tilde{a}_i^{(k)T} d^{(k)} > 0$ dan wordt voor alle $\alpha > 0$ aan de ongelijkheid voldaan. Van belang voor de stapgroottebegrenzing zijn daarom alleen die beperkingen met een index in de verzameling

$$I_p^-(x^{(k)}) := \{i \in I_p(x^{(k)}) \mid \tilde{a}_i^{(k)T} d^{(k)} < 0\} \qquad (3.2.96)$$

In het geval dat $I_p^-(x^{(k)}) = \emptyset$ dan bestaat geen stapgroottebegrenzing (dus: $\bar{\alpha} := \infty$), in het andere geval geldt dat

$$\bar{\alpha} := \min \left\{ -\frac{\tilde{c}_i^{(k)}}{\tilde{a}^{(k)T} d^{(k)}} \;\middle|\; i \in I_p^-(x^{(k)}) \right\} \qquad (3.2.97)$$

Wordt een stap ter grootte $\bar{\alpha}$ gerealiseerd dan wordt de i-de passieve beperking waarvoor de minimale waarde van het quotiënt in (3.2.97) wordt bereikt actief. Bij toepassingen in de practijk is het nuttig naast de stapgroottebegrenzing $\bar{\alpha}$ ook te onthouden welke beperking daarmee actief wordt.

3.2.22. De hier geschetste bepaling van de stapgroottebegrenzing is in essentie gelijk aan de bepaling van de spilrij (engl.: pivot row) in de simplex methode voor lineaire programmeringsproblemen (vgl. [3.2.12]). Om dit te laten zien is het nodig in te gaan op de relatie tussen de simplex methode (welke bekend wordt verondersteld) en de in het voorgaande besproken gereduceerde-gradiënt methode. Uitgangspunt voor de simplex methode is de gemodificeerde versie van het LP-probleem (3.1.16)

$$\min \{c^T x \mid [I \mid B^{-1}D]\begin{bmatrix} x_B \\ x_D \end{bmatrix} = B^{-1}b \ , \ x \geq 0\} \tag{3.2.98}$$

en de daarmee corresponderende zoekrichting wordt gegeven door (vgl. (3.2.44))

$$\begin{aligned} d_B &:= -B^{-1}De_s \\ d_D &:= e_s \end{aligned} \tag{3.2.99}$$

waar e_s de eenheids-(n-m)-vector is, d.i.

$$e_s := (0,0, \ldots 0, 1, 0, \ldots 0) \tag{3.2.100}$$

en s de index is van de grootste niet-basis-variabele component van de principiële zoekrichting (= negatieve gereduceerde gradiënt (vgl. (3.2.40))) van de simplex methode

$$\begin{aligned} \bar{d}_D &:= -c_D + D^T B^{-T} c_B \\ \bar{d}_B &:= -B^{-1}D\bar{d}_D \end{aligned} \tag{3.2.101}$$

In het geval dat geen degeneratie optreedt zijn er bij toepassing van de simplex methode in iedere stap steeds juist m passieve beperkingen, nl. de coördinaat-beperkingen corresponderend met de (van nul verschillende)

basisvariabelen

$$x_B \geq 0 \tag{3.2.102}$$

Voor deze passieve beperkingen geldt in terminologie van het voorgaande (pt. 3.2.21) dat

$$\tilde{a}_i^{(k)} := e_i \tag{3.2.103}$$

waar e_i de i-de eenheids-(n)-vector is en

$$-\tilde{c}_i^{(k)} := -x_{B,i} := -(B^{-1}b)_i \tag{3.2.104}$$

d.i. juist het i-de element van het rechterlid in de probleemformulering (3.2.15). Het inwendig product tussen de normaal van de i-de passieve beperking en de zoekrichting wordt gegeven door

$$\tilde{a}_i^{(k)} T_d^{(k)} := -(B^{-1}D)_{i,s} \tag{3.2.105}$$

d.i. het i-de element van de s-de kolom van de matrix $B^{-1}D$, de z.g. spilkolom (engl.: pivot column) in het simplex-tableau. Voor de maximale stap die kan worden gezet voordat een van de passieve beperkingen actief wordt volgt daarmee overeenkomstig (3.2.97)

$$\bar{\alpha} := \min\left\{\frac{(B^{-1}b)_i}{(B^{-1}D)_{i,s}} \;\middle|\; (B^{-1}D)_{i,s} > 0, i = 1, \ldots, m\right\} \tag{3.2.106}$$

Dit voorschrift is juist de bekende uitdrukking (vgl. [3.2.12] p. 91) voor de bepaling van de spil rij(engl.: pivot row) in de simplex methode. In het geval r de rij-index is waarvoor de minimumwaarde van $\bar{\alpha}$ wordt gerealiseerd dan volgt dat een stap in de zoekrichting d (3.2.99) met deze stapgrootte $\bar{\alpha}$ resulteert in het actief (= gelijk aan nul) worden van de r-de beperking

$$x_{B,r} = 0 \tag{3.2.107}$$

en het passief worden van de s-de actieve beperking

$$x_{D,s} := \bar{\alpha} \qquad (3.2.108)$$

In de simplex-methode wordt in dit geval $x_{B,r}$ tot niet-basis variabele en $x_{D,s}$ tot basis variabele verklaard waarna volgens de bekende methode van "vegen" een nieuw simplex tableau wordt gegenereerd. De bij de simplex-methode gebruikelijke basiswisseling is dus equivalent met het realiseren van de grootst mogelijke stap in de zoekrichting (3.2.99) van de simplex-methode. Er geldt dat bij de simplex algorithme in iedere iteratieslag de grootst mogelijke stap $\bar{\alpha}$ (3.2.106) inderdaad wordt gezet en dat daarom ook in iedere slag een basiswisseling als hierboven omschreven plaats vindt. Bij de primale methoden wordt deze grootst mogelijke stap zeker niet altijd gezet en bij de gereduceerde-gradiënt methode zijn mede als gevolg van de omstandigheid dat niet alle niet-basisvariabelen gelijk aan nul zijn (en mogelijk niet alle begrensd) andere criteria nodig in de practijk om te bepalen wanneer en welke basiswisseling moet plaats vinden (vgl. [3.2.13], [3.2.19]).

Matrix-aanpassingen bij toevoeging of loslaten van actieve beperkingen (stap (iii) (en (i)))

3.2.23. In het geval bij toepassing van primale methoden nieuwe lineaire beperkingen aan de verzameling van actieve beperkingen moeten worden toegevoegd dan wel oude daaruit verwijderd dan is het niet nodig alle berekeningen voor de nieuwe matrix van normalen van voren af aan opnieuw uit te voeren. Door diverse auteurs werden algorithmen ontwikkeld om in deze gevallen zo effectief mogelijk gebruik te maken van de informatie die aanwezig was in de uitdrukkingen berekend voor de oude matrix van normalen. Het bekendst in dit verband zijn de aanpassingsformules van Rosen [3.2.7] voor het recursief bepalen van de matrix-inverse $(N^{(k+1)T}N^{(k+1)})^{-1}$ met behulp van de gegeven veronderstelde matrix-inverse $(N^{(k)T}N^{(k)})^{-1}$. Daarnaast zijn van veel belang in de practijk de aanpassingsformules voor de bepaling van nieuwe QR-decompositie van de matrix $N^{(k+1)}$ gegenereerd door toevoeging of weglating van een kolom van de matrix $N^{(k)}$. Deze laatste werden ontwikkeld door Bartels, Golub en Saunders [3.2.22] en toegepast door o.a. Gill & Murray [3.2.23] (en [3.2.24]). Hieronder zullen de ideeën achter beide soorten aanpassingen kort worden toegelicht.

3.2.24. De aanpassingsformules van Rosen zijn, zowel voor het geval dat een kolom wordt toegevoegd als voor het geval dat een kolom wordt verwijderd uit de matrix $N^{(k)}$, gebaseerd op de volgende uit de matrixtheorie bekende uitspraak voor de inverse van een "omrande" matrix (vgl. [3.2.5])

Stelling 3.2.24 : Als A een symmetrische positief definiete matrix is en $[a^T \mid \alpha]^T \neq 0$ dan geldt

$$\left[\begin{array}{c|c} A & a \\ \hline a^T & \alpha \end{array}\right]^{-1} = \left[\begin{array}{c|c} B & b \\ \hline b^T & \beta \end{array}\right] \tag{3.2.109}$$

met

$$\beta = (\alpha - a^T A^{-1} a)^{-1} \tag{3.2.110}$$

$$b = -\beta A^{-1} a = -(\alpha - a^T A^{-1} a)^{-1} A^{-1} a \tag{3.2.111}$$

$$B = A^{-1} + \beta A^{-1} a a^T A^{-1} = A^{-1} + \frac{1}{\beta} b b^T \tag{3.2.112}$$

Bewijs :
Eenvoudig kan worden geverifieerd dat door de uitdrukkingen wordt voldaan aan de noodzakelijke relaties

$$\begin{array}{ll} AB + ab^T = I & Ab + \beta a = 0 \\ a^T B + \alpha b^T = 0 & a^T b + \alpha\beta = 1 \end{array} \tag{3.2.113}$$

De voorwaarde dat A niet-singulier is en dat $[a^T \mid \alpha]^T \neq 0$ maken het mogelijk de gewenste uitdrukkingen (3.2.110) t/m (3.2.112) uit deze laatste relaties af te leiden. □

Gevolg 3.2.24 : Op grond van de voorgaande stelling geldt als $(N \mid n)$ een matrix is met volle rang dat

$$\left(\left[\begin{array}{c} N^T \\ \hline n^T \end{array}\right]\left[\begin{array}{c|c} N & n \end{array}\right]\right)^{-1} = \left(\begin{array}{c|c} N^TN & N^Tn \\ \hline n^TN & n^Tn \end{array}\right)^{-1} = \left(\begin{array}{c|c} B & b \\ \hline b^T & \beta \end{array}\right) \qquad (3.2.114)$$

met

$$\beta = \| \bar{P}n \|^{-2} \qquad (3.2.115)$$

$$b = -\beta(N^TN)^{-1}N^Tn \qquad (3.2.116)$$

$$B = (N^TN)^{-1} + \frac{1}{\beta} bb^T \qquad (3.2.117)$$

waar

$$\bar{P}n = (I - N(N^TN)^{-1}N^T)n \qquad (3.2.118)$$

Bewijs :
Het resultaat volgt onmiddellijk uit de voorgaande stelling samen met de overweging dat

$$(n^Tn - n^TN(N^TN)^{-1}N^Tn) = (n^Tn - (n^TN(N^TN)^{-1}N^T)(N(N^TN^{-1})N^Tn))$$

$$= n^T(I - N(N^TN)^{-1}N^T)(I - N(N^TN)^{-1}N^T)n$$

De overige formules resulteren direct door substitutie. □

3.2.25. Wordt de verzameling van normalen van actieve beperkingen uitgebreid met een nieuwe normaal n_{q+1} , d.i, als

$$N^{(k+1)} = [N^{(k)} \mid n_{q+1}]$$

dan kan de inverse matrix $(N^{(k+1)T}N^{(k+1)})^{-1}$ direct worden gevonden met behulp van de uitdrukkingen (3.2.115) (3.2.116) en (3.2.117):

$$\beta^{(k+1)} := \| \bar{P}^{(k)}n_{q+1} \|^{-2}$$

$$b^{(k+1)} := -\beta^{(k+1)}(N^{(k)T}N^{(k)})^{-1}N^{(k)T}n_{q+1}$$

$$B^{(k+1)} := (N^{(k)T}N^{(k)})^{-1} + \frac{1}{\beta^{(k+1)}} b^{(k+1)}b^{(k+1)T}$$

Moet de verzameling van normalen worden ingekrompen dan wordt een rij- en kolomverwisseling toegepast zodat de te verwijderen normaal juist de laatste normaal n_q is in de matrix van normalen $N^{(k)}$, d.i.

$$N^{(k+1)} := [n_1,\ldots,n_{q-1}]$$

De inverse matrix $(N^{(k+1)T}N^{(k+1)})^{-1}$ volgt daarna onmiddellijk met behulp van de van uitdrukking (3.2.117) afgeleide relatie

$$(N^{(k+1)T}N^{(k+1)})^{-1} = B^{(k)} - \frac{1}{\beta^{(k)}} b^{(k)}b^{(k)T} \qquad (3.2.119)$$

waar $B^{(k)}$, $b^{(k)}$ en $\beta^{(k)}$ onder matrices zijn van de inverse matrix $(N^{(k)T}N^{(k)})^{-1}$ volgens het schema

$$(N^{(k)T}N^{(k)})^{-1} = \left(\begin{array}{c|c} B^{(k)} & b^{(k)} \\ \hline b^{(k)T} & \beta^{(k)} \end{array}\right) \qquad (3.2.120)$$

3.2.26. De hierboven weergegeven aanpassingsformules van Rosen zijn van belang indien gebruik gemaakt wordt van de projectieformule (3.2.20) voor het genereren van zoekrichtingen en van de corresponderende formule (3.2.93) voor het bepalen van de Lagrange multiplicatoren. In veruit de meeste toepassingen tegenwoordig wordt echter zowel voor het genereren van zoekrichtingen (vgl. pt. 3.2.12)

$$d^{(k)} := -(I - Q_1^{(k)}Q_1^{(k)T})\nabla f(x^{(k)}) = -Q_2^{(k)}Q_2^{(k)T}\nabla f(x^{(k)}) \qquad (3.2.121)$$

als voor de bepaling van de Lagrange-multiplicatoren vector

$$\lambda^{(k)} := R^{(k)-1}Q_1^{(k)T}\nabla f(x^{(k)}) \qquad (3.2.122)$$

gebruik gemaakt van de QR-decompositie van de matrix $N^{(k)}$ (vgl. (3.2.26))

$$N^{(k)} = \left[\begin{array}{c|c} Q_1^{(k)} & Q_2^{(k)} \end{array}\right]\left[\begin{array}{c} R^{(k)} \\ \hline 0 \end{array}\right]$$

Wordt in dit geval de verzameling van actieve (lineaire) beperkingen uitgebreid of ingekrompen dan dient overeenkomstig de QR-decompositie

te worden aangepast. Twee mogelijke procedures hiervoor worden hieronder toegelicht zonder dat wordt ingegaan op de numerieke details. Opgemerkt zij slechts dat het bij het gebruik van deze aanpassingsprocedures belangrijk is dat de gunstige numerieke eigenschappen van de QR-decompositie zo goed mogelijk gehandhaafd blijven.

3.2.27. In het geval dat de matrix van normalen $N^{(k)}$ van de actieve beperkingen met de normaal n_{q+1} moet worden uitgebreid dan geldt voor de nieuwe matrix $N^{(k+1)}$ dat

$$N^{(k+1)} = \left[\begin{array}{c:c} N^{(k)} & n_{q+1} \end{array}\right] = \left[\begin{array}{c:c} Q_1^{(k)} & Q_2^{(k)} \end{array}\right] \left[\begin{array}{c:c} R^{(k)} & r_1 \\ \hdashline 0 & r_2 \end{array}\right] \qquad (3.2.123)$$

waar

$$\left[\begin{array}{c} r_1 \\ \hdashline r_2 \end{array}\right] = \left[\begin{array}{c} Q_1^{(k)T} n_{q+1} \\ \\ Q_2^{(k)T} n_{q+1} \end{array}\right] \qquad (3.2.124)$$

Voor het bepalen van de QR-decompositie van $N^{(k+1)}$ is het voldoende om een orthogonale $(n-q) \times (n-q)$ matrix $\tilde{Q}$ te bepalen zodat

$$\tilde{Q} r_2 = \begin{bmatrix} \|r_2\| \\ 0 \\ \vdots \\ 0 \end{bmatrix} \qquad (3.2.125)$$

en waarmee

$$N^{(k+1)} = \left[\begin{array}{c:c} Q_1^{(k)} & Q_2^{(k)} \end{array}\right] \left[\begin{array}{c:c} I & 0 \\ \hdashline 0 & \tilde{Q}^T \end{array}\right] \left[\begin{array}{c:c} I & 0 \\ \hdashline 0 & \tilde{Q} \end{array}\right] \left[\begin{array}{c:c} R^{(k)} & r_1 \\ \hdashline 0 & r_2 \end{array}\right]$$

$$= \left[\begin{array}{c:c} Q_1^{(k)} & Q_2^{(k)} \tilde{Q}^T \end{array}\right] \left[\begin{array}{c:c} R^{(k)} & r_1 \\ \hdashline 0 & \|r_2\| \\ \hdashline 0 & 0 \end{array}\right] \qquad (3.2.126)$$

$$= \left[Q_1^{(k+1)} \;\vdots\; Q_2^{(k+1)} \right] \left[\begin{array}{c} R^{(k+1)} \\ \hline 0 \end{array} \right]$$

waar

$$Q_1^{(k+1)} := \left[Q_1^{(k)} \;\vdots\; q_{q+1}^{(k+1)} \right] \qquad Q_2^{(k+1)} := \left[q_{q+2}^{(k+1)} \cdots q_n^{(k+1)} \right]$$

$$R^{(k+1)} := \left[\begin{array}{c|c} R^{(k)} & r_1 \\ \hline 0 & \|r_2\| \end{array} \right] \tag{3.2.127}$$

en

$$\left[q_{q+1}^{(k+1)} q_{q+2}^{(k+1)} \cdots q_n^{(k+1)} \right] := Q_2^{(k)} \tilde{Q}^T$$

Een orthogonale $(n-q) \times (n-q)$ matrix $\tilde{Q}$ die de gewenste eigenschap (3.2.125) bezit is de Householder-matrix (vgl. [3.2.6])

$$\tilde{Q} = I - 2 \frac{uu^T}{u^T u} \tag{3.2.128}$$

met

$$u = r_2 - \|r_2\| e_1 \tag{3.2.129}$$

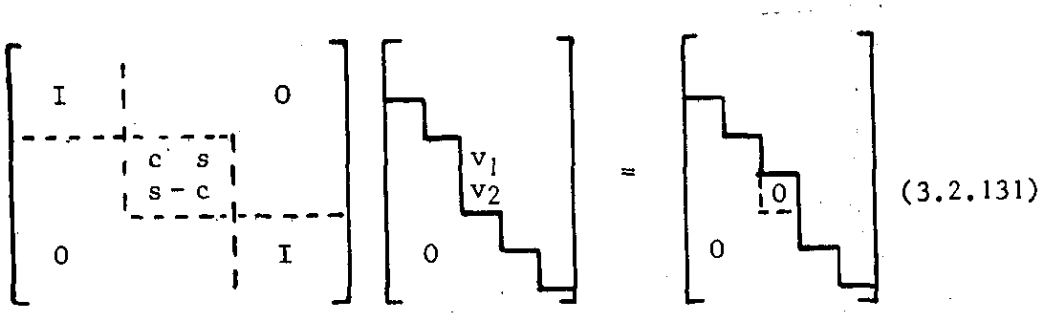
Het gebruik van deze matrix sluit uitstekend aan bij het uit het oogpunt van numerieke stabiliteit aanbevolen gebruik van Householder-matrices voor het genereren van de QR-decompositie van de voorgaande matrices $N^{(k)}$.

3.2.28. In het geval dat een normaal n_s uit de matrix van normalen moet worden verwijderd dan geldt

$$\begin{aligned} N^{(k+1)} &= [n_1 \ldots n_{s-1} n_{s+1} \ldots n_q] \\ &= \left[Q_1^{(k)} \;\vdots\; Q_2^{(k)} \right] \left[\begin{array}{c} D^{(k+1)} \\ \hline 0 \end{array} \right] \end{aligned} \tag{3.2.130}$$

waarin $D^{(k+1)}$ de $q \times (q-1)$-matrix voorstelt die resulteert wanneer uit de bovendriehoeksmatrix $R^{(k)}$ (vgl. (3.2.26)) de s-de kolom ver-

wijderd wordt. De matrix $D^{(k+1)}$ kan weer in bovendriehoeksvorm worden gebracht door successievelijke voorvermenigvuldiging ervan met (orthogonale) Givens' matrices (zie [3.2.25]) die telkens het eerste element onder de diagonaal tot nul reduceren volgens het schema



(3.2.131)

Dit laatste is mogelijk indien voor c en s respectievelijk gekozen wordt

$$c = \frac{v_1}{(v_1^2 + v_2^2)^{1/2}} \qquad s = \frac{v_2}{(v_1^2 + v_2^2)^{1/2}} \tag{3.2.132}$$

Uitwerking geeft dan

$$\begin{aligned} N^{(k+1)} &= \left[\begin{array}{c|c} Q_1^{(k)} & Q_2^{(k)} \end{array}\right] G_{s+1}^T G_{s+2}^T \ldots G_q^T G_q \ldots G_{s+2} G_{s+1} \left[\begin{array}{c} D^{(k+1)} \\ \hline 0 \end{array}\right] \\ &= \left[\begin{array}{c|c} Q_1^{(k)} & Q_2^{(k)} \end{array}\right] \tilde{G}^T \left[\begin{array}{c} R^{(k+1)} \\ \hline 0 \end{array}\right] \\ &= \left[\begin{array}{c|c} Q_1^{(k+1)} & Q_2^{(k+1)} \end{array}\right] \left[\begin{array}{c} R^{(k+1)} \\ \hline 0 \end{array}\right] \end{aligned} \tag{3.2.133}$$

waar

$$\begin{aligned} Q_1^{(k+1)} &= \left[q_1, \ldots q_{s-1}, \tilde{q}_s \ldots \tilde{q}_{q-1}\right] \\ Q_2^{(k+1)} &= \left[\begin{array}{c|c} \tilde{q}_q & Q_2^{(k)} \end{array}\right] \end{aligned} \tag{3.2.134}$$

en

$$\left[\tilde{q}_s, \ldots \tilde{q}_q\right] = \left[q_s^{(k)} \ \ldots q_q^{(k)}\right] \tilde{G}^T$$

In de practijk zijn de hier geschetste procedures ingewikkelder omdat de diverse matrices niet altijd expliciet aanwezig zijn en omdat in een aantal gevallen ter vergroting van de numerieke stabiliteit gebruik wordt gemaakt van rij- en kolomverwisselingen. Details hierover kunnen worden gevonden in de publicaties van de eerder genoemde auteurs ([3.2.3] en [3.2.25] [3.2.23] en [3.2.24]) alsook in de recente literatuur op het gebied van de numerieke lineaire algebra.

3.2.29. In niet alle gevallen is het noodzakelijk om de gehele QR-decompositie aan te passen. Soms is het bijvoorbeeld voldoende om alleen de nieuwe matrix

$$Q_2^{(k+1)} =: Z^{(k+1)} \tag{3.2.135}$$

te bepalen. In het geval van toevoeging van een nieuwe beperking kan dit relatief eenvoudig geschieden door navermenigvuldiging van de matrix $Q_2^{(k)} =: Z^{(k)}$ met de in pt. 3.2.27 besproken matrix $\tilde{Q}^T$ (3.2.128) en weglating van de eerste vector $q_{q+1}^{(k+1)}$ van het resultaat, d.i.

$$Z^{(k+1)} := Z^{(k)}\tilde{Q}^{(k)T}\begin{bmatrix} 0 \\ \hline I_{n-q-1} \end{bmatrix} \tag{3.2.136}$$

In het geval van het laten vallen van een eerdere actieve beperking kan dit eveneens relatief eenvoudig. Er moet een nieuwe kolom $\tilde{z}$ voor de matrix

$$Z^{(k+1)} := \left[\tilde{z} \;\vdots\; Z^{(k)}\right] \tag{3.2.137}$$

worden bepaald die voldoet aan de voorwaarden

$$\begin{aligned} n_j^{(k)T}\tilde{z} &= 0 \qquad j = 1,\ldots,s-1,s+1,\ldots q \\ z_j^{(k)T}\tilde{z} &= 0 \qquad j = 1,\ldots,n-q \end{aligned} \tag{3.2.138}$$

en

$$\|\tilde{z}\| = (\tilde{z}^T\tilde{z})^{1/2} = 1$$

De vector die voldoet aan deze voorwaarde is de vector

$$\tilde{z} = n_s^+ / \|n_s^+\| \tag{3.2.139}$$

waar $(n_s^+)^T$ de s-de rij is van de gegeneraliseerde of pseudo-inverse $N^{(k)+}$ van de matrix $N^{(k)}$ waarvoor geldt

$$\begin{aligned} N^{(k)+}N^{(k)} &= I_q \\ N^{(k)+}Z^{(k)} &= O_{q\times n-q} \end{aligned} \tag{3.2.140}$$

In het tot dusver steeds veronderstelde geval dat de matrix $N^{(k)}$ maximum rang heeft geldt voor de s-de rij van $N^{(k)+}$

$$(n_s^+)^T = e_s^T N^{(k)+} = e_s^T (N^{(k)T} N^{(k)})^{-1} N^{(k)T} \tag{3.2.141}$$

of equivalent

$$n_s^+ = Q_1^{(k)} R^{(k)-T} e_s = N^{(k)} R^{(k)-1} R^{(k)-T} e_s \tag{3.2.142}$$

waar e_s de s-de kolom is van de eenheidsmatrix I_q.

De vector n_s^+ kan worden bepaald als oplossing van de vergelijking

$$R^{(k)T} R^{(k)} v = e_s \tag{3.2.143}$$

en substitutie van het resultaat in

$$n_s^+ = N^{(k)} v \tag{3.2.144}$$

Gezien de speciale vorm van e_s en de driehoeksvorm van $R^{(k)}$ kost de uitwerking van deze oplossing relatief weinig moeite. Opgemerkt kan worden dat naast de hier besproken methode nog een aantal ander alternatieve methoden bekend zijn voor het bepalen van de nieuwe kolom $\tilde{z}$ (zie [3.2.23]).

Referenties

3.2.30. Voor meer details over de in deze paragraaf besproken methoden moet worden verwezen naar de volgende publikaties.

[3.2.1] : Zie [1.1.1] Luenberger (1973)

[3.2.2] : Zie [1.1.2] Jacoby, Kowalik and Pizzo (1972)

[3.2.3] : Zie [1.1.4] Gill & Murray (1974)

[3.2.4] : Zie [2.1.3] Zangwill (1969)

[3.2.5] : Zie [2.10.6] Eilers (1975)

[3.2.6] : Zie [2.10.9] Lawson and Hanson (1974)

[3.2.7] : Zie [2.10.16] Rosen (1960)

[3.2.8] : Abadie, J.: Application of the GRG algorithm to optimal control problems, Ch. 8 in : Abadie, J. (Ed): Integer and nonlinear programming North Holland, Publ. Cy, Amsterdam (1970)

[3.2.9] : Abadie, J. and Carpentier, J.: Generalization of the Wolfe reduced gradient method to the case of nonlinear constraints, in Fletcher, J. (Ed): "Optimization", Academic Press, New York (1969)

[3.2.10] : Abadie, J. and Guigou, J.: Numerical experiments with the GRG method, App. III in Abadie J.(Ed). Integer and nonlinear programming, North Holland Publ. Cy, Amsterdam (1970)

[3.2.11] : Arts, J. : Een Algol-procedure voor de geprojecteerde-gradiënt methode voor minimaliseringsproblemen met lineaire beperkingen, COSOR Notitie R 74-06 (mei 1974)

[3.2.12] : Benders, J.F.: Syllabus bij het College "Optimaliserings-methoden I", (najaar 1973)

[3.2.13] : Dirkx, C.J.B.: NONLINMIN, een Algol-procedure voor het minimaliseren van niet-lineaire functies onder niet-lineaire nevenvoorwaarden, Afstudeerverslag, Technische Hogeschool Eindhoven, Onderafdeling der Wiskunde, (december 1975)

[3.2.14] : Faure, P. et Huard, P.: Résulution de programmes mathématique à fonction non linéaire par la méthode du gradient réduit, Rev. Franc. de Rech. Oper. 36 (1965) p.p. 167-206

[3.2.15] : Fletcher, R.: Minimizing general functions subject to linear constraints, in Lootsma, F.A. (Ed) Numerical methods for nonlinear optimization, Academic Press, London (1972)

[3.2.16] : Powell, M.J.D.: Introduction to constrained optimization,Ch. I in [3.2.3] Gill & Murray (1974)

[3.2.17] : Sargent, R.W.H.: Reduced-gradient and projection methods for nonlinear programming, Ch. V in [3.2.3] Gill & Murray (1974)

[3.2.18] : Swann, W.H.: Constrained optimization by direct search, Ch. VII in [3.2.3] Gill & Murray (1974)

[3.2.19] : van der Velden, J.G.: De gereduceerde gradiënt methode voor het oplossen van mathematische programmeringsproblemen, Afstudeerverslag Technische Hogeschool Eindhoven, Onderafdeling der Wiskunde, (september 1973)

[3.2.20]: Wolfe, Ph.: Methods of nonlinear programming, in Graves, R.L. and Wolfe, Ph. (Eds): Recent advances in mathematical programming, McGraw-Hill, New York (1963)

[3.2.21]: Wolfe, Ph.: Methods for linear constraints, in : J. Abadie (Ed) Nonlinear programming, North Holland Publ. Cy (1967)

[3.2.22]: Bartels, R.H., Golub, G.H. and Saunders, M.A. "Numerical techniques in mathematical programming" in Rosen, J.B. etal, ed: "Nonlinear programming" pp 123-176, Academic Press, New York (1970)

[3.2.23]: Gill, P.E. and Murray, W.: "Two methods for the solution of linearly constrained and unconstrained optimization problems" Nat. Phys. Lab. Teddington; Rept. NAC 25 (November 1972)

[3.2.24]: Gill, P.E., and Murray, W.: "Quasi-Newton methods for linearly constrained optimization", Nat. Phys. Lab., Teddington, Rept. NAC 32 (May 1973)

[3.2.25]: Gill, P.E., Golub, G.H., Murray, W. and Saunders, M.A. "Methods for modifying matrix factorizations" Stanford University, Computer Science Department Report STAN-CS-72-322 (November 1972)

[3.2.26]: Daniel, J.W., Gragg, W.B., Kaufman, L., Stewart, G.W.: "Reorthogonalisation and stable algorithms for updating the Gram-Schmidt QR-factorization" Math. Comp., 30 (1976), pp. 772-795.

§ 3.3. Primale methoden II: Hoger-orde methoden en niet-lineare beperkingen

3.3.1. De in de voorgaande paragraaf besproken methoden maakten uitsluitend gebruik van eerste-orde informatie over de objectfunctie en veronderstelden uitsluitend lineaire beperkingen. Zij zijn als zodanig vergelijkbaar met de gradiënt-methoden voor onbeperkte minimalisering (vgl.§ 2.4) en hebben met die methoden ook de langzame convergentie in de omgeving van het locale optimum gemeen. Gezien de gunstige ervaringen met hoger-orde methoden bij de onbeperkte minimalisering, en wel in het bijzonder met de Newton-methode (vgl.§ 2.5) en de quasi-Newton methoden (vgl.§ 2.7. - 2.9), is het niet verwonderlijk dat men ook voor de minimaliseringsproblemen met (lineaire) nevenvoorwaarden heeft getracht analoge methoden te ontwikkelen. Enkele van de bekendste daarvan zullen in deze paragraaf nader worden beschouwd. Daarna zal ook nog kort aandacht worden besteed aan de generalisatie van deze methoden en die van de voorgaande paragraaf voor de toepassing op minimaliseringsproblemen waarbij ook niet-lineaire beperkingen voorkomen. De gebruikelijke procedure in dit laatste geval is dat in iedere iteratiestap een zoekrichting wordt gegenereerd en een stap wordt gezet gebaseerd op de locale linearisatie van de beperkingen. Resulteert deze stap in een niet-toegelaten punt dan wordt eerst een correctie- of "restauratie"-stap gezet die leidt tot een toegelaten punt voordat met een nieuwe iteratieslag begonnen wordt. Op een aantal practische aspecten van deze restauratieprocedure zal aan het einde van deze paragraaf nader worden ingegaan.

Oplossing van het probleem QLE (I)

3.3.2. Uitgangspunt voor de meeste hoger-orde primale methoden is het probleem dat ontstaat door tweede-orde benadering van de objectfunctie en eerste-orde benadering van de actieve beperkingen rond het laatste gevonden iteratiepunt, d.i. het met (3.2.15) vergelijkbare probleem

$$\min \{ f(x^{(k)}) + \nabla^T f(x^{(k)})(x - x^{(k)}) + \frac{1}{2} (x - x^{(k)})^T G(x^{(k)})(x - x^{(k)}) \mid N^{(k)T}(x - x^{(k)}) = 0 \ , \ \tilde{A}^{(k)T}(x - x^{(k)}) \geq - \tilde{c}^{(k)} \ , \ x \in \mathbb{R}^n \} \tag{3.3.1}$$

Voor het bepalen van de zoekrichting beperkt men zich tot de actieve beperkingen en gebruikt men als uitgangspunt de oplossing van het corresponderende probleem van het type QLE (vgl. (3.1.18))

$$\min\{\tfrac{1}{2}\, y^T Q y + q^T y \mid N^T y - b = 0\} \tag{3.3.2}$$

waar

$$\begin{aligned} y &:= x - x^{(k)} \qquad Q := G(x^{(k)}) \qquad q := \nabla f(x^{(k)}) \\ N &:= N^{(k)} \qquad\qquad b := 0 \end{aligned} \tag{3.3.3}$$

In de veronderstelling dat Q een positief definiete matrix is, bestaan er een tweetal verschillende formulering en van de oplossing van dit probleem, die beide hebben geleid tot een klasse van hoger-orde primale methoden. In verband daarmee worden hieronder eerst beide oplossingen afgeleid en wordt daarna aandacht besteed aan de van deze oplossingen afgeleide klassen van primale methoden.

3.3.3. De eerste, meest voor de hand liggende formulering van de oplossing van het probleem QLE (3.3.2) kan worden gevonden door het bepalen van een punt dat aan de noodzakelijke voorwaarden voldoet. In termen van de afgeleiden van de Lagrange-functie

$$L(y,\lambda) := \tfrac{1}{2}\, y^T Q y + q^T y - \lambda^T (N^T y - b) \tag{3.3.4}$$

luiden deze noodzakelijk voorwaarden voor optimaliteit

$$\begin{aligned} \nabla_y L &:= Qy + q - N\lambda = 0 \\ -\nabla_\lambda L &:= N^T y - b = 0 \end{aligned} \tag{3.3.5}$$

In het geval Q niet-singulier is volgt hieruit dat

$$y = -Q^{-1} q + Q^{-1} N\lambda \tag{3.3.6}$$

hetgeen bij substitutie in de tweede vergelijking leidt tot

$$N^TQ^{-1}N\lambda - N^TQ^{-1}q - b = 0$$

In de veronderstelling dat het optimale punt een regulier punt is (zodat de matrix N volle rang heeft) volgt dan als oplossing van het stelsel (3.3.5)

$$\lambda = (N^TQ^{-1}N)^{-1}N^TQ^{-1}q + (N^TQ^{-1}N)^{-1}b \tag{3.3.7}$$

en

$$y = -(I - Q^{-1}N(N^TQ^{-1}N)^{-1}N^T)Q^{-1}q + Q^{-1}N(N^TQ^{-1}N)^{-1}b \tag{3.3.8}$$

Deze laatste oplossing is voor het geval dat m = 1 en n = 2 geïllustreerd in Figuur 3.3.3. Op te merken valt dat de matrix

$$\bar{P}_{(Q^{-1})} := (I - Q^{-1}N(N^TQ^{-1}N)^{-1}N^T) \tag{3.3.9}$$

een scheve-projectie matrix is die vectoren loodrecht op de kolommen van de matrix N onveranderd laat en vectoren in de richting van de kolommen van de matrix $Q^{-1}N$ tot nul reduceert. De oplossing y bestaat uit een stap Δy_1 in de richting van de vector $Q^{-1}N$ tot aan de beperking

$$\Delta y_1 := Q^{-1}N(N^TQ^{-1}N)^{-1}b \tag{3.3.8a}$$

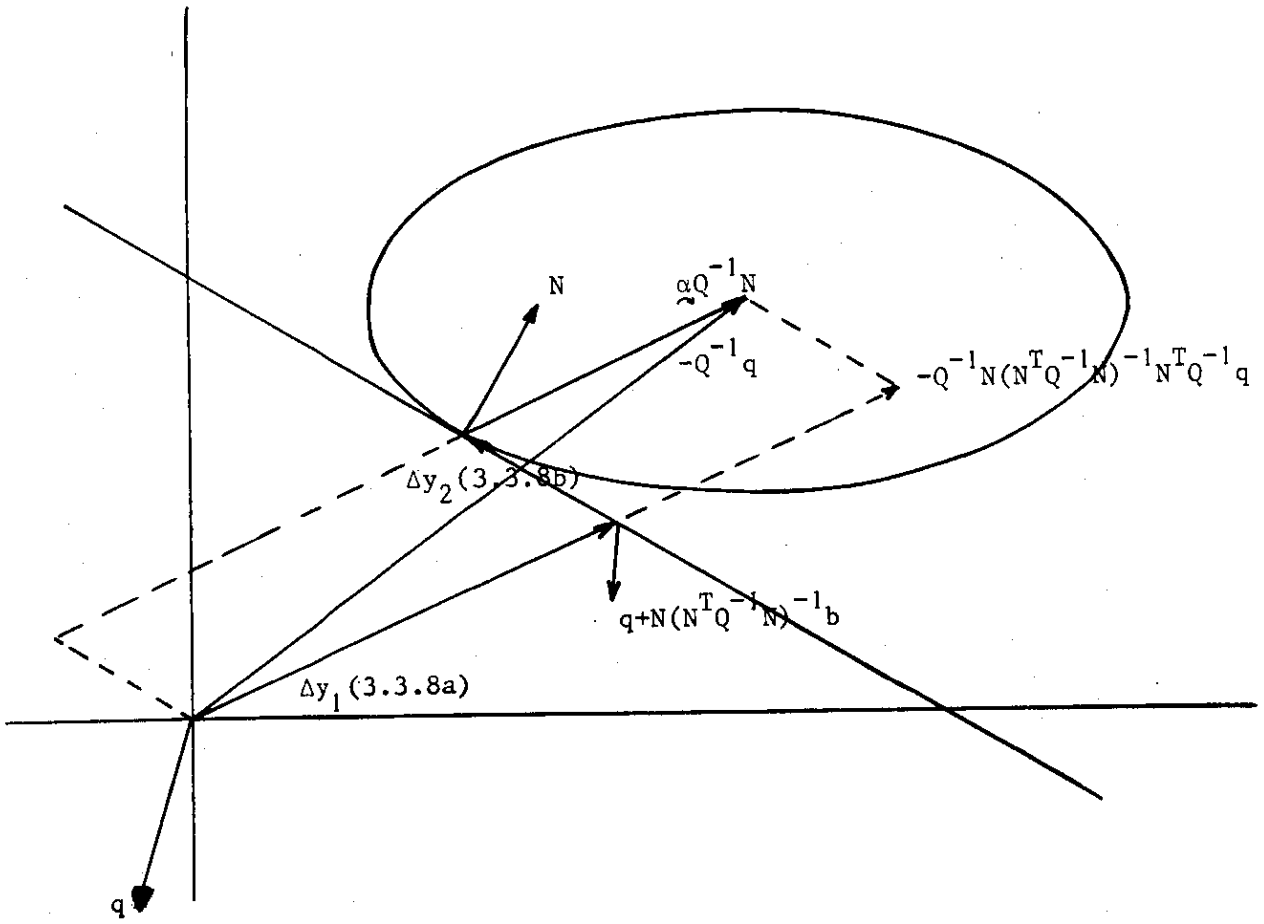
gevolgd door een stap Δy_2 in het vlak van de beperking gelijk aan

$$\begin{aligned}\Delta y_2 &:= -(I - Q^{-1}N(N^TQ^{-1}N)^{-1}N^T)Q^{-1}(q + N(N^TQ^{-1}N)^{-1}b) \\ &= -(I - Q^{-1}N(N^TQ^{-1}N)^{-1}N^T)Q^{-1}\nabla f(\Delta y_1)\end{aligned} \tag{3.3.8b}$$

waarin

$$\nabla f(\Delta y_1) := q + Q\Delta y_1 = q + N(N^TQ^{-1}N)^{-1}b$$

de locale gradiënt is ter plaatse waar de stap Δy_1 eindigt.



Figuur 3.3.3 : Eerste oplossing van het probleem QLE.

3.3.4. De uitdrukkingen (3.3.7) en (3.3.8) voor de oplossing van het lineaire stelsel (3.3.5)

$$\begin{pmatrix} Q & -N \\ N^T & 0 \end{pmatrix} \begin{pmatrix} y \\ \lambda \end{pmatrix} = \begin{pmatrix} -q \\ b \end{pmatrix}$$

hadden ook gevonden kunnen worden door directe invertering van de coëfficiënten matrix. Zoals eenvoudig kan worden geverifieerd geldt daarvoor

$$\begin{pmatrix} Q & -N \\ N^T & 0 \end{pmatrix}^{-1} = \begin{pmatrix} Q^{-1} - Q^{-1}N(N^TQ^{-1}N)^{-1}N^TQ^{-1} & Q^{-1}N\,(N^TQ^{-1}N)^{-1} \\ -(N^TQ^{-1}N)^{-1}N^TQ^{-1} & (N^TQ^{-1}N)^{-1} \end{pmatrix}$$

(3.3.10)

Opgemerkt kan worden dat er tussen deze uitdrukking en de uitdrukkingen (3.2.109) t/m (3.2.112) uit Stelling 3.2.24 (uiteraard) een grote verwantschap bestaat.

Oplossing van het probleem QLE (II)

3.3.5. De tweede bekende formulering van de oplossing van het probleem QLE (3.3.2)

$$\min\{q^T y + \tfrac{1}{2} y^T Q y \mid N^T y - b = 0\}$$

maakt gebruik van de coördinaten transformatie gebaseerd op de aan het probleem gerelateerde, in pt. 3.1.8 besproken matrix $[N \mid Z]$, d.w.z. van de transformatie

$$y := Nv + Zw \tag{3.3.11}$$

Herschrijven van het probleem QLE (3.3.2) in termen van de q-vector v en (n-q) vector w geeft na enige rangschikking

$$\min\{\, q^T Nv + \tfrac{1}{2} v^T N^T Q N v + (q + QNv)^T Zw + \tfrac{1}{2} w^T Z^T Q Z w \mid \mid N^T Nv - b = 0\} \tag{3.3.12}$$

uit welke formulering direct als oplossing volgt voor v en w

$$\begin{aligned} v &= (N^T N)^{-1} b \\ w &= -(Z^T Q Z)^{-1} Z^T (q + QN(N^T N)^{-1} b) \end{aligned} \tag{3.3.13}$$

In de terminologie van het oorspronkelijke probleem volgt hieruit met (3.3.11) als oplossing

$$y = N(N^T N)^{-1} b - Z(Z^T Q Z)^{-1} Z^T (q + QN(N^T N)^{-1} b) \tag{3.3.14}$$

Een illustratie van deze oplossing voor het geval dat $m = 1$ en $n = 2$ is gegeven in Figuur 3.3.5. Opgemerkt kan worden dat de oplossing in dit geval opgebouwd gedacht kan worden uit een stap Δy_1 vanuit de oorsprong naar de lineaire variëteit bepaald door de actieve beperkingen

$$\Delta y_1 = N(N^T N)^{-1} b \tag{3.3.15}$$

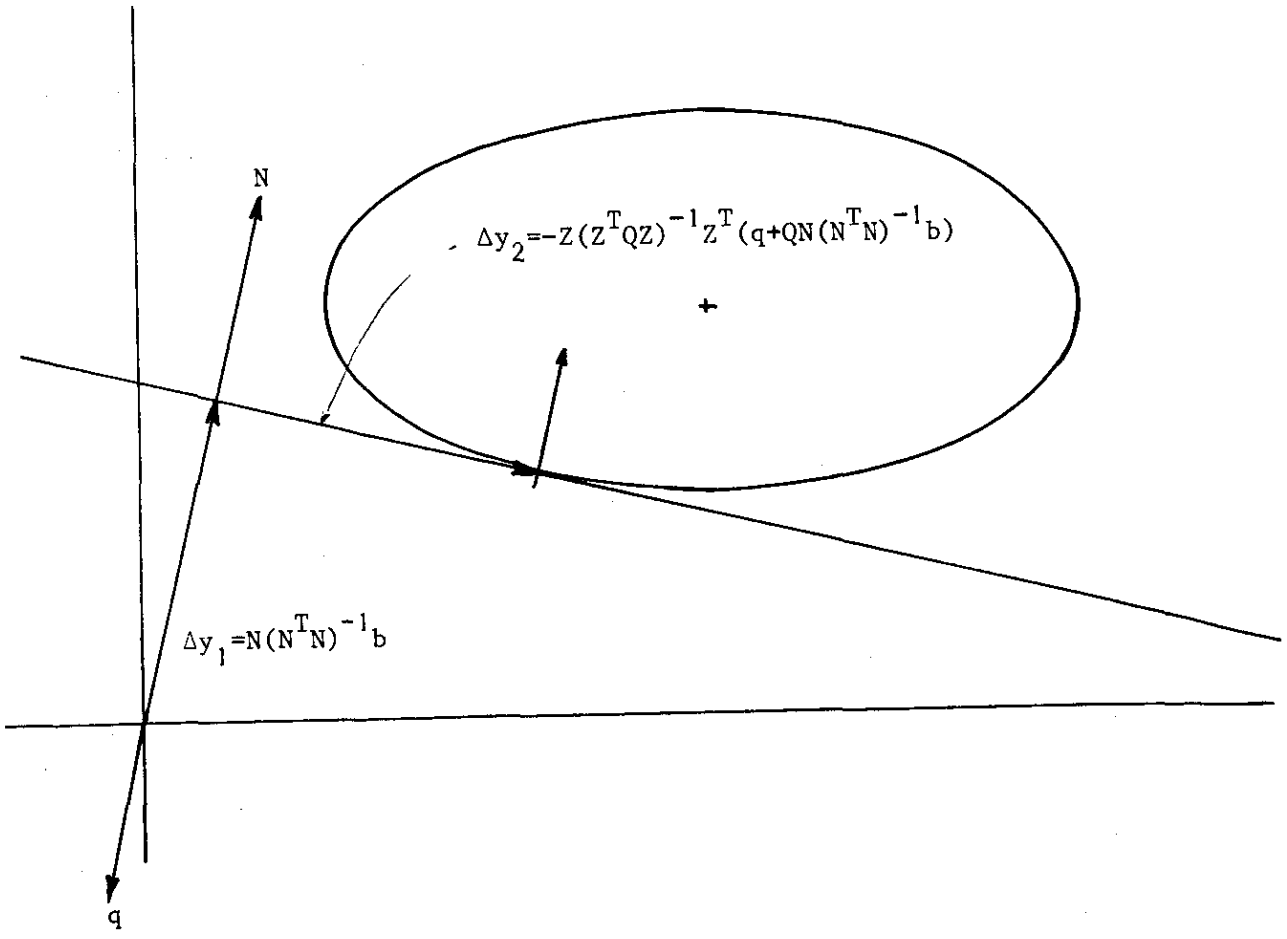
gevolgd door een stap Δy_2 binnen de lineaire variëteit gelijk aan

$$\Delta y_2 := -Z(Z^T Q Z)^{-1} Z^T \nabla f(\Delta y_1) \tag{3.3.16}$$

waarin

$$\nabla f(\Delta y_1) := q + Q\Delta y_1 = q + QN(N^T N)^{-1} b$$

weer de locale gradiënt is ter plaatse waar de stap Δy_1 eindigt



Figuur 3.3.5 : Tweede oplossing van het probleem QLE.

3.3.6. Inplaats van de coördinaten transformatie (3.3.11) gebaseerd op de matrix [N ¦ Z] waarin Z de matrix is met als kolommen de orthonormale basisvectoren van het orthogonale complement van de deelruimte opgespannen door de kolommen van de matrix N had evengoed een analoge coördinaten transformatie gebruikt kunnen worden gebaseerd op de matrix [N ¦ M] waarin M de in verband met de gereduceerde-gradiënt methode in pt. 3.2.17 besproken matrix voorstelt die werd gedefinieerd door (3.2.54)

$$M := \begin{pmatrix} -B^{-1}D \\ I \end{pmatrix}$$

en waarvoor eveneens geldt dat (vgl. (3.2.55))

$$N^T M = 0_{q\times(n-q)}$$

De kolommen $m_i, i = 1,\ldots,n-q$ van deze matrix vormen een andere basis voor dezelfde deelruimte als waarvoor de kolommen van de matrix Z een basis zijn. Het voornaamste verschil tussen de matrices M en Z is dat de kolommen van M niet genormaliseerd zijn zodat in het algemeen

$$M^T M \neq I_{(n-q)\times(n-q)} \tag{3.3.17}$$

Normaliseren van de matrix M door navermenigvuldiging met de matrix $(M^T M)^{-\frac{1}{2}}$ geeft een matrix die juist de eigenschappen van de niet-eenduidig gedefinieerde matrix Z bezit en deze in alle voorgaande en nog volgende afleidingen kan vervangen

$$M(M^T M)^{-\frac{1}{2}} \leftrightarrow Z \tag{3.3.18}$$

Ter vermijding van duplicatie wordt hierna bij de nog te bespreken toepassingen van de tweede oplossing van het probleem QLE (3.3.14) en de daarmee samenhangende theoretische beschouwingen steeds de matrix Z gebruikt in die situaties waar zowel de matrix Z als de matrix M kunnen worden gebruikt met dien verstande dat waar nodig ook de matrix $Z^T Z$ expliciet genoteerd wordt als gold niet dat $Z^T Z = I$

Newton-methoden (en modificaties daarvan) bij lineaire nevenvoorwaarden

3.3.7. In het geval dat b = 0 in het probleem QLE, d.i. (vgl. (3.3.3)) in het geval dat het startpunt voor het betreffende probleem toegelaten is, gaan de beide oplossingen (3.3.8) en (3.3.14) voor het probleem QLE over in respectievelijk

$$y := -(I - Q^{-1}N(N^TQ^{-1}N)^{-1}N^T)Q^{-1}q \qquad (3.3.19)$$

en

$$y := -Z(Z^TQZ)^{-1}Z^Tq \qquad (3.3.20)$$

Deze uitdrukkingen vormen, analoog aan de situatie bij de onbeperkte minimalisering (vgl. (2.5.6)), de basis voor de formulering van zoekrichtingen gebruikt in de methode van Newton voor minimaliseringsproblemen met (lineaire) nevenvoorwaarden. Voor deze zoekrichtingen, ter onderscheiding aangeduid met de indices I en II, volgt (vgl. (2.5.3)) respectievelijk

$$d_I^{(k)} := -(I - G^{(k)-1}N^{(k)}(N^{(k)T}G^{(k)-1}N^{(k)})^{-1}N^{(k)T})G^{(k)-1}\nabla f(x^{(k)}) \qquad (3.3.21)$$

waar

$G^{(k)} := G(x^{(k)})$: Hessiaan van de objectfunctie in $x^{(k)}$

$N^{(k)}$: matrix van normalen van actieve beperkingen in $x^{(k)}$

en

$$d_{II}^{(k)} := -Z^{(k)}(Z^{(k)T}G^{(k)}Z^{(k)})^{-1}Z^{(k)T}\nabla f(x^{(k)}) \qquad (3.3.22)$$

waar

$Z^{(k)}$: matrix met als kolommen de eventueel orthonormale (vgl. pt. 3.3.6) basisvectoren van het orthogonale complement van de ruimte opgespannen door de kolommen van $N^{(k)}$

3.3.8. Analoog aan de situatie m.b.t. de twee verschillende formuleringen voor dezelfde projectie matrix bij de gewone geprojecteerde-gradiënt-methode (vgl. pt. 3.2.11) waar gold dat

$$(I - N(N^TN)^{-1}N^T) = Z(Z^TZ)^{-1}Z^T \tag{3.3.23}$$

kan ook hier worden aangetoond dat de matrices uit (3.3.21)

$$\bar{P}_I := (I - G^{-1}N(N_Q^{T\,-1}N)^{-1}N^T)G^{-1} \tag{3.3.24}$$

en (3.3.22)

$$\bar{P}_{II} := Z(Z^TGZ)^{-1}Z^T \tag{3.3.25}$$

(waarin in beide gevallen de indices (k) werden weggelaten) in het veronderstelde geval dat G positief definiet is, beide representaties zijn van dezelfde scheve projectieoperator. Deze projectie operator reduceert vectoren in ruimte opgespannen door de kolommen van de matrix N tot nul en opereert op vectoren in de deelruimte opgespannen door de kolommen van de matrix GZ als de inverse matrix G^{-1}, ofwel in formule vorm (vgl. ook pt. 2.6.22)

$$\bar{P}_I[N\,|\,GZ] = \bar{P}_{II}[N\,|\,GZ] = [O\,|\,Z] \tag{3.3.26}$$

Aangezien de kolommen van de matrix [N|GZ] in het geval dat G niet-singulier is evengoed een basis voor de ruimte $\mathbb{R}^n$ vormen als de kolommen van de matrix [N|Z] volgt dat inderdaad ook geldt

$$\bar{P}_I := (I - G^{-1}N(N^TG^{-1}N)^{-1}N^T)G^{-1} = Z(Z^TGZ)^{-1}Z^T =: \bar{P}_{II} \tag{3.3.27}$$

Hieruit volgt dat beide zoekrichtingen d_I (3.3.21) en d_{II} (3.3.22) in het geval dat G niet-singulier is exact aan elkaar gelijk zijn. De op deze zoekrichtingen gebaseerde Newton-methoden verschillen in dat geval alleen van elkaar in de manier waarop de informatie over het raakvlak aan de actieve beperkingen in het punt $x^{(k)}$ wordt bewaard.

3.3.9. Van de beide formuleringen (3.3.21) en (3.3.22) voor de zoekrichting van de methode van Newton verdient de tweede in de practijk de voorkeur en wel om reden van o.a. grotere toepasbaarheid, grotere eenvoud en sterkere analogie met de formulering van de methode van Newton voor onbeperkte minimalisering (zie paragraaf 2.5). De grotere toepasbaarheid is duidelijk (en van belang) in de niet zelden voorkomende gevallen waar de matrix Z^TGZ positief definiet is en de matrix G singulier of bijna-singulier. De grotere eenvoud en de sterkere analogie van de tweede formulering met de formulering van de methode van Newton voor onbeperkte minimalisering maken het mogelijk dezelfde procedures te volgen en dezelfde modificaties van de originele methode te gebruiken als besproken in paragraaf 2.5. (pt. 2.5.12-2.5.22) in die gevallen waar de matrix Z^TGZ niet strikt positief definiet is. Uitgangspunt daarbij is dat de zoekrichting $d_{II}^{(k)}$ gegeven wordt door de uitdrukking

$$d_{II}^{(k)} := Z^{(k)} d_Z^{(k)} \tag{3.3.28}$$

waar $d_Z^{(k)}$ in principe bepaald wordt als oplossing van de vergelijking

$$G_Z^{(k)} d_Z^{(k)} = -g_Z^{(k)} \tag{3.3.29}$$

waar

$$G_Z^{(k)} := Z^{(k)T} G(x^{(k)}) Z^{(k)} \tag{3.3.30}$$

$$g_Z^{(k)} := Z^{(k)T} \nabla f(x^{(k)}) \tag{3.3.31}$$

De "geprojecteerde Hessiaan" $G_Z^{(k)}$ is een $(n-q)\times(n-q)$-matrix en de "geprojecteerde gradiënt" $g_Z^{(k)}$ een $(n-q)$-vector en de bepaling van $d_Z^{(k)}$ verloopt dan ook geheel analoog aan de situatie als betrof het een toepassing van de methode van Newton in $\mathbb{R}^{n-q}$. In het bijzonder geldt dit ook voor de drie besproken modificaties van Goldfeld, Quandt en Trotter (pt. 2.5.14), Greenstadt (pt. 2.5.16) en Fiacco-McCormick (pt. 2.5.20) in die gevallen waar de locale geprojecteerde Hessiaan $G_Z^{(k)}$ niet voldoende positief definiet is.

3.3.10. Voor het geval dat voor de oplossing van (3.3.29) gebruik gemaakt wordt van een Choleski-decompositie (vgl.pt. 2.5.21)

$$G_Z^{(k)} := L_Z^{(k)} D_Z^{(k)} L_Z^{(k)T} \tag{3.3.32}$$

met als op te lossen driehoeksstelsels (als D_Z niet singulier is)

$$\begin{aligned} L_Z^{(k)} v &= -g_Z^{(k)} \\ L_Z^{(k)T} d_Z^{(k)} &= D_Z^{(k)-1} v \end{aligned} \tag{3.3.33}$$

werden door Gill en Murray ([3.3.4] en [3.3.5]) procedures ontwikkeld om direct de factorisatie matrices $D_Z^{(k)}$ en $L_Z^{(k)}$ aan te passen in het geval in de k-de stap een oude beperking uit de verzameling van actieve beperkingen moet worden verwijderd dan wel een nieuwe beperking daaraan toegevoegd. Voor de details van die procedures als ook voor de details van een door dezelfde auteurs ontwikkelde modificatie van de methode van Newton die het midden houdt tussen de modificatie van Goldfeld, Quandt en Trotter en de modificatie van Fiacco-McCormick wordt de lezer verwezen naar [3.3.4] . Van de modificatie van de methode van Newton kan worden opgemerkt dat deze erop neerkomt dat tijdens de Choleski-decompositie de diagonaal elementen van de matrix $D_Z^{(k)}$ zodanig worden verhoogd dat een decompositie resulteert van een positief definiete matrix $\bar{G}_Z^{(k)}$ gegeven door

$$\bar{G}_Z^{(k)} := L_Z^{(k)} D_Z^{(k)} L_Z^{(k)T} := G_Z^{(k)} + E_Z^{(k)} \tag{3.3.34}$$

waarin $E_Z^{(k)}$ een diagonaal matrix is.

De genoemde publicatie [3.3.4] van Gill en Murray bevat naast de beschrijving van deze modificatie ook een voorschrift voor een actieve-set-strategie en een voorschrift voor de berekening van de matrix Z met QR-decompositie en daarmee alle bestanddelen van een compleet practisch algorithme voor de toepassing van de methode van Newton in het geval van lineair beperkte problemen. Een vergelijkbare beschrijving voor een soortgelijk compleet algorithme van de methode van Newton voor lineaire beperkingen waarbij gebruik gemaakt wordt van de matrix M (3.2.54) i.p.v. de matrix Z (vgl. pt. 3.3.6) werd gegeven door McCormick [3.3.11].

Als modificatie in het geval van een niet voldoende positief definiete matrix $\bar{G}^{(k)}$ (of beter $G^{(k)}$) koos McCormick een methode die gebaseerd is op een analyse van de eigenwaarden en als actieve-set-strategie een door hem zelf (vgl. [3.3.12]) ontwikkelde strategie die hij de "bending" strategie noemde. Als naam voor zijn algorithme koos McCormick de naam " variable-reduction-method". Voor de practijk (vgl. [3.3.1]) lijkt de eerste van deze complete algorithmen, nl. die van Gill en Murray de meeste voordelen te bieden.

Quasi-Newton methoden bij lineaire nevenvoorwaarden

3.3.11. Om dezelfde redenen als in het geval van onbeperkte minimaliseringsproblemen (vgl. pt. 2.7.1),d.w.z. in het bijzonder ter omzeiling van het probleem van het evalueren van alle elementen van de Hessiaan in iedere iteratiestap, heeft men getracht op analoge wijze quasi-Newton methoden te ontwikkelen voor het geval dat er lineaire nevenvoorwaarden zijn. Deze representeren in meerdere opzichten een tussenvorm tussen de in de vorige paragraaf besproken eerste-orde geprojecteerde- (en gereduceerde) gradiënt methoden en de hiervoor in deze paragraaf besproken methoden van Newton. Juist als bij deze laatste zijn ook de bekende quasi-Newton methoden voor minimaliseringsproblemen met lineaire beperkingen onder te verdelen in twee klassen afhankelijk of voor het genereren van een zoekrichting een benadering gebruikt wordt van de formule voor $d_I^{(k)}$ (3.3.21)

$$d_I^{(k)} := -(I - G^{(k)-1}N_q^{(k)}(N_q^{(k)T}G^{(k)-1}N_q^{(k)})^{-1}N_q^{(k)T})G^{(k)-1}g^{(k)}$$

of van de formule voor $d_{II}^{(k)}$ (3.3.22)

$$d_{II}^{(k)} := -Z_q^{(k)}(Z_q^{(k)T}G^{(k)}Z_q^{(k)})^{-1}Z_q^{(k)T}g^{(k)}$$

in welke uitdrukkingen gebruik werd gemaakt van een onder index q om aan te geven dat de "huidige" actieve set q lineair onafhankelijke beperkingen omvat. In plaats van de echte Hessiaan of zijn inverse maken de quasi-Newton methoden gebruik van benaderingen daarvan of van benaderingen van bepaalde producten van matrices in de gegeven uitdrukkingen. Deze benaderingen worden in iedere iteratiestap aangepast en behoeven in het algemeen niet helemaal opnieuw berekend te worden. Hieronder zullen drie van

de meest bekende quasi-Newton methoden voor het geval van lineaire nevenvoorwaarden worden besproken en wel de methode van Goldfarb [3.3.8] en de methode van Murtagh en Sargent [3.3.13], beide behorend tot de categorie van methoden die de uitdrukking voor $d_I^{(k)}$ trachten te benaderen en de methode van Gill en Murray [3.3.5] die het genereren van de zoekrichting $d_{II}^{(k)}$ tracht te benaderen. Daarna wordt nog ingegaan op het verband tussen deze in de literatuur zeer bekende methoden en de klasse van methoden die bestaat uit het direct toepassen van geconjugeerde-richtingen en quasi-Newton methoden in $\mathbb{R}^{n-q}$. Verondersteld bij al deze benaderingen wordt dat het uitgangspunt, waar de zoekrichting wordt bepaald, een toegelaten punt is.

Methode van Goldfarb

3.3.12. Teneinde het in elke stap berekenen van de Hessiaan en zijn inverse te vermijden suggereerde Goldfarb in 1969 (zie [3.3.8]) de symmetrische matrix

$$\bar{P}_{I,q}^{(k)} := (I - G^{(k)-1}N_q^{(k)}(N_q^{(k)T}G^{(k)-1}N_q^{(k)})^{-1}N_q^{(k)T})G^{(k)-1} \qquad (3.3.35)$$

te vervangen door een benaderings matrix $K_q^{(k)}$ en deze te gebruiken voor het genereren van de zoekrichting volgens de formule

$$d^{(k)} := -K_q^{(k)}g^{(k)} \qquad (3.3.36)$$

Het probleem wordt dan een start matrix $K_q^{(0)}$ te definiëren voor de eerste stap en vervolgens met behulp van de in iedere stap verkregen informatie de matrix $K_q^{(k)}$ zo aan te passen dat deze een zo goed mogelijk benadering wordt voor de matrix $\bar{P}_{I,q}^{(k)}$ in het punt $x^{(k)}$. Als start matrix ligt voor de hand de keuze de orthogonale projectie matrix

$$K_q^{(0)} := I - N_q^{(0)}(N_q^{(0)T}N_q^{(0)})^{-1}N_q^{(0)T} \qquad (3.3.37)$$

De zoekrichting in de eerste stap is dan, geheel analoog aan de situatie bij de onbeperkte minimalisering, gelijk aan de gewone geprojecteerde gradiënt. Voor de aanpassing van $K_q^{(k)}$ na de (k+1)-de stap worden drie gevallen onderscheiden:

(i) De verzameling van actieve (lineaire) beperkingen is dezelfde aan het einde van de k-de stap als aan het begin van die stap

(ii) De verzameling van actieve (lineaire) beperkingen moet na de k-de stap worden uitgebreid van q tot (q + 1) beperkingen

(iii) De verzameling van actieve (lineaire) beperkingen moet aan het begin van de k-de stap worden ingekrompen van q tot (q - 1) beperkingen.

Voor ieder van deze drie gevallen werd door Goldfarb een andere aanpassingsformule voorgeschreven en wel in geval (i):

$$K_q^{(k+1)} := K_q^{(k)} + \frac{s^{(k)} s^{(k)T}}{s^{(k)T} y^{(k)}} - \frac{K_q^{(k)} y^{(k)} y^{(k)T} K_q^{(k)}}{y^{(k)T} K_q^{(k)} y^{(k)}} \qquad (3.3.38)$$

in geval (ii) (als n_{q+1} de normaal is van de nieuwe actieve beperking):

$$K_{q+1}^{(k+1)} := K_q^{(k)} - \frac{K_q^{(k)} n_{q+1} n_{q+1}^T K_q^{(k)}}{n_{q+1}^T K_q^{(k)} n_{q+1}} \qquad (3.3.39)$$

en in geval (iii) (als n_q de normaal is van de niet langer actief verondersteld beperking):

$$K_{q-1}^{(k)} := K_q^{(k)} + \frac{\hat{P}_{q-1} n_q n_q^T \hat{P}_{q-1}}{n_q^T \hat{P}_{q-1} n_q} \qquad (3.3.40)$$

waarbij

$$\hat{P}_{q-1} := I - N_{q-1}^{(k)} (N_{q-1}^{(k)T} N_{q-1}^{(k)})^{-1} N_{q-1}^{(k)T} \qquad (3.3.41)$$

waar $N_{q-1}^{(k)}$ gelijk is aan de matrix $N_q^{(k)}$ met daaruit de kolom n_q verwijderd.

Voor de genoemde aanpassingsformules werd door Goldfarb de volgende argumentatie gegeven

(i) De aanpassingsformule (3.3.38) is juist de uit de onbeperkte minimalisering bekende DFP-aanpassingsformule (2.8.2) die gebruikt wordt

voor het incorporeren van informatie over de inverse van de Hessiaan in de benaderingsmatrix H. Precies als in het onbeperkte minimaliseringsgeval geldt voor $K_q^{(k+1)}$ de quasi-Newton-relatie

$$K_q^{(k+1)} y^{(k)} = s^{(k)} \tag{3.3.42}$$

Bovendien geldt dat als

$$K_q^{(k)} n_j = 0 \qquad j = 1,\ldots q \tag{3.3.43a}$$

dat dan ook

$$K_q^{(k+1)} n_j = 0 \qquad j = 1,\ldots q \tag{3.3.43b}$$

Aangetoond kan worden dat het gebruik van deze aanpassingsformules met als startmatrix de matrix (3.3.37) en van stapgroottebepaling door exacte lijnminimalisering, in het geval van de minimalisering van een positief definiete kwadratische vorm met q lineaire nevenvoorwaarden in ten hoogste n-q stappen tot de oplossing van het beperkte minimaliseringsprobleem leidt. Opgemerkt kan worden dat de keuze van de DFP-aanpassingsformule in de methode van Goldfarb nog stamt uit de tijd toen de DFP-aanpassingsformule als de beste quasi-Newtonformule gold. In principe kan voor iedere symmetrische aanpassingsformule uit de klasse van Huang (vgl. pt. 2.7.10) de eigenschap (3.3.43) worden bewezen. Ieder van deze aanpassingsformules kan dan daarom gebruikt worden i.p.v. (3.3.38). Aan het einde van deze paragraaf (pt. 3.3.16) zal hierop nader worden ingegaan.
(ii) De aanpassingsformule (3.3.39) wordt gebruikt in die gevallen waarbij het eendimensionale zoekproces stopt omdat een eerdere passieve beperking actief wordt. Omdat geen lijnminimum gevonden wordt, wordt geen gebruik gemaakt van de DFP-aanpassingsformule. In plaats daarvan wordt een aanpassingsformule gebruikt die de eerdere informatie over de inverse Hessiaan in de richting n_{q+1} elimineert uit de benaderingsmatrix en iedere vector in $\mathbb{R}^n$ projecteert op de doorsnede van het raakvlak aan de oude actieve beperkingen met de nieuwe actieve beperking. Aangetoond kan worden dat bij toepassing van de aanpassingsformule (3.3.39) bij de minimalisering van een positief definiet kwadratische vorm met lineaire nevenvoorwaarden (Probleem QLI (3.1.19)) geldt dat als (vgl. (3.3.8))

$$K_q^{(k)} = \bar{P}_{I,q}^{(k)} := (I - Q^{-1}N_q^{(k)}(N_q^{(k)T}Q^{-1}N_q^{(k)})^{-1}N_q^{(k)T})Q^{-1} \tag{3.3.44}$$

dat dan ook voor $K_{q+1}^{(k+1)}$ bepaald met (3.3.39) geldt

$$K_{q+1}^{(k+1)} = \bar{P}_{I,q+1}^{(k+1)} := (I - Q^{-1}N_{q+1}^{(k+1)}(N_{q+1}^{(k+1)T}Q^{-1}N_{q+1}^{(k+1)})^{-1}N_{q+1}^{(k+1)T})Q^{-1} \tag{3.3.45}$$

Verder is eenvoudig in te zien dat als

$$K_q^{(k)}N_q^{(k)} = 0_{n\times q} \tag{3.3.46}$$

dat dan ook

$$K_{q+1}^{(k+1)}[N_q^{(k)} \mid n_{q+1}] = K_{q+1}^{(k+1)}N_{q+1}^{(k+1)} = 0_{n\times(q+1)} \tag{3.3.47}$$

(iii) De aanpassingsformule (3.3.40) wordt toegepast indien in het begin van de k-de iteratiestap blijkt dat geen of slechts relatief geringe vooruitgang kan worden geboekt langs de in eerste instantie met (3.3.36) bepaalde zoekrichting $-K_q^{(k)}g^{(k)}$. In dat geval kan beter worden overgegaan tot het passief maken van een tot dan toe actieve beperking. Goldfarb [3.3.8] besluit in navolging van Rosen [3.3.2] daartoe indien blijkt dat

$$\|K_q^{(k)}g^{(k)}\| \leq -\tfrac{1}{2}\min_{j\in I_A}\{0,\alpha_j b_{jj}^{-\frac{1}{2}}\} \approx -\tfrac{1}{2}\min_j\{g^{(k)T}n_j/\|n_j\|\} \tag{3.3.48}$$

waar

α_j : j-de component van $(N_q^{(k)T}N_q^{(k)})^{-1}N_q^{(k)T}g^{(k)}$

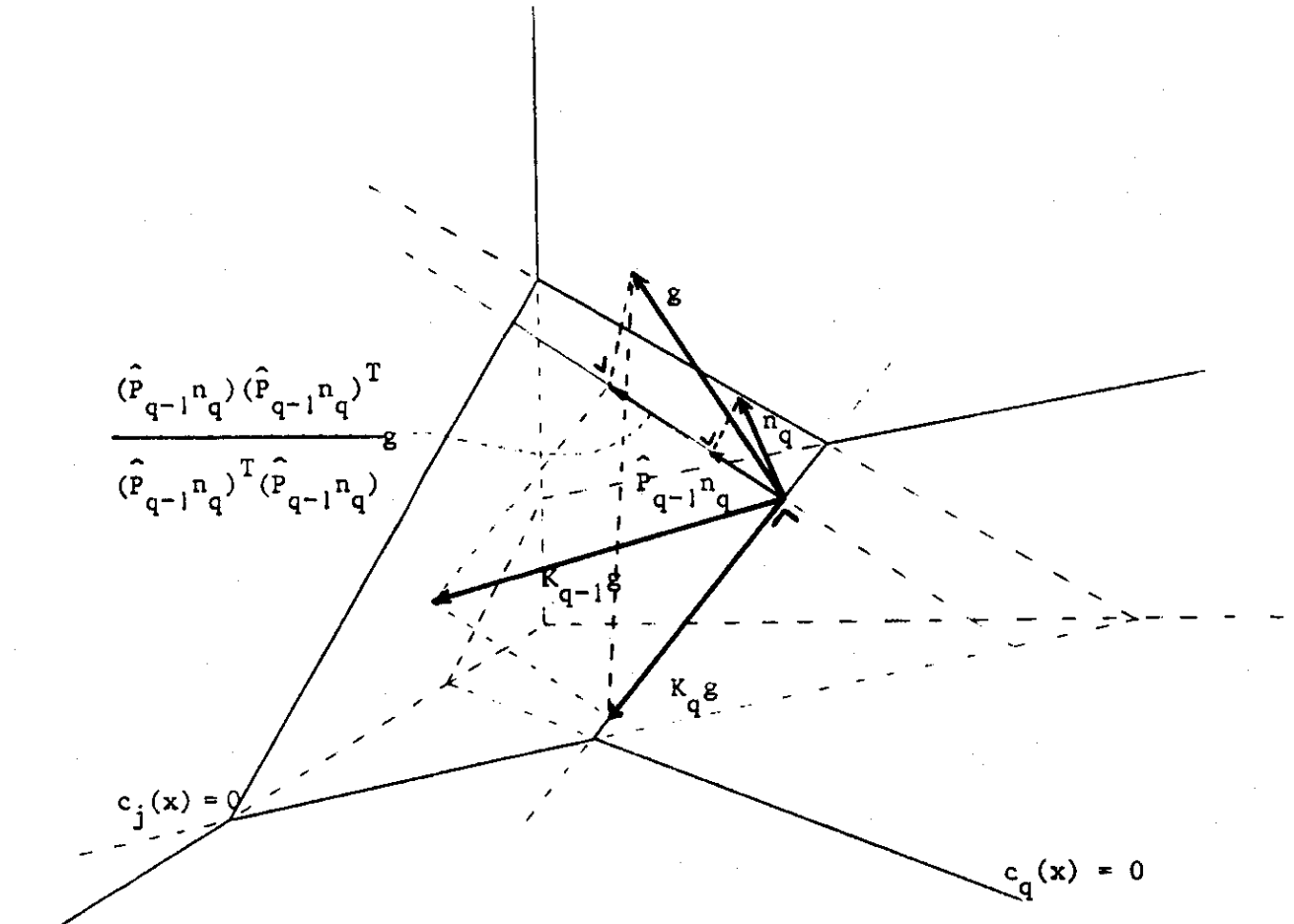
b_{jj} : j-de diagonaal-element van $(N_q^{(k)T}N_q^{(k)})^{-1}$

(Voor aanvullende opmerkingen over deze test, zie [3.3.1] p. 83)

De aanpassingsformule zorgt ervoor dat niet langer op de verlaten beperking wordt geprojecteerd. Juist als de matrix $K_q^{(k)}$ bevat de nieuwe matrix $K_{q-1}^{(k)}$ echter geen informatie over het gedrag van de inverse Hessiaan in de richting van de normaal n_q. Voor de nieuwe zoekrichting met de in (3.3.40) gedefinieerde orthogonale projectiematrix $\hat{P}_{q-1}$ geldt

$$d^{(k)} := -K_{q-1}{}^{(k)} g^{(k)} = -\left(K_q{}^{(k)} g^{(k)} + \frac{(\hat{P}_{q-1} n_q)(\hat{P}_{q-1} n_q)^T}{(\hat{P}_{q-1} n_q)^T (\hat{P}_{q-1} n_q)} g^{(k)} \right) \tag{3.3.49}$$

Dat wil zeggen dat de zoekrichting gebaseerd op de informatie in de doorsnede van de originele beperkingen wordt aangevuld met de orthogonale projectie van de gradiënt op de orthogonale projectie van de q-de normaal op de doorsnede van de q-1 resterende actieve beperkingen. Figuur 3.3.12 illustreert deze gang van zaken aan de hand van een uit [3.3.3] overgenomen schets.



Figuur 3.3.12.: Het inaktief (passief) maken van de q^e beperking volgens de methode van Goldfarb.

Methode van Murtagh en Sargent

3.3.13. Een tweede manier om de matrix $\bar{P}_{I,q}^{(k)}$ (3.3.35) te benaderen werd in 1969 gesuggereerd door Murtagh en Sargent in [3.3.13]. Deze manier bestaat daaruit dat gebruik gemaakt wordt van een benadering $H^{(k)}$ voor $G^{(k)-1}$ zoals bij de quasi-Newton-methoden voor onbeperkte minimalisering. De benaderingsmatrix wordt daarmee

$$K_q^{(k)} := (I - H^{(k)}N_q^{(k)}(N_q^{(k)T}H^{(k)}N_q^{(k)})^{-1}N_q^{(k)T})H^{(k)} \qquad (3.3.50)$$

De matrix $H^{(k)}$ wordt na iedere stap aangepast met behulp van de algemene quasi-Newton-aanpassingsformule (vgl. pt. 2.7.7)

$$H^{(k+1)} := H^{(k)} + \frac{(s^{(k)} - H^{(k)}y^{(k)})v^{(k)T}}{v^{(k)T}y^{(k)}} \qquad (3.3.51)$$

waar in het bijzonder twee keuzen voor $v^{(k)}$ door Murtagh en Sargent werden beschouwd

(i) $v^{(k)}$ zodat $N^{(k)T}v^{(k)} = 0$ (3.3.52)

(ii) $v^{(k)} := s^{(k)} - H^{(k)}y^{(k)}$ (3.3.53)

De eerste keuze heeft als voordeel dat in het geval dat geen verandering optreedt in de verzameling van actieve beperkingen (d.i. als $N_q^{(k+1)} = N_q^{(k)}$) dat dan zowel

$$H^{(k+1)}N_q^{(k+1)} = H^{(k)}N_q^{(k)} \qquad (3.3.54)$$

als

$$\left(N_q^{(k+1)T}H^{(k+1)}N_q^{(k+1)}\right)^{-1} = \left(N_q^{(k)T}H^{(k)}N_q^{(k)}\right)^{-1} \qquad (3.3.55)$$

zodat de nieuwe matrix $K_q^{(k+1)}$ eenvoudig berekend kan worden uit

$$K_q^{(k+1)} := P_q^{(k+1)}H^{(k+1)} = P_q^{(k)}H^{(k+1)} \qquad (3.3.56)$$

De matrix $H^{(k+1)}$ zal bij deze keuze in het algemeen niet symmetrisch zijn met alle problemen vandien.

De tweede keuze voor $v^{(k)}$ in (3.3.51) resulteert in de bekende in pt. 2.8.1 t/m 2.8.5 besproken rang-één-aanpassingsformule(2.8.1)

$$H^{(k+1)} := H^{(k)} + \frac{(s^{(k)} - H^{(k)}y^{(k)})(s^{(k)} - H^{(k)}y^{(k)})^T}{(s^{(k)} - H^{(k)}y^{(k)})^T y^{(k)}}$$

Deze heeft het voordeel (vgl. pt. 2.8.2) dat bij de minimalisering van een positief definiet kwadratische vorm steeds A-geconjugeerde richtingen worden gegenereerd zonder lijnminimalisering. Dit resultaat is vooral van belang omdat de stapgrootte factoren bij problemen met ongelijkheidsvoorwaarden vaak worden bepaald door het actief worden van eerdere passieve beperkingen.

Voor het geval de verzameling van actieve beperkingen veranderd moet de matrix $P_{q+1}^{(k+1)}$ worden aangepast. Hiervoor werden door Murtagh en Sargent in [3.3.13] in navolging van Rosen [3.3.2] recursieformules gegeven die het niet nodig maken de matrix $P_{q+1}^{(k+1)}$ helemaal opnieuw te bepalen. Juist als bij de methode van Goldfarb geldt dat bij uitbreiding van de verzameling van actieve beperkingen de aanwezige informatie "geprojecteerd" wordt op de nieuwe beperkingen en dat bij het passief worden van een eerdere actieve beperking geen (of vrijwel geen) informatie voorhanden is over het gedrag van de inverse Hessiaan in de richting van de normaal van de te verlaten beperking.

Methode van Gill en Murray

3.3.14. In plaats van een quasi-Newton benadering van de matrix $\bar{P}_I$ suggereerden Gill en Murray in 1972 in [3.3. 4] een quasi-Newton benadering van de matrix (3.3.25)

$$\bar{P}_{II} := Z^{(k)}(Z^{(k)T}G^{(k)}Z^{(k)})^{-1}Z^{(k)T}$$

en wel door benadering van de matrix $G^{(k)}$ door een matrix $B^{(k)}$ gevolgd door de bepaling van een zoekrichting $d^{(k)}$ door oplossing van het lineaire stelsel

$$(Z^{(k)T}B^{(k)}Z^{(k)})d_Z^{(k)} = -Z^{(k)T}g^{(k)} \tag{3.3.57}$$

en het substitueren van de oplossing $d_Z^{(k)} \in \mathbb{R}^{n-q}$ in de uitdrukking voor de zoekrichting in $\mathbb{R}^n$

$$d^{(k)} := Z^{(k)}d_Z^{(k)} \tag{3.3.58}$$

of, equivalent met behulp van de aan (3.3.30) en (3.3.31) analoge definities

$$B_Z^{(k)} := Z^{(k)T}B^{(k)}Z^{(k)} \tag{3.3.59}$$

$$g_Z^{(k)} := Z^{(k)T}\nabla f(x^{(k)}) \tag{3.3.60}$$

door oplossing van het herschreven stelsel (3.3.57)

$$B_Z^{(k)}d_Z^{(k)} = -g_Z^{(k)} \tag{3.3.61}$$

en invulling van de oplossing daarvan in (3.3.58)

$$d^{(k)} := Z^{(k)}d_Z^{(k)}$$

Het bepalen van de zoekrichting $d_Z^{(k)}$ met behulp van (3.3.61) verloopt geheel analoog als betrof het een toepassing van de in het voorgaande hoofdstuk, pt. 2.9.21 t/m 2.9.23 besproken quasi-Newton-algorithme van Gill en Murray voor onbeperkte minimalisering in $\mathbb{R}^{n-q}$. Voor de oplossing van het stelsel (3.3.61) suggereren Gill en Murray het gebruik van een Choleski-decompositie van de matrix $B_Z^{(k)}$

$$B_Z^{(k)} = L_Z^{(k)}D_Z^{(k)}L_Z^{(k)T} \tag{3.3.62}$$

waarna (3.3.61) opgelost kan worden door successievelijke oplossing van de driehoeksstelsels (vgl. (3.3.33))

$$\begin{aligned} L_Z^{(k)}v &= -g_Z^{(k)} \\ L_Z^{(k)T}d_Z^{(k)} &= D_Z^{(k)-1}v \end{aligned} \tag{3.3.63}$$

3.3.15. Juist als bij de in het voorgaande besproken methoden moeten bij problemen met lineaire nevenvoorwaarden drie verschillende gevallen m.b.t. veranderingen in de verzameling van actieve (lineaire) beperkingen worden onderscheiden die gevolgen hebben voor de matrices $B^{(k)}$ en $Z^{(k)}$ (en de daarvan afgeleide matrices $B_Z^{(k)}$, $L_Z^{(k)}$ en $D_Z^{(k)}$). Deze drie gevallen betreffen de volgende situaties (vgl. pt. 3.3.12):

(i) de verzameling van actieve (lineaire) beperkingen is dezelfde aan het einde van de k-de stap als aan het begin van de k-de stap
(ii) de verzameling van actieve (lineaire) beperkingen moet na de k-de stap worden uitgebreid van q tot q+1 beperkingen
(iii) de verzameling van actieve(lineaire)beperkingen moet aan het begin van de k-de stap worden ingekrompen van q tot q-1 beperkingen

Voor ieder van deze drie gevallen werden door Gill en Murray speciale aanpassingsformules gesuggereerd in [3.3.4] en [3.3.5] en wel als volgt:

(i) In het geval dat de matrix van normalen van actieve beperkingen dezelfde is in $x^{(k)}$ als in $x^{(k+1)}$ dan ondergaat de matrix $Z^{(k)}$ geen verandering en geldt

$$Z^{(k+1)} := Z^{(k)} \tag{3.3.64}$$

De in de k-de stap verworven informatie over de verandering van de gradiënt kan in dit geval worden gebruikt ter verbetering van de benadering van de Hessiaan $B^{(k)}$ (en de benadering van de geprojecteerde Hessiaan $B_Z^{(k)}$) op dezelfde manier als besproken in pt. 2.8.14 voor het geval van onbeperkte minimalisering. In principe kunnen daarom alle duale formuleringen van de bekende quasi-Newton aanpassingsformules worden toegepast. Gill en Murray vonden langs experimentele weg (vgl. [3.3.5]) dat daarbij de voorkeur dient te worden gegeven aan de complementaire DFP-formule (2.8.23)

$$B^{(k+1)} := B^{(k)} + \frac{y^{(k)}{}_y{}^{(k)T}}{y^{(k)T}{}_s{}^{(k)}} - \frac{B^{(k)}{}_s{}^{(k)}{}_s{}^{(k)T}{}_B{}^{(k)}}{s^{(k)T}{}_B{}^{(k)}{}_s{}^{(k)}} \tag{3.3.65}$$

die met de relatie

$$B^{(k)}s^{(k)} = H^{(k)-1}(-\alpha^{(k)}H^{(k)}g^{(k)}) = -\alpha^{(k)}g^{(k)} \qquad (3.3.66)$$

te schrijven is (zonder matrix-vector producten) als

$$B^{(k+1)} := B^{(k)} + \frac{y^{(k)}y^{(k)T}}{y^{(k)T}s^{(k)}} + \alpha^{(k)}\frac{g^{(k)}g^{(k)T}}{g^{(k)T}s^{(k)}} \qquad (3.3.67)$$

Voor vermenigvuldiging met $Z^{(k+1)T} = Z^{(k)T}$ en navermenigvuldiging met $Z^{(k+1)} = Z^{(k)}$ en gebruikmaking van de voor de hand liggende definities

$$y_Z^{(k)} := Z^{(k)T}y^{(k)} = Z^{(k)T}(g^{(k+1)} - g^{(k)}) = g_Z^{(k+1)} - g_Z^{(k)} \qquad (3.3.68)$$

$$Z^{(k)}s_Z^{(k)} := s^{(k)} \qquad (3.3.69)$$

geeft onmiddellijk de aanpassingsformule

$$B_Z^{(k+1)} := B_Z^{(k)} + \frac{y_Z^{(k)}y_Z^{(k)T}}{y_Z^{(k)T}s_Z^{(k)}} + \alpha^{(k)}\frac{g_Z^{(k)}g_Z^{(k)T}}{g_Z^{(k)}s_Z^{(k)}} \qquad (3.3.70)$$

of equivalent de formule

$$B_Z^{(k+1)} := B_Z^{(k)} + \frac{y_Z^{(k)}y_Z^{(k)T}}{y_Z^{(k)T}s_Z^{(k)}} - \frac{B_Z^{(k)}s_Z^{(k)}s_Z^{(k)T}B_Z^{(k)}}{s_Z^{(k)T}B_Z^{(k)}s_Z^{(k)}} \qquad (3.3.71)$$

Deze laatste formules illustreren duidelijk hoe de aanpassing van de matrix $B_Z^{(k)}$ direct kan worden uitgevoerd als betrof het een aanpassing in een ruimte $\mathbb{R}^{n-q}$. In verband met de speciale vorm van de aanpassingsformule (3.3.70) is het mogelijk (vgl. pt. 2.9.22) deze op te vatten als het resultaat van twee successievelijke rang-één-correcties die het mogelijk maken direct de Choleski-factoren van de matrix B_Z aan te passen. Met deze door Gill en Murray in meer detail [3.3.5] beschreven aanpassing kan een grotere numerieke stabiliteit van het rekenproces worden bereikt.

(ii) In het geval dat aan het einde van de k-de stap een eerdere passieve beperking actief wordt neemt de dimensie van de matrix N toe van q naar q + 1 en moet de dimensie van de matrix Z afnemen van n - q naar n - q - 1. De nieuwe matrix $Z^{(k+1)}$ kan worden bepaald door opnieuw een QR-decompositie uit te voeren dan wel, zoals besproken in pt. 3.2.27, door gebruik te maken van de informatie aanwezig in de oude QR-decompositie. Nodig in dat geval is bepaling van een orthogonale (n - q) × (n - q)-matrix $\hat{Q}^{(k)}$ zo dat (vgl. (3.2.136))

$$Z^{(k+1)} = Z^{(k)}\hat{Q}^{(k)T}\begin{bmatrix} 0 \\ \hline I_{n-q-1} \end{bmatrix}$$

waar (vgl.pt. 3.2.27) $\hat{Q}$ een Householder matrix is

$$\hat{Q}^{(k)} = I - 2\frac{uu^T}{u^Tu}$$

waarin

$$u = r_2 - \|r_2\| e_1$$

en (vgl. (3.2.124))

$$r_2 = Q_2^{(k)T} n_{q+1}$$

Gegeven de matrix $\hat{Q}^{(k)}$ is het mogelijk om direct de Choleski-factoren te bepalen van de nieuwe matrix $\bar{B}_Z^{(k)}$ met behulp van de bekende Choleski-factoren van de matrix $B_Z^{(k)}$

$$\begin{aligned} \bar{B}_Z^{(k)} &= Z^{(k+1)T} B^{(k)} Z^{(k+1)} \\ &= \left[0 \;\vdots\; I_{n-q-1}\right] \hat{Q}^{(k)} Z^{(k)T} B^{(k)} Z^{(k)} \hat{Q}^{(k)T} \begin{bmatrix} 0 \\ \hline I_{n-q-1} \end{bmatrix} \qquad (3.3.72) \\ &= \left[0 \;\vdots\; I_{n-q-1}\right] \hat{Q}^{(k)} L_Z^{(k)} D_Z^{(k)} L_Z^{(k)T} \tilde{Q}^{(k)T} \begin{bmatrix} 0 \\ \hline I_{n-q-1} \end{bmatrix} \end{aligned}$$

Voor verdere details van de niet-triviale uitwerking hiervan moet worden verwezen naar [3.3.5] of [3.3.9].
Anders dan Goldfarb suggereren Gill en Murray in [3.3.5] ook nog om

de in deze stap (die eindigde op een nieuwe beperking) verkregen informatie te verwerken in een nieuwe benadering voor de geprojecteerde Hessiaan $B_Z^{(k+1)}$. Hiertoe kan gebruik gemaakt worden van de volgende aangepaste versie van de complementaire DFP-formule waarin de matrix $\bar{B}_Z^{(k)}$ de plaats inneemt van de matrix $B_Z^{(k)}$

$$B_Z^{(k+1)} := \bar{B}_Z^{(k)} + \frac{\bar{y}_Z^{(k)} \bar{y}_Z^{(k)T}}{\bar{y}_Z^{(k)T} \bar{s}_Z^{(k)}} + \alpha^{(k)} \frac{\bar{g}_Z^{(k)} \bar{g}_Z^{(k)T}}{\bar{g}_Z^{(k)T} \bar{s}_Z^{(k)}} \tag{3.3.74}$$

waarin

$$\bar{y}_Z^{(k)} := Z^{(k+1)T} y^{(k)} = \left[0 \mid I_{n-q-1}\right] \hat{Q}^{(k)} y_Z^{(k)} \tag{3.3.75}$$

$$\bar{s}_Z^{(k)} := Z^{(k+1)T} s^{(k)} = \left[0 \mid I_{n-q-1}\right] \hat{Q}^{(k)} s_Z^{(k)} \tag{3.3.76}$$

$$\bar{g}_Z^{(k)} := Z^{(k+1)T} g^{(k)} = \left[0 \mid I_{n-q-1}\right] \hat{Q}^{(k)} g_Z^{(k)} \tag{3.3.77}$$

(iii) In het geval dat aan het begin van de k-de stap een tot dusver actief verondersteld beperking passief wordt verondersteld neemt het aantal kolommen van de matrix $N^{(k)}$ af van q naar q - 1 en moet het aantal kolommen van de corresponderende matrix $Z^{(k)}$ toenemen van n - q naar n - q + 1. Dit laatste kan worden gerealiseerd, zoals besproken in pt. 3.2.29, door toevoeging van een kolom $\tilde{z}$ aan de matrix $Z^{(k)}$, zodat (vgl. (3.2.137))

$$Z^{(k+1)} := [\tilde{z} \mid Z^{(k)}] \tag{3.3.78}$$

Deze vector $\tilde{z}$ moet voldoen aan zowel (vgl. (3.2.138))

$$Z^{(k)T} \tilde{z} = 0$$

als

$$N_{q-1}^{(k)T} \tilde{z} = 0$$

waar, als n_s de nieuwe passieve normaal is, $N_{q-1}^{(k)}$ gelijk is aan de matrix van normalen $N^{(k)}$ met daaruit de normaal n_s verwijderd (vgl. (3.2.130))

$$N^{(k+1)} := N^{(k)}_{q-1} := [n_1, n_2 \ldots n_{s-1}, n_{s+1}, \ldots n_q] \tag{3.3.79}$$

De vector $\tilde{z}$ die juist gelijk blijkt te zijn aan de genormaliseerde s-de rij n_s^+ van de gegeneraliseerde of pseudo-inverse $N^{(k)+}$ van de matrix $N^{(k)}$, d.i. (vgl. (3.2.139))

$$\tilde{z} = n_s^+ \; / \; \| n_s^+ \|$$

kan als zodanig o.a. worden bepaald als oplossing van de vergelijkingen (3.2.143) en (3.2.144)

$$R^{(k)T} R^{(k)} v = e_s$$

en

$$n_s^+ = N^{(k)} v$$

waarin $R^{(k)}$ de bovendriehoeksmatrix is van de QR-decompositie van $N^{(k)}$ en e_s de s-de kolom van de eenheidsmatrix I_q

Het simpele verband (3.3.78) tussen de nieuwe matrix $Z^{(k+1)}$ en de oude matrix $Z^{(k)}$ maakt het mogelijk op eenvoudige manier een nieuwe benadering te construeren voor de geprojecteerde Hessiaan

$$B_Z^{(k+1)} := \bar{B}_Z^{(k)} := Z^{(k+1)T} B^{(k)} Z^{(k+1)} \tag{3.3.80}$$

Met behulp van een enkele gradiënt-evaluatie langs de lijn $x(\alpha) := x^{(k)} + \alpha\tilde{z}$ is het namelijk mogelijk om een benadering r te bepalen voor de vector $B^{(k)}\tilde{z}$

$$r := (\nabla f(x^{(k)} + \tau\tilde{z}) - \nabla f(x^{(k)})) \; / \tau \tag{3.3.81}$$

waarmee

$$\bar{B}_Z^{(k)} := \left[\begin{array}{c|c} \tilde{z}^T r & r^T Z^{(k)} \\ \hline Z^{(k)T} r & B_Z^{(k)} \end{array}\right] \tag{3.3.82}$$

Men kan deze extra gradiënt-evaluatie ook achterwege laten en in plaats daarvan $Z^{(k)T}B^{(k)}\tilde{z} = 0$ en $\tilde{z}^{T}B^{(k)}\tilde{z} = 1$ stellen, waarmee de nieuwe benadering $\bar{B}_Z^{(k)}$ voor de geprojecteerde Hessiaan gelijk wordt aan

$$\bar{B}_Z^{(k)} = \left[\begin{array}{c|c} 1 & 0 \\ \hline 0 & B_Z^{(k)} \end{array}\right] \tag{3.3.83}$$

Dit laatste komt overeen met de gebruikelijke strategie om de Hessiaan bij gebrek aan betere informatie gelijk te stellen aan de eenheidsmatrix. De nieuwe zoekrichting $\bar{d}_Z^{(k)}$ wordt nu bepaald als oplossing van de vergelijking

$$\left[\begin{array}{c|c} 1 & 0 \\ \hline 0 & B_Z^{(k)} \end{array}\right] \bar{d}_Z = -\left[\begin{array}{c} \tilde{z}^{T}g^{(k)} \\ \\ Z^{(k)T}g^{(k)} \end{array}\right] \approx -\left[\begin{array}{c|c} \tilde{z}^{T}g^{(k)} \\ \hline 0 \end{array}\right] \tag{3.3.84}$$

met als resultaat dat de nieuwe zoekrichting

$$\bar{d}^{(k)} := Z^{(k+1)}\bar{d}_Z = [\tilde{z} \mid Z^{(k)}]\bar{d}_Z \approx -(\tilde{z}^{T}g^{(k)})\tilde{z} \tag{3.3.85}$$

bij benadering gelijk zal zijn aan de vector $\tilde{z}$ d.i. juist gelijk aan de richting waarin (tweede-orde) informatie over de verandering van de gradiënt gewenst is. Ongeacht welke van de twee benaderingen voor de matrix $\bar{B}_Z^{(k)}$ wordt gekozen, het is in beide gevallen relatief eenvoudig om de Choleski-factoren van de nieuwe matrix $\bar{B}_Z^{(k)}$ te bepalen door aanpassing van de Choleski-factoren van de voorgaande matrix $B_Z^{(k)}$. Voor meer details daarover wordt weer verwezen naar [3.3.5].

Geconjugeerde-gradiënt en quasi-Newton algorithmen in $\mathbb{R}^{n-q}$

3.3.16. Zoals reeds eerder opgemerkt (pt. 3.2.11) kan, zolang de verzameling van actieve beperkingen niet verandert, ieder minimaliseringsprobleem met q lineaire nevenvoorwaarden worden opgevat als een onbeperkt minimaliseringsprobleem in $\mathbb{R}^{n-q}$

$$\min\{\varphi^{(\ell)}(w) \mid \varphi^{(\ell)}(w) := f(x^{(\ell)} + Z^{(\ell)}w) \ , \ w \in \mathbb{R}^{n-q}\} \tag{3.3.86}$$

waarin $\varphi^{(\ell)}(w)$ de restrictie voorstelt van de functie f(x) tot de lineaire variëteit

$$T(x^{(\ell)}) := \{x \in \mathbb{R}^n \mid x := x^{(\ell)} + Z^{(\ell)}w,\ w \in \mathbb{R}^{n-q}\} \tag{3.3.87}$$

In deze uitdrukkingen is $x^{(\ell)}$ een steunpunt en $Z^{(\ell)}$ de matrix met (als kolommen de niet noodzakelijke orthonormale) basisvectoren van het orthogonale complement van de ruimte opgespannen door de normalen van de actieve beperkingen in $x^{(\ell)}$. Met ieder punt $w^{(\ell,k)} \in \mathbb{R}^{n-q}$ correspondeert een punt $x^{(k)} \in \mathbb{R}^n$ overeenkomstig de relatie

$$x^{(k)} := x^{(\ell)} + Z^{(\ell)}w^{(\ell,k)} \tag{3.3.89}$$

Verandert in een punt $x^{(\bar{\ell})} \in \underline{T}(x^{(\ell)})$ de verzameling van actieve beperkingen dan wordt het punt $x^{(\bar{\ell})}$ als nieuw steunpunt gekozen voor een nieuwe lineaire variëteit $T(x^{(\bar{\ell})})$. De aanwezige informatie over de restrictie $\varphi^{(\ell)}(\underline{w})$ kan overeenkomstig worden aangepast voor de nieuwe restrictie $\varphi^{(\bar{\ell})}(w)$.

3.3.17. Voor het minimaliseren van de functie $\varphi^{(\ell)}(w)$ (3.3.86) kunnen alle in het voorgaande hoofdstuk besproken algorithmen voor onbeperkte minimalisering worden toegepast. Naast de gradiënt methode (vgl. § 2.4 en § 3.2) en de methode van Newton (§ 2.5 en deze paragraaf) komen in het bijzonder ook alle geconjugeerde-richtingen algorithmen (§ 2.6) en quasi-Newton methoden (§ 2.7 - § 2.9) daarvoor in aanmerking. Nodig voor de toepassing van deze methoden is de kennis van de gradiënt van de te minimaliseren functie in ieder iteratie punt $w^{(\ell,k)}$

$$\nabla_w \varphi(w^{(\ell,k)}) = Z^{(\ell)T}\nabla f(x^{(\ell)} + Z^{(\ell)}w^{(\ell,k)}) = Z^{(k)T}\nabla f(x^{(k)}) = g_Z^{(k)} \tag{3.3.90}$$

en eventueel de Hessiaan

$$\begin{aligned}\nabla_w^2 \varphi(w^{(\ell,k)}) &= Z^{(\ell)T}G(x^{(\ell)} + Z^{(\ell)}w^{(\ell,k)})Z^{(\ell)} \\ &= Z^{(k)T}G(x^{(k)})Z^{(k)} = G_Z^{(k)}\end{aligned} \tag{3.3.91}$$

Alle overige in de theorie van deze algorithmen voorkomende grootheden kunnen (voorzien van een index Z) op analoge wijze worden gedefinieerd in $\mathbb{R}^{n-q}$. Bijvoorbeeld met $Z^{(k+1)} = Z^{(k)} = Z^{(\ell)}$ geldt (vgl. (3.3.68))

$$y_Z^{(k)} := g_Z^{(k+1)} - g_Z^{(k)} = \nabla_w \varphi^{(\ell)}(w^{(\ell,k+1)}) - \nabla_w \varphi^{(\ell)}(w^{(\ell,k)})$$

$$= Z^{(\ell)T}(g^{(k+1)} - g^{(k)}) = Z^{(k)T} y^{(k)} \qquad (3.3.92)$$

en (3.3.69)

$$s_Z^{(k)} := w^{(\ell,k+1)} - w^{(\ell,k)} = (Z^{(\ell)T} Z^{(\ell)})^{-1} Z^{(\ell)T}(x^{(k+1)} - x^{(k)})$$

$$= (Z^{(k)T} Z^{(k)})^{-1} Z^{(k)T} s^{(k)} \qquad (3.3.93)$$

Wordt met een van de onbeperkte minimaliserings-algorithmen een zoekrichting $d_Z^{(k)}$ gegenereerd dan volgt daaruit onmiddellijk een zoekrichting $d^{(k)} \in \mathbb{R}^n$ met behulp van de transformatie

$$d^{(k)} := Z^{(k)} d_Z^{(k)} \qquad (3.3.94)$$

Een tweetal voorbeelden (vgl. [3.3.6]) van het gebruik van algorithmen in $\mathbb{R}^{n-q}$ voor het genereren van zoekrichtingen voor minimaliseringsproblemen met lineaire nevenvoorwaarden zal hieronder in het kort worden besproken.

3.3.18. Een van de meest simpele en daarom meest toegepaste geconjugeerde-gradiënt algorithmen is de algorithme van Fletcher en Reeves (pt. 2.6.18) met als zoekrichting (2.6.32)

$$d^{(0)} := -g^{(0)} \qquad k = 0$$

$$d^{(k)} := -g^{(k)} + \frac{g^{(k)T} g^{(k)}}{g^{(k-1)T} g^{(k-1)}} s^{(k-1)} \qquad k = 1,2,3,\ldots$$

Voor de minimalisering van de functie $\varphi^{(\ell)}(w)$ uitgaande van het *start*-punt $x^{(\ell)}$ in $T(x^{(\ell)})$ geeft dat als zoekrichtingen in $\mathbb{R}^{n-q}$

$$d_Z^{(k)} := -g_Z^{(\ell)} = -g_Z^{(k)} \qquad k = \ell.$$

$$d_Z^{(k)} := -g_Z^{(k)} + \frac{g_Z^{(k)T} g_Z^{(k)}}{g_Z^{(k-1)T} g_Z^{(k-1)}} s_Z^{(k-1)} \qquad k = \ell+1, \ell+2, \ldots \tag{3.3.95}$$

en, equivalent met $Z^{(k)} = Z^{(k-1)} = Z^{(\ell)}$ als zoekrichtingen in $\mathbb{R}^n$

$$d^{(k)} := -Z^{(\ell)} Z^{(\ell)T} g^{(\ell)} = -Z^{(k)} Z^{(k)T} g^{(k)} \qquad k = \ell.$$

$$d^{(k)} := -Z^{(k)} Z^{(k)T} g^{(k)} + \frac{g^{(k)T} Z^{(k)} Z^{(k)T} g^{(k)}}{g^{(k-1)T} Z^{(k-1)} Z^{(k-1)T} g^{(k-1)}} s^{(k-1)}$$

$$k = \ell+1, \ell+2, \ldots \tag{3.3.96}$$

In het geval dat de kolommen van $Z^{(\ell)}$ orthonormaal zijn geldt dat $d^{(\ell)}$ gelijk is aan de (negatieve) geprojecteerde gradiënt (vgl. (3.2.25))

$$d^{(\ell)} := -\bar{P}^{(\ell)} g^{(\ell)} := -Z^{(\ell)} Z^{(\ell)T} g^{(\ell)} \tag{3.3.97}$$

en dat overeenkomstig voor $d^{(k)}$ $(k>\ell)$, geschreven kan worden

$$d^{(k)} := -\bar{P}^{(k)} g^{(k)} + \frac{\|\bar{P}^{(k)} g^{(k)}\|^2}{\|\bar{P}^{(k-1)} g^{(k-1)}\|^2} s^{(k-1)} \tag{3.3.98}$$

In het geval dat (vgl. (3.2.54))

$$Z^{(\ell)} = M^{(\ell)} = \begin{bmatrix} -B^{-1}D \\ I \end{bmatrix}^{(\ell)}$$

dan geldt dat $d^{(\ell)}$ de principiële zoekrichting is van de gereduceerde-gradiënt methode (vgl. (3.2.59))

$$d^{(\ell)} := -\bar{R}^{(\ell)} g^{(\ell)} := -M^{(\ell)} M^{(\ell)T} g^{(\ell)} \tag{3.3.99}$$

en dat voor $d^{(k)}$ geschreven kan worden als $k > \ell$

$$d^{(k)} := -\bar{R}^{(k)} g^{(k)} + \frac{g^{(k)T}(\bar{R}^{(k)} g^{(k)})}{g^{(k-1)T}(\bar{R}^{(k-1)} g^{(k-1)})} s^{(k-1)} \qquad (3.3.100)$$

3.3.19. De meest bekende quasi-Newton algorithme is zonder twijfel de algorithme van Davidon-Fletcher-Powell, d.i. de standaard quasi-Newton algorithme (pt. 2.7.2) met de DFP-aanpassingsformule (2.8.2). Hierbij wordt de zoekrichting gegenereerd met behulp van de formule (2.7.2)

$$d^{(k)} := -H^{(k)} g^{(k)}$$

waarbij de matrix $H^{(k)}$ een benadering is voor de inverse $G^{(k)-1}$ van de Hessiaan, welke benadering aanvankelijk gelijk is aan een positief definiete startmatrix

$$H^{(0)} := H_0 (= I)$$

en vervolgens in iedere iteratiestap wordt aangepast met de (DFP-) aanpassingsformule (2.8.2)

$$H^{(k+1)} := H^{(k)} + \frac{s^{(k)} s^{(k)T}}{s^{(k)T} y^{(k)}} - \frac{H^{(k)} y^{(k)} y^{(k)T} H^{(k)}}{y^{(k)T} H^{(k)} y^{(k)}}$$

Toepassing van DFP-algorithme voor de minimalisering van de functie $\varphi^{(\ell)}(w)$ geeft als zoekrichting in $\mathbb{R}^{n-q}$

$$d_Z^{(k)} := -H_Z^{(k)} g_Z^{(k)} \qquad (3.3.101)$$

waar $H_Z^{(k)}$ een benadering voorstelt van de inverse van de Hessiaan $G_Z^{(k)-1}$

$$H_Z^{(k)} := B_Z^{(k)-1} = (Z^{(k)T} B^{(k)} Z^{(k)})^{-1} \qquad (3.3.102)$$

De aanpassingsformule in $\mathbb{R}^{n-q}$ krijgt de vorm

$$H_Z^{(\ell)} = H_{Z,\ell} (= I_{n-q})$$

en

$$H_Z^{(k+1)} := H_Z^{(k)} + \frac{s_Z^{(k)} s_Z^{(k)T}}{s_Z^{(k)T} y_Z^{(k)}} - \frac{H_Z^{(k)} y_Z^{(k)} y_Z^{(k)T} H_Z^{(k)}}{y_Z^{(k)T} H_Z^{(k)} y_Z^{(k)}} \qquad k = \ell, \ell+1, \ldots$$

(3.3.103)

De overeenkomstige zoekrichting in $\mathbb{R}^n$ wordt gegeven door

$$d^{(k)} := -Z^{(k)}H_Z^{(k)}Z^{(k)T}g^{(k)} = -\bar{H}^{(k)}g^{(k)} \qquad (3.3.104)$$

waar

$$\bar{H}^{(k)} = Z^{(k)}H_Z^{(k)}Z^{(k)T} \qquad (3.3.105)$$

Voor en na vermenigvuldiging van de aanpassingsformule voor de matrix $H_Z^{(k)}$ met respectievelijk $Z^{(k)}$ en $Z^{(k)T}$ en gebruikmaking van de definities (3.3.92) en (3.3.93) voor $y_Z^{(k)}$ en $s_Z^{(k)}$ leert dat deze laatste matrix $\bar{H}^{(k)}$ ook direct verkregen had kunnen worden met de aanpassingsformules

$$\bar{H}^{(\ell)} := Z^{(\ell)}H_{Z,\ell}Z^{(\ell)T} (= Z^{(\ell)}Z^{(\ell)T})$$

en

$$\bar{H}^{(k+1)} := \bar{H}^{(k)} + \frac{s^{(k)}s^{(k)T}}{s^{(k)T}y^{(k)}} - \frac{\bar{H}^{(k)}y^{(k)}y^{(k)T}\bar{H}^{(k)}}{y^{(k)T}\bar{H}^{(k)}y^{(k)}} \quad k = \ell,\ell+1,\ldots \qquad (3.3.106)$$

Deze laatste formule is juist gelijk aan de aanpassingsformule van Goldfarb (3.3.38) voor de matrix $K_q^{(k)}$ in het geval dat geen verandering in de verzameling van actieve beperkingen optreedt. In theorie zijn de methode van Goldfarb en de hier geschetste methode op dit punt dus equivalent. In de practijk verdient de hier geschetste methode de voorkeur omdat een kleinere matrix $H_Z^{(k)}$ i.p.v. $\bar{H}^{(k)}$ moet worden aangepast en omdat de aanpassing geen effect heeft op de projectieeigenschappen van de matrix $\bar{H}^{(k)}$.

3.3.20. Bij de hiervoor besproken toepassing van de geconjugeerde-gradiënt methode van Fletcher en Reeves (pt. 3.3.18), kan de formule voor het genereren van de zoekrichting niet zinvol worden gebruikt bij een verandering van de verzameling van actieve beperkingen, omdat de in de formule expliciet voorkomende voorgaande stap $s_Z^{(k-1)}$ betrekking heeft op een andere lineaire variëteit, dan waarover in de komende stap moet worden geminimaliseerd. Bij de toepassing van quasi-Newton

methoden ligt dat anders en is het beperkt mogelijk om bij een verandering van de verzameling van actieve beperkingen de in de matrix $H_Z^{(k)}$ verwerkte informatie over de inverse van de geprojecteerde Hessiaan van de objectfunctie (vgl. (3.3.102)) te gebruiken voor een nieuwe benaderingsmatrix $H_Z^{(k+1)}$ die betrekking heeft op de nieuwe lineaire variëteit bepaald door de actieve beperkingen in $x^{(k+1)}$. In het geval het aantal actieve beperkingen groter wordt, wordt de dimensie van de matrix $H_Z^{(k+1)}$ overeenkomstig kleiner en kan voor de bepaling van de matrix $H_Z^{(k+1)}$ in principe gebruik gemaakt worden van de stelling voor de inverse van "omrande" matrices (Stelling 3.2.24). In het geval het aantal actieve beperkingen afneemt, moet de dimensie van de matrix $H_Z^{(k+1)}$ overeenkomstig groeien. Bij gebrek aan informatie over de (inverse van de geprojecteerde) Hessiaan in de nieuwe toegelaten richting moet voor de uitbreiding van de matrix $H_Z^{(k+1)}$ gebruik gemaakt worden van ruwe schattingen. In beide gevallen, uitbreiding en inkrimping van de verzameling van actieve beperkingen, zijn de resulterende aanpassingsformules voor de matrix $H_Z^{(k+1)}$ in principe equivalent met de corresponderende aanpassingsformules van Goldfarb ((3.3.39) en (3.3.40)).

3.3.21. Van belang voor de convergentiesnelheid in de practijk van zowel geconjugeerde richtingen als quasi-Newton methoden is de mate waarin het lijnminimum benaderd wordt bij de bepaling van de stapgrootte. Indien bij problemen met nevenvoorwaarden de stapgrootte bepaald wordt door het actief worden van een nieuwe beperking is de relatie met het lijnminimum in het algemeen afwezig en is het de vraag wat de waarde is van het aan passen van H_Z-matrix met de informatie opgedaan in die stap (Goldfarb verkoos in dit geval daarom ook alleen de projectieeigenschappen van zijn matrix K_q aan te passen). De voorkeur indien toch aangepast wordt verdienen dan die aanpassingsformules waarvan de convergentie in de practijk minder van de nauwkeurigheid van de lijnminimalisering afhankelijk is. Dat betreft dan in de eerste plaats de rang-één-aanpassingsformule (2.8.1) waarvan in theorie de convergentieeigenschappen onafhankelijk zijn van het al dan niet lijnminimaliseren

$$H_Z^{(k+1)} := H_Z^{(k)} + \frac{(s_Z^{(k)} - H_Z^{(k)} y_Z^{(k)})(s_Z^{(k)} - H_Z^{(k)} y_Z^{(k)})^T}{(s_Z^{(k)} - H_Z^{(k)} y_Z^{(k)})^T y_Z^{(k)}} \qquad (3.3.107)$$

In de tweede plaats zou gebruik gemaakt kunnen worden van de aanpassingsformule van de zelf-schalende variabele metriek algorithmen van

Oren en Luenberger (pt. 2.9.16)

$$H_Z^{(k+1)} := (H_Z^{(k)} - \frac{H_Z^{(k)} y_Z^{(k)} y_Z^{(k)T} H_Z^{(k)}}{y_Z^{(k)T} H_Z^{(k)} y_Z^{(k)}} + \phi^{(k)} v_Z^{(k)} v_Z^{(k)T}) \gamma^{(k)} + \frac{s_Z^{(k)} s_Z^{(k)T}}{s_Z^{(k)T} y_Z^{(k)}}$$

waarin (3.3.108)

$$v_Z^{(k)} := (y_Z^{(k)T} H_Z^{(k)} y_Z^{(k)})^{\frac{1}{2}} \left(\frac{s_Z^{(k)}}{s_Z^{(k)T} y_Z^{(k)}} - \frac{H_Z^{(k)} y_Z^{(k)}}{y_Z^{(k)T} H_Z^{(k)} y_Z^{(k)}} \right)$$

en waarin $\phi^{(k)} \in [0,1]$ en $\gamma^{(k)} \in [0,1]$ in overeenstemming met de door Oren gesuggereerde regels te kiezen (vgl. pt. 2.9.19) parameters zijn. Eenvoudig kan worden aangetoond dat voorvermenigvuldiging met $Z^{(k+1)} = Z^{(k)}$ en navermenigvuldiging met $Z^{(k)T}$ resulteert in gelijkvormige en equivalente aanpassingsformules in $\mathbb{R}^n$ voor de met de matrices $K_q^{(k)}$ van Goldfarb vergelijkbare matrices $\bar{H}^{(k)}$.

Niet-lineaire beperkingen: Restauratie procedures

3.3.22. In het geval dat alle beperkingen lineair zijn en gebruik wordt gemaakt van de in het voorgaande besproken zoekrichtingen en stapgroottebegrenzingen (pt. 3.2.21) dan zullen in theorie (d.i. indien wordt afgezien van eventuele numerieke onnauwkeurigheden) als het startpunt toegelaten is ook alle volgende iteratiepunten toegelaten zijn. Bij aanwezigheid van niet-lineaire beperkingen verandert die situatie omdat de actieve beperkingen niet noodzakelijk actief blijven en omdat de met linearisaties berekende stapgroottebegrenzingen passieve beperkingen niet noodzakelijk actief maken. Het voorlopige eindresultaat van iedere iteratiestap

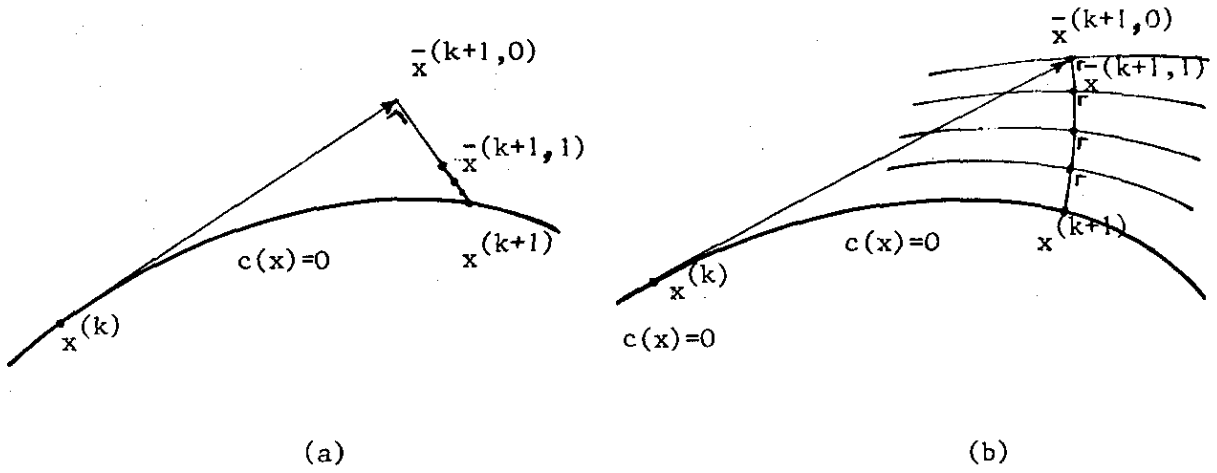
$$\bar{x}^{(k+1,0)} = x^{(k)} + \alpha^{(k)} d^{(k)} \qquad (3.3.109)$$

zal in het algemeen dan ook een niet-toegelaten punt zijn en speciale procedures zijn nodig om uitgaande van het niet-toegelaten punt (3.3.109) een toegelaten punt te bepalen. Zoals gezegd in pt. 3.3.1 wordt zo'n procedure aangeduid als restauratieprocedure. Men kan bij

het bepalen van een toegelaten punt vanuit het laatst bereikte niet-toegelaten punt uitgaan van twee essentieel verschillende ideeën over het verkrijgen van de benodigde informatie in het niet-toegelaten punt:

a) Men gaat uit van de informatie, zoals deze bepaald is in het laatst bepaalde toegelaten punt:
b) Men doet voor elke restauratiestap alsof het niet-toegelaten punt een eerste niet-toegelaten punt is en bepaalt in dat punt alle nodige informatie.

Ook allerlei tussenvormen zijn mogelijk.Bijvoorbeeld door een deel van de oude informatie te gebruiken en de rest opnieuw te bepalen óf door alleen in het eerste niet-toegelaten punt in de betreffende iteratie-slag alle informatie opnieuw te bepalen en deze in eventuele volgende restauratiestappen te gebruiken. De mogelijkheden genoemd onder a en b worden geïllustreerd in Figuur 3.3.22.



Figuur 3.3.22 : Twee mogelijkheden om vanuit een niet-toegelaten punt $\bar{x}^{(k+1,0)}$ een toegelaten punt $x^{(k+1)}$ te vinden.

3.3.23. In het geval men alleen gebruik wil maken van de informatie die in het laatste toegelaten punt bekend is, maakt het verschil of de zoekrichting $d^{(k)}$ bepaald is met de gereduceerde-gradiëntmethode of met de geprojekteerde-gradiëntmethode. Wanneer $d^{(k)}$ bepaald is met de gereduceerde-gradiëntmethode dan kan men de niet-basiscomponenten aanpassen aan de niet-lineaire beperkingen. Dat wil zeggen dat de restauratie alleen via de basisvariabelen wordt uitgevoerd. Als geldt dat tijdens het restauratieproces geen beperkingen overschreden worden, behalve diegene die in $x^{(k)}$ actief waren, kan de restauratiestap $s^{(k+1,r)}$ met een Newton-achtige methode worden berekend uit (vgl. (3.2.84))

$$B^{(k)} s_B^{(k+1,r)} = -c(\bar{x}^{(k+1,r)})$$

$$s_D^{(k+1,r)} = 0$$

waar $s^{(k+1,r)} = \Delta\bar{x}^{(k+1,r)}$ de (r+1)-de restauratiestap is, $\bar{x}^{(k+1,r)}$ de r-de benadering is voor het toegelaten punt $x^{(k+1)}$, en $B^{(k)}$ de in het toegelaten punt $x^{(k)}$ bepaalde basismatrix. Als wel nieuwe beperkingen overschreden worden, heeft het weinig nut vast te houden aan de oude splitsing in basisvariabelen en niet-basisvariabelen en kan de restauratie evengoed worden uitgevoerd via alle variabelen.

Als $d^{(k)}$ bepaald is met de geprojecteerde-gradiëntmethode kan men een toegelaten punt zoeken door een lineaire combinatie te bepalen van de normalen van de actieve en overschreden beperkingen, berekend in het laatst bepaalde toegelaten punt $x^{(k)}$. In dat geval kan de restauratiestap worden bepaald uit het Newton-achtige stelsel (vgl. pt. 3.2.7)

$$\bar{N}^{(k)T} s^{(k+1,r)} = \bar{N}^{(k)T} \bar{N}^{(k)} \kappa^{(k+1,r)} = -c(\bar{x}^{(k+1,r)}) \qquad (3.3.110)$$

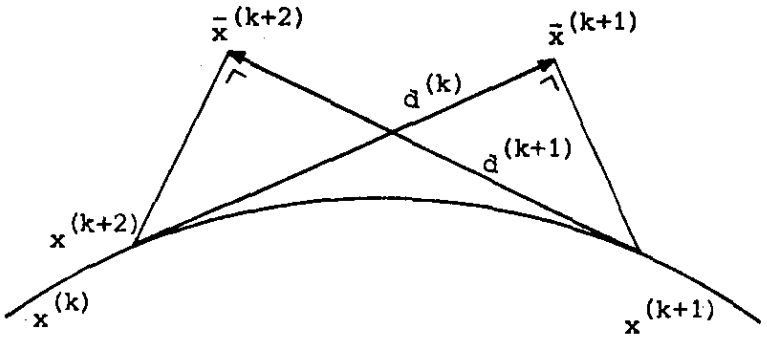
waaruit volgt (vgl. (3.2.12))

$$s^{(k+1,r)} = -\bar{N}^{(k)} (\bar{N}^{(k)T} \bar{N}^{(k)})^{-1} c(\bar{x}^{(k+1,r)}) = -(\bar{N}^{(k)T})^{+} c(\bar{x}^{(k+1,r)})$$

(3.3.111)

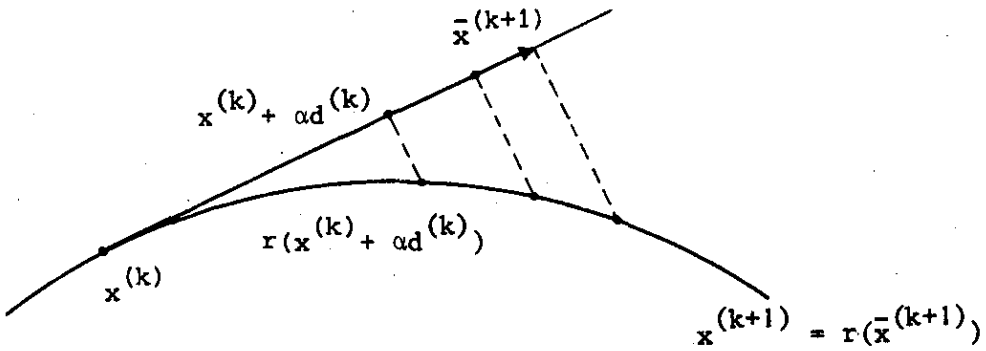
De matrix $\bar{N}^{(k)}$ die vergelijkbaar is met de in pt. 3.2.7 besproken matrix A bevat in eerste instantie de normalen van de actieve beperkingen in $x^{(k)}$. Als tijdens de hele restauratieprocedure geen andere beperkingen overschreden worden dan die, die in $x^{(k)}$ actief waren, kan daarvoor de eerder toegepaste matrix $N^{(k)}$ worden gebruikt. In andere gevallen moet de matrix $\bar{N}^{(k)}$ aangepast worden, zo mogelijk door gebruik te maken van de normalen van de beperkingen zoals deze berekend zijn in het punt $x^{(k)}$. In het geval we restauratie willen uitvoeren door in elk bereikt niet-toegelaten punt alle informatie opnieuw te bepalen wordt in elke stap precies gehandeld zoals in pt. 3.2.7 beschreven voor het geval van een toegelaten startpunt.

3.3.24. Een heel ander probleem dat op kan treden ten gevolge van het niet-lineair zijn van de beperkingen (met de daaruit voortvloeiende noodzaak tot restauratie) is het volgende. De meeste gebruikte methoden garanderen, dat in het gevonden niet-toegelaten punt $\bar{x}^{(k+1)}$ geldt $f(\bar{x}^{(k+1)}) < f(x^{(k)})$. Tengevolge van de restauratie is het echter mogelijk dat in het uit $\bar{x}^{(k+1)}$ bepaalde toegelaten punt $x^{(k+1)}$ geldt $f(x^{(k+1)}) > f(x^{(k)})$. Dit houdt de mogelijkheid tot cykelen in (zie Figuur 3.3.24a). Voorkomen kan men dat door in de gebruikte lijnminimaliseringsprocedure een andere strategie



Figuur 3.3.24a : Een mogelijk geval van cykelen.

toe te passen zoals geïllustreerd in Figuur 3.3.24b. Zij r(x) het toegelaten punt, dat gevonden wordt bij restauratie uitgaande van het punt x. Dan moet in de gebruikte lijnminimaliseringsprocedure telkens het minimum bepaald worden van $f(r(x^{(k)} + \alpha d^{(k)}))$ in plaats van van $f(x^{(k)} + \alpha d^{(k)})$



Figuur 3.3.24b : Lijnminimaliseringsmethode om cykelen te voorkomen.

3.3.25. Bij toepassing van geconjugeerde-richtingen- en quasi-Newton methoden treedt er in het geval dat geen verandering optreedt in de verzameling van actieve beperkingen doch wel restauratie nodig is omdat de actieve beperkingen niet-lineair zijn nog een kleine extra complicatie op. Deze complicatie wordt veroorzaakt doordat de expliciet in de aanpassingsformules voorkomende grootheden (H,y en s) uit de voorgaande stap betrekking hebben op een andere lineaire variëteit dan waarover in het huidige iteratiepunt moet worden geminimaliseerd. In het geval dat gebruik wordt gemaakt van de aanpassingsformules voor $\mathbb{R}^{n-q}$ (zoals (3.3.95) (3.3.103) (3.3.107) en (3.3.108)) dan zal deze complicatie in de practijk verwaarloosd kunnen worden. In het geval, zoals bij de methode van Goldfarb, echter gebruik gemaakt wordt van de bij uitsluitend lineaire beperkingen equivalente formuleringen in $\mathbb{R}^n$ (zoals (3.3.96) en (3.3.106))

dan worden zoekrichtingen gegenereerd die niet langer in het raakvlak aan de huidige beperkingen liggen. In dit laatste geval is de formulering van de aanpassingsformules in $\mathbb{R}^n$ niet meer equivalent met die in $\mathbb{R}^{n-q}$ en is een extra projectie van de gegenereerde zoekrichting noodzakelijk.

3.3.26. Als laatste opmerking in verband met niet-lineaire beperkingen zij vermeld dat de aanwezigheid van niet-lineaire beperkingen in de practijk tot aanzienlijk meer rekenwerk en een groot aantal kleinere complicaties leidt bij de realisering van primale minimaliseringsalgorithmen. In het bijzonder geldt dat de normalen van alle niet-lineaire beperkingen in ieder iteratiepunt opnieuw moeten worden geëvalueerd. Geen van de hiervoor uitgebreid besproken aanpassingen voor veranderingen in de verzameling van actieve beperkingen kunnen meer worden toegepast en met de restauratieprocedure wordt een in principe oneindig iteratieproces geïntroduceerd in iedere iteratieslag. In verband daarmee gaat ook de nauwkeurigheid waarmee aan ieder van de beperkingen moet worden voldaan een grote rol spelen. De primale methoden verliezen bij niet-lineaire beperkingen veel van hun voordelen t.o.v. de andere methoden voor de minimalisering onder nevenvoorwaarden, zoals de in de volgende paragraaf te bespreken boete-functie-methoden. In de literatuur wordt daarom wel gesuggereerd om gebruik te maken van een gemengde methode waarbij de niet-lineaire beperkingen via een boete-functie aanpak worden verdisconteerd in de objectfunctie en waarbij dan de minimalisering van deze objectfunctie onder uitsluitend lineaire nevenvoorwaarden wordt uitgevoerd met behulp van primale algorithmen. Ook op de THE wordt hiermee geëxperimenteerd [3.3.7] in het kader van de verdere uitwerking van het in eerste instantie door Dirkx ontwikkelde Algol programma NONLINMIN ([3.3.3] en [3.3.10]) waarin voor het minimaliseren van niet-lineaire functies onder niet-lineaire nevenvoorwaarden gebruik wordt gemaakt van een aantal van de hiervoor besproken primale algorithmen.

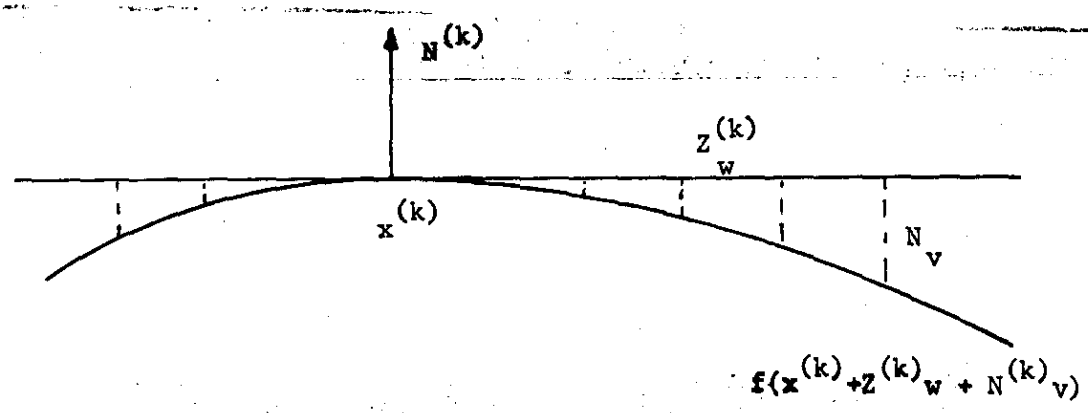
3.3.27. De convergentiesnelheid van primale methoden in het geval van niet-lineaire nevenvoorwaarden blijkt nauw samen te hangen met de conditie, ofwel verhouding van grootste tot kleinste eigenwaarde, van de restrictie van de Hessiaan van de Lagrange functie tot het raakvlak aan de beperkingen, d.i. de matrix

$$Z^{(k)T} \nabla^2_{xx} \mathcal{L}(x^{(k)}, \lambda^{(k)}) Z^{(k)} \qquad (3.3.112)$$

waarin (vgl(3.1.41))

$$\lambda^{(k)} := (N^{(k)T} N^{(k)}) N^{(k)T} \nabla f(x^{(k)}) \qquad (3.3.113)$$

De reden hiervoor, die door Luenberger in [1.1.1] in meer detail werd uitgewerkt, is het feit dat de primale methoden in het geval van niet-lineaire beperkingen kunnen worden opgevat als toepassingen van gradiëntmethoden in het locale raakvlak aan de beperkingen op een objectfunctie die ontstaat door de functiewaarden op het werkelijke oppervlak van de beperkingen tenzij te projecteren op het raakvlak in kwestie (zie figuur 3.3.27)



Figuur 3.3.27: Projectie van de objectfunctiewaarden op de gelineariseerde beperkingen

In tweede-orde-ontwikkeling leidt dit tot het probleem in de k-de iteratiestap

$$\min\{f(x^{(k)}) + \nabla^T f(x^{(k)})(Z^{(k)} w + N^{(k)} v) + \tfrac{1}{2} w^T Z^{(k)T} G^{(k)} Z^{(k)} \mid$$

$$\mid (N^{(k)T} N^{(k)} v)_i = -\tfrac{1}{2} w^T Z^{(k)T} H_i^{(k)} Z^{(k)} w, \; i=1,\ldots q\}$$

Uitwerking van de nevenvoorwaarden geeft dat

$$\nabla^T f(x^{(k)}) N^{(k)} v = \nabla^T f(x^{(k)}) N^{(k)} (N^{(k)T} N^{(k)})^{-1} \begin{pmatrix} -\tfrac{1}{2} w^T Z^{(k)T} H_1^{(k)} Z^{(k)} w \\ -\tfrac{1}{2} w^T Z^{(k)T} H_2^{(k)} Z^{(k)} w \\ \vdots \\ \vdots \end{pmatrix}$$

ofwel indien gebruik gemaakt wordt van de definitie (3.3.113)

$$\nabla^T f(x^{(k)}) N^{(k)} v = -\tfrac{1}{2} w^T Z^{(k)} \left(\sum_{i=1}^{q} \lambda_i^{(k)} H_i^{(k)} \right) Z^{(k)} w$$

Substitie hiervan is de tweede-orde ontwikkeling van het probleem leidt dan inderdaad tot het equivalente onbeperkte minimaliseringsprobleem

$$\min\{f(x^{(k)}) + \nabla^T f(x^{(k)}) Z^{(k)} w + \tfrac{1}{2} w^T Z^{(k)T} \nabla^2_{xx} \mathcal{L}^{(k)} Z^{(k)} w\}$$

Voor meer informatie over deze aanpak zij verwezen naar Luenberger ([1.1.1], § 11.5 en § 11.7).

Referenties

3.3.27. Specifieke details over de in deze paragraaf besproken hoger-orde-primale algorithmen en restauratie procedures kunnen o.a. worden gevonden in de volgende publicaties:

[3.3.1] : Zie [1.1.4] Gill & Murray (1974)

[3.3.2] : Zie [2.10.16] Rosen (1960)

[3.3.3] : Zie [3.2.13] Dirkx (1975)

[3.3.4] : Zie [3.2.23] Gill & Murray (1972)

[3.3.5] : Zie [3.2.24] Gill & Murray (1973)

[3.3.6] : Fischer, J.: On minimization under linear equality constraints in: Oettli, W. and Ritter, K.(Eds): "Optimization and Operations Research", Lecture Notes in Economics and Mathematical Systems nr. 117, Springer Verlag, Berlin (1976) pp. 77-82

[3.3.7] : Frederix, G.H.M.: "Het gecombineerd gebruik van boetefunctie- en projectiemethoden in een algorithme voor het minimaliseren van niet-lineaire functies onder niet-lineaire nevenvoorwaarden". Stageverslag, Technische Hogeschool Eindhoven, Onderafdeling der Wiskunde, (October 1976)

[3.3.8] : Goldfarb, D.: Extension of Davidon's variable metric method to maximization under linear inequality and equality constraints, SIAM J. Appl. Math. 17 (1969) pp. 739-764

[3.3.9] : Goldfarb, D.: Matrix factorisations in optimization of non-linear constraints, Math. Progr. 10 (1976) pp. 1-31

[3.3.10]: de Jong, J.L.: NONLINMIN, een procedure voor het minimaliseren van niet-lineaire functies onder niet-lineaire nevenvoorwaarden, Syllabus Colloquium Numerieke Programmatuur, Deel 2, Mathematisch Centrum, Amsterdam (1977)

[3.3.11] : McCormick, G.P.: A second order method for the linearly constrained nonlinear programming problem, in : Rosen, J.B., Mangasarian, O.L. en Ritter, K. (Eds) "Nonlinear Programming" Academic New York (1970)

[3.3.12] : McCormick, G.P.: Anti-zig-zagging by bending.Management Science, 15 (1969) pp. 315-320

[3.3.13] : Murtagh, B.A.and Sargent, R.W.H.: A constrained minimization method with quadratic convergence, in : R. Fletcher (Ed) :"Optimization",Academic Press, New York (1972) pp. 215-246

§ 3.4. Boete- en barrièrefunctiemethoden

3.4.1. Tot de oudst bekende methoden voor het numeriek oplossen van niet-lineaire minimaliseringsproblemen met niet-lineaire nevenvoorwaarden behoren de in deze paragraaf te bespreken boete- en barrièrefunctiemethoden. Het idee achter deze methoden is dat het beperkte minimaliseringsprobleem wordt vervangen door een rij van onbeperkte minimaliseringsproblemen waarvan de oplossingen convergeren naar de oplossing van het originele beperkte probleem. (Omdat men deze vervanging ook kan opvatten als een transformatie spreekt men in de literatuur (bv. [3.4.2]) ook wel over transformatiemethoden). De rij van onbeperkte minimaliseringsproblemen kan daarbij ofwel afhankelijk zijn van een in principe vooraf te definiëren rij van parameters, in welk geval men spreekt van parametrische boete- en barrièrefunctiemethoden, ofwel iteratief worden gegenereerd, in welk geval men spreekt van niet-parametrische boete- en barrièrefunctiemethoden. Van beide klassen van methoden zullen hierna een aantal voorbeelden en een aantal eigenschappen worden besproken. Aan het einde van de paragraaf zal daarna nog aandacht worden besteed aan de methoden voor het bepalen van de oplossingen van de gegenereerde onbeperkte minimaliseringsproblemen.

Boetefuncties

3.4.2. Het min of meer klassieke voorbeeld van het gebruik van boetefuncties is de reeds in 1943 door Courant [3.4.10] gepropageerde toepassing van de boetefunctie aanpak voor de oplossing van het algemene niet-lineaire minimaliseringsprobleem met uitsluitend gelijkheidsvoorwaarden, d.i. het probleem van het type GNLE (vgl. pt. 3.1.4)

$$\min \{f(x) \mid c_i(x) = 0,\ i = 1,\ldots,m,\ x \in \mathbb{R}^n\} \qquad (3.4.1)$$

Dit beperkte minimaliseringsprobleem wordt daarbij met behulp van een monotoon naar 0 convergerende rij $\{r_1, r_2 \ldots\}$ van (positieve) getallen vervangen door de rij van onbeperkte minimaliseringsproblemen

$$\min \{f(x) + \frac{1}{r_k} \Psi(x) \mid x \in \mathbb{R}^n\} \qquad (3.4.2)$$

waarin $\Psi(x)$ gebruikelijk een functie is van de gedaante

$$\Psi(x) = \sum_{i=1}^{m} \psi_i(c_i(x)) \tag{3.4.3}$$

en de functies $\psi_i(t)$ continue reëel-waardige (en bij voorkeur convexe) functies zijn van één variabele met de eigenschappen dat

$$\begin{aligned} \psi_i(t) &= 0 \quad \text{als} \quad t = 0 \\ &> 0 \quad \text{als} \quad t \neq 0 \\ &\to \infty \quad \text{als} \quad |t| \to \infty \end{aligned} \tag{3.4.5}$$

De functie

$$P(x,r) := f(x) + \frac{1}{r}\Psi(x) := f(x) + \frac{1}{r}\sum_{i=1}^{m} \psi_i(c_i(x)) \tag{3.4.6}$$

die gelijk is aan de som van de objectfunctie en een aantal termen die kunnen worden opgevat als "boeten" voor het overschrijden van de beperkingen wordt toepasselijk de <u>boetefunctie</u> (eng. : <u>penalty function</u>) genoemd. De functie $\Psi(x)$ (3.4.3) wordt in dit verband vaak de verliesfunctie (eng. : loss function) genoemd. Voorbeelden van functies $\psi_i(t)$ die de bouwstenen voor de functie $\Psi(x)$ vormen zijn de bekende kwadratische functie

$$\psi_i(t) = \frac{1}{2} k_i t^2 \tag{3.4.7}$$

en de modulus functie

$$\psi_i(t) = k_i |t| \tag{3.4.8}$$

waarin de k_i eventuele gewichtsfactoren voorstellen. Vooral de eerste van deze functies, die in feite ook reeds door Courant [3.4.10] werd gebruikt, wordt veelvuldig in de practijk toegepast.

3.4.3. In het geval de beperkingen in de originele probleemformulering uitsluitend ongelijkheidsbeperkingen zijn, d.i. bij een probleemformulering van de gedaante

$$\min \{f(x) \mid g_i(x) \geq 0,\ i = 1,\ldots,m,\ x \in \mathbb{R}^n\} \tag{3.4.9}$$

moet een oplossing $\hat{x}$ bepaald worden die behoort tot het toegelaten gebied (vgl. (3.4.3))

$$S := \{x \in \mathbb{R}^n \mid g_i(x) \geq 0\} \tag{3.4.10}$$

Toepassing van het idee van de boetefuncties is mogelijk in dit geval door de definitie $P(x,r)$ met de eigenschappen

$$\begin{aligned} P(x,r) &= f(x) \qquad \text{als} \qquad x \in S \\ &> f(x) \qquad \text{als} \qquad x \notin S \end{aligned} \tag{3.4.11}$$

Bij een probleemformulering met ongelijkheden zoals (3.4.9) wordt een dergelijke boetefunctie gegeven door de aan (3.4.6) verwante definitie

$$P(x,r) := f(x) + \frac{1}{r}\ \Psi^-(x) \tag{3.4.12}$$

waar

$$\Psi^-(x) := \sum_{i=1}^{m} \psi_i(g_i^-(x)) \tag{3.4.13}$$

in welke uitdrukking de functies ψ_i mogelijk dezelfde reëelwaardige (en mogelijk convexe) functies van één variabele zijn als boven gedefinieerd ((3.4.7), (3.4.8)) en de functies $g_i^-(x)$ de negatieve componenten zijn van de functies $g_i(x)$ volgens de definitie

$$g_i^-(x) := \min [0, g_i(x)] \quad := -\max [0, -g_i(x)] \tag{3.4.14}$$

Opgemerkt kan worden dat de aldus gedefinieerde verliesfunctie $\Psi^-(x)$, ondanks het feit dat de functies $g_i^-(x)$ meestal discontinu zijn in de eerste afgeleide (daar waar $g_i(x) = 0$), toch meestal continu differentieerbaar zal zijn als, zoals gebruikelijk, de functies $\psi_i(t)$ continu differentieerbaar zijn met afgeleide gelijk aan nul in het punt $t = 0$.

Barrièrefuncties

3.4.4. Bij minimaliseringsproblemen met ongelijkheden zoals besproken in het voorgaande punt kan in het geval het toegelaten gebied (3.4.10) een inwendig punt heeft, d.i. in het geval

$$S^0 := \{x \in \mathbb{R}^n \mid g_i(x) > 0,\ i = 1,\ldots m\} \tag{3.4.15}$$

niet leeg is, ook gebruik gemaakt worden van de nauw aan de boetefuncties verwante barrièrefuncties. In plaats van boetetermen bij het overschrijden van de beperkingen voegt men dan barrièretermen toe aan de objectfunctie welke termen groter worden naarmate de rand van het toegelaten gebied dichter benaderd wordt. Uitgaande van een toegelaten punt minimaliseert men in plaats van het beperkte minimaliseringsprobleem (3.4.9)

$$\min \{f(x) \mid g_i(x) \geq 0,\ i = 1,\ldots m,\ x \in \mathbb{R}^n\}$$

de met de monotoon naar 0 convergerende rij $\{r_1, r_2 \ldots\}$ corresponderende rij van onbeperkte minimaliseringsproblemen

$$\min \{f(x) + r_k\, \Phi(x) \mid x \in S^0\} \tag{3.4.16}$$

waarin $\Phi(x)$ gebruikelijk een functie is van de gedaante

$$\Phi(x) = \sum_{i=1}^{m} \varphi_i(g_i(x)) \tag{3.4.17}$$

en de functies $\varphi_i(t)$ voor $t > 0$ gedefinieerde continue reëelwaardige functies zijn van een variabele met de eigenschappen dat

$$\begin{aligned} \varphi_i(t) &\to \infty && t \to 0 \\ &> 0 && t > 0 \\ &\to 0 && t \to \infty \end{aligned} \tag{3.4.18}$$

Van deze eigenschappen impliceert de eerste dat als het ware een barrière wordt opgeworpen zodra het argument (d.i. de waarde van de functie $g_i(x)$) nul nadert. Om die reden wordt de functie

$$B(x,r) := f(x) + r\, \Phi(x) \tag{3.4.19}$$

dan ook barrièrefunctie (eng. : barrierfunction) genoemd. Voorbeelden van in de practijk toegepaste functies $\varphi_i(t)$ zijn o.a. de inverse barrièrefunctie, welke o.a. werd gepropageerd door Fiacco & McCormick bij hun z.g. SUMT-(= Sequential Unconstrained Minimization Technique)-methode [3.4.7]

$$\varphi_i(t) = k_i \frac{1}{t} \tag{3.4.20}$$

en de kwadratische inverse functie (vgl. [3.4.12])

$$\varphi_i(t) = k_i \frac{1}{t^2} \tag{3.4.21}$$

waarin de k_i weer eventuele gewichtsfactoren voorstellen.

Ook onderzocht en met veel succes toegepast in de practijk is de logarithmische barrièrefunctie (vgl. [3.4.17] en [3.4.18]) die gebruik maakt van functies $\varphi_i(t)$ die niet voldoen aan de twee laatste in (3.4.18) genoemde eigenschappen

$$\varphi_i(t) = -k_i \ln t \tag{3.4.22}$$

Voor deze keuze voor de functies $\varphi_i(t)$ gelden analoge convergentie-bewijzen als hierna te geven voor het geval dat $\varphi_i(t)$ alle in (3.4.18) genoemde eigenschappen bezit.

Classificatie van boete- en barrièrefunctiemethoden

3.4.5. In verband met de bovengegeven verschillende mogelijkheden voor de functies $\varphi_i(t)$ definieerde Lootsma in [3.4.17] ten behoeve van de classificatie van barrièrefunctiemethoden de orde van barrièrefuncties als volgt:

Definitie 3.4.5a: Een barrièrefunctie B(x,r) (3.4.19) heet van de orde λ als de afgeleiden $\varphi_i'(t)$ van de corresponderende functies $\varphi_i(t)$ een pool hebben van de orde λ in het punt $t = 0$.

Op grond van deze definitie volgt onmiddellijk dat de barrière-functies die corresponderen met de in (3.4.20) (3.4.21) en (3.4.22) genoemde functies $\varphi_i(t)$ respectievelijk van de orde 2, 3 en 1 zijn.

De orde van boetefuncties kan op analoge wijze gedefinieerd: De definitie van Lootsma [3.4.17] daarvoor luidt

Definitie 3.4.5b: Een boetefunctie $P(x,r)$ ((3.4.6) of (3.4.12)) heet van de orde λ als de afgeleiden $\psi_i'(t)$ van de corresponderende functies $\psi_i(t)$ een nulpunt hebben van orde λ in het punt $t = 0$.

Volgens deze definitie geldt dat de kwadratische boetefunctie (3.4.7) van de orde 1 is. De modulusfunctie (3.4.8) valt buiten deze classificatie.

Opgemerkt kan worden dat Lootsma in [3.4.17] in de door hem beschouwde standaarduitdrukkingen voor de boete-en barrièrefunctiemethoden de parameters van een exponent-array gelijk aan de orde van de betreffende functies. Zijn standaarduitdrukking voor een boetefunctie van orde λ werd daarmee (vgl. (3.4.6))

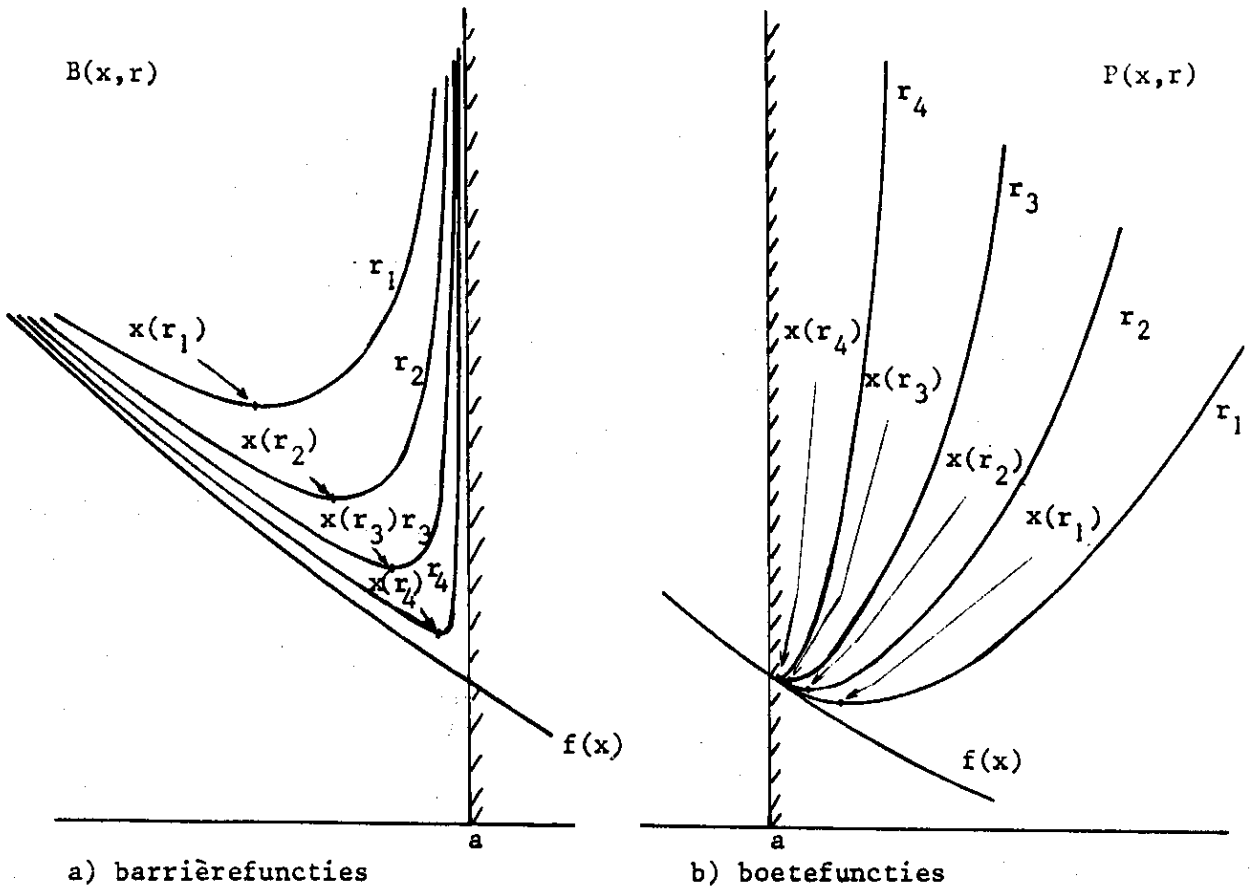
$$P(x,r) := f(x) + r^{-\lambda}\psi(x) \tag{3.4.23}$$

en die voor een barrièrefunctie van orde λ analoog (vgl. (3.4.19))

$$B(x,r) := f(x) + r^{\lambda}\phi(x) \tag{3.4.24}$$

Van deze uitdrukkingen zal hieronder geen gebruik worden gemaakt.

3.4.6. Een illustratie van het verschil in gebruik tussen boetefunctie- en barrièrefunctiemethoden is gegeven in Figuur 3.4.6. Opgemerkt kan worden dat in beide gevallen de opvolgend bepaalde minima $x(r_k)$ continu afhankelijk zijn van de parameter r. Hierop zal in pt. 3.4.20 worden teruggekomen in verband met de mogelijkheid van het gebruik van extrapolatie van deze continue functie.



Figuur 3.4.6.: Vergelijking van het gebruik van boetefuncties en van barrièrefuncties bij het probleem min $\{f(x) \mid a - x \geq 0\}$

Figuur 3.4.6 illustreert ook duidelijk dat in het geval van boetefuncties deze opvolgende minima $x(r_k)$ buiten het toegelaten gebied liggen, terwijl dezelfde minima bij barrièrefuncties in het inwendige van het toegelaten gebied gevonden worden. In verband hiermee worden de boetefunctiemethoden die gebruik maken van de in pt. 3.4.2 en pt. 3.4.3 besproken boetefuncties in de literatuur ook wel aangeduid als de "parametric exterior-point-methods" terwijl de barrièrefunctiemethoden die gebruik maken van de in pt. 3.4.4 besproken barrièrefuncties de "parametric interior-point-methods" worden genoemd.

3.4.7. In het geval dat naast elkaar zowel gelijkheids- als ongelijkheids-beperkingen voorkomen, d.i. in het geval van een probleem van het type GNLI (vgl. pt. 3.1.4)

$$\min\{f(x) \mid g_i(x) \geq 0,\ i \in I,\ h_j(x) = 0,\ j \in E,\ x \in \mathbb{R}^n\} \qquad (3.4.25)$$

kan gebruik gemaakt worden van een z.g. gemengde boetefunctie $M(x,r)$ waarin zowel boetefunctie- als barrièrefunctietermen voorkomen. De barrièrefunctietermen betreffen in de practijk die ongelijkheidsbeperkingen die passief zijn in het startpunt. Het onbeperkte minimaliseringsprobleem dat in de k^0 iteratie moet worden opgelost luidt dan

$$\min \{f(x) + r_k \sum_{i\in I_1} \varphi_i(g_i(x)) + \frac{1}{r_k} \sum_{i\in I_2} \psi_i(g_i^-(x))$$

$$+ \frac{1}{r_k} \sum_{j\in E} \psi_j(h_j(x)) \mid x \in S_k^0 \subset \mathbb{R}^n\} \qquad (3.4.26)$$

waarin

$$\begin{aligned} I_1 &:= \{i \in I \mid g_i(x_{k-1}) > 0\} \\ I_2 &:= \{i \in I \mid g_i(x_{k-1}) \leq 0\} \end{aligned} \qquad (3.4.27)$$

en

$$S_k^0 := \{x \in \mathbb{R}^n \mid g_i(x) > 0\ ,\ i \in I_1\} \qquad (3.4.28)$$

Het probleem (1.1.4) is een voorbeeld van een dergelijke algemene gemengde boetefunctie.

Convergentie van parametrische boetefunctie- en barrièrefunctiemethoden

3.4.8. Alhoewel boete- zowel als barrièrefunctiemethoden vanuit theoretisch standpunt bezien strikt genomen geen iteratieve methoden zijn - het minimum $x(r_{k+1})$ is in geen enkel opzicht afhankelijk van het eerder gevonden minimum $x(r_k)$ - worden boetefunctiemethoden in de practijk wel als iteratieve methoden opgevat. Uitgangspunt daarbij is de (mogelijk gegenereerde) monotoon naar 0 convergerende rij van parameters $\{r_1, r_2, \ldots\}$ waarvoor geldt

$$0 <\ldots< r_{k+1} < r_k < r_{k-1} <\ldots< r_1 \tag{3.4.29}$$

De oplossing $x_k := x(r_k)$ die correspondeert met de parameter waarde r_k en daarmee met het onbeperkte minimaliseringsprobleem

$$\min \{f(x) + \frac{1}{r_k} \Psi(x) \mid x \in \mathbb{R}^n\} \tag{3.4.30}$$

wordt daarbij dan steeds gebruikt als startpunt voor de oplossing van het onbeperkte minimaliseringsprobleem corresponderend met de parameter r_{k+1} , d.i. het probleem

$$\min \{f(x) + \frac{1}{r_{k+1}} \Psi(x) \mid x \in \mathbb{R}^n\} \tag{3.4.31}$$

Voor de opvolgende minima $x_k := x(r_k)$ gelden een aantal interessante relaties samengevat in het volgende lemma:

Lemma 3.4.8 (vgl. [3.4. 5]) Als x_k het minimum is van het probleem (3.4.30) corresponderend met de parameter r_k en de parameters geordend zijn volgens (3.4.29) dan gelden als de functie P(x,r) gedefinieerd is door (3.4.6)

$$P(x,r) := f(x) + \frac{1}{r} \Psi(x)$$

met $\Psi(x)$ gegeven door (3.4.3) de volgende relaties

$$P(x_{k+1},r_{k+1}) \geq P(x_k,r_k) \tag{3.4.32}$$

$$\Psi(x_{k+1}) \leq \Psi(x_k) \tag{3.4.33}$$

$$f(x_{k+1}) \geq f(x_k) \tag{3.4.34}$$

Bovendien geldt als $\hat{x}$ een oplossing is van het originele probleem (3.4.1) dat

$$f(\hat{x}) \geq P(x_k,r_k) \geq f(x_k) \tag{3.4.35}$$

Bewijs: Omdat $\Psi(x) \geq 0$ voor alle $x \in \mathbb{R}^n$ en $r_{k+1} < r_k$ geldt

$$f(x_{k+1}) + \frac{1}{r_{k+1}} \Psi(x_{k+1}) \geq f(x_{k+1}) + \frac{1}{r_k} \Psi(x_{k+1}) \geq f(x_k) + \frac{1}{r_k} \Psi(x_k)$$

waarmee (3.4.32) is aangetoond. Verder geldt zowel

$$f(x_k) + \frac{1}{r_k} \Psi(x_k) \leq f(x_{k+1}) + \frac{1}{r_k} \Psi(x_{k+1}) .$$

als

$$f(x_{k+1}) + \frac{1}{r_{k+1}} \Psi(x_{k+1}) \leq f(x_k) + \frac{1}{r_{k+1}} \Psi(x_k)$$

hetgeen bij optelling en herrangschikking leidt tot

$$\left(\frac{1}{r_{k+1}} - \frac{1}{r_k}\right)\Psi(x_{k+1}) \leq \left(\frac{1}{r_{k+1}} - \frac{1}{r_k}\right)\Psi(x_k)$$

waaruit onmiddellijk (3.4.33) volgt. Combinatie van dit resultaat met de ongelijkheid

$$f(x_{k+1}) + \frac{1}{r_k} \Psi(x_{k+1}) \geq f(x_k) + \frac{1}{r_k} \Psi(x_k)$$

ofwel

$$f(x_{k+1}) - f(x_k) \geq \frac{1}{r_k} (\Psi(x_k) - \Psi(x_{k+1})) \geq 0$$

geeft de gezochte relatie (3.4.34). In het optimale punt $\hat{x}$ geldt $e_i(\hat{x}) = 0$ waarmee $\Psi(\hat{x}) = 0$ en

$$f(\hat{x}) = f(\hat{x}) + \frac{1}{r_k} \Psi(\hat{x}) \geq f(x_k) + \frac{1}{r_k} \Psi(x_k) \geq f(x_k)$$

waaruit de relatie (3.4.35) direct volgt. □

3.4.9. De relaties (3.4.32) en (3.4.34) vormen de voornaamste peilers van het bewijs van de volgende convergentiestelling voor parametrische boete-functiemethoden.

Stelling 3.4.9 : Als $\{r_k\}$ een monotoon naar 0 convergerende rij van parameters is en $\{x_k\}$ de rij van daarmee corresponderende oplossingen van onbeperkte minimaliseringsproblemen (3.4.30) dan geldt dat ieder verdichtingspunt van de rij $\{x_k\}$ een oplossing is van het originele beperkte minimaliseringsprobleem (3.4.1).

Bewijs : Als $\{x_k\}_{k\in K}$ een deelrij is die convergeert naar $\bar{x}$, dan geldt op grond van de continuïteit van f dat

$$\lim_{k\in K} f(x_k) = f(\bar{x})$$

De rij $\{P(x_k,r_k)\}_{k\in K}$ is een naar boven door $f(\hat{x})$ begrensde niet-dalende rij die (daarom) een limiet heeft waarvoor geldt

$$\lim_{k\in K} P(x_k,r_k) = \bar{P} \leq f(\hat{x})$$

waar $\hat{x}$ een oplossing is van het originele probleem. Aftrekken van beide limieten levert

$$\lim_{k\in K} \frac{1}{r_k} \Psi(x_k) = \bar{P} - f(\bar{x}) \leq f(\hat{x}) - f(\bar{x})$$

Omdat $\Psi(x) \geq 0$ en $r_k \to 0$ volgt noodzakelijkerwijs dat

$$\lim_{k\in K} \Psi(x_k) = 0$$

waaruit op grond van de continuïteit van Ψ weer volgt dat

$$\Psi(\bar{x}) = 0$$

en derhalve dat (op grond van (3.4.3) en (3.4.5)

$$c_i(\bar{x}) = 0 \qquad i = 1,\ldots,m$$

Het verdichtingspunt $\bar{x}$ is dus een toegelaten punt waarvoor geldt dat

$$f(\bar{x}) = \lim_{k\in K} f(x_k) \leq f(\hat{x})$$

hetgeen impliceert dat (ook) $\bar{x}$ een oplossing is van het originele probleem. □

3.4.10. Voor de toepassing van barrièrefunctiemethoden in het geval van minimaliseringsproblemen met uitsluitend ongelijkheidsbeperkingen en een toegelaten gebied met een niet leeg inwendige gelden analoge resultaten als voor de hiervoor besproken boetefunctiemethode. In het bijzonder geldt dat als $\{r_1, r_2, \ldots\}$ een monotoon naar 0 convergerende rij van parameters is waarvoor geldt (3.4.29)

$$0 < \ldots < r_{k+1} < r_k < r_{k-1} < \ldots < r_0$$

en $x_k := x(r_k)$ de oplossing is van het met parameterwaarde r_k corresponderende onbeperkte minimaliseringsprobleem

$$\min \{f(x) + r_k \Phi(x) \mid x \in S^0 (x_{k-1})\} \tag{3.4.36}$$

dan gelden voor de opvolgende oplossingen x_k de in het volgende lemma samengevatte relaties:

Lemma 3.4.10 : Als x_k het minimum is van het probleem (3.4.36) corresponderend met de parameter r_k en de parameters r_k een monotoon naar 0 convergerende rij vormen dan gelden als de functie B(x,r) is gedefinieerd door (3.4.19)

$$B(x,r) := f(x) + r\Phi(x)$$

met $\Phi(x)$ gegeven door (3.4.17) de volgende relaties

$$B(x_{k+1}, r_{k+1}) \leq B(x_k, r_k) \tag{3.4.37}$$

$$\Phi(x_{k+1}) \geq \Phi(x_k) \tag{3.4.38}$$

$$f(x_{k+1}) \leq f(x_k) \tag{3.4.39}$$

en bovendien

$$f(\hat{x}) \leq f(x_k) \leq B(x_k, r_k) \tag{3.4.40}$$

waar $\hat{x}$ de oplossing is van het originele probleem (3.4.9).

Bewijs : Het bewijs verloopt analoog aan het bewijs van Lemma 3.4.8. Omdat $\Phi(x) > 0$ voor $x \in S^0$ en $r_{k+1} < r_k$ geldt

$$f(x_k) + r_k \Phi(x_k) \geq f(x_k) + r_{k+1}\Phi(x_k) \geq f(x_{k+1}) + r_{k+1}\Phi(x_{k+1})$$

waarmee (3.4.37) is aangetoond. Verder geldt dat

$$f(x_{k+1}) + r_{k+1} \Phi(x_{k+1}) \leq f(x_k) + r_{k+1} \Phi(x_k)$$

en

$$f(x_k) + r_k\Phi(x_k) \leq f(x_{k+1}) + r_k\Phi(x_{k+1})$$

waaruit na optelling en rangschikking volgt

$$(r_k - r_{k+1}) \Phi(x_k) \leq (r_k - r_{k+1}) \Phi(x_{k+1})$$

en daarmee ongelijkheid (3.4.38). Combinatie van dit laatste resultaat met

$$f(x_{k+1}) + r_{k+1} \Phi(x_{k+1}) \leq f(x_k) + r_{k+1}\Phi(x_k)$$

geeft onmiddellijk de derde ongelijkheid (3.4.39)

$$f(x_{k+1}) - f(x_k) \leq r_{k+1}(\Phi(x_k) - \Phi(x_{k+1})) \leq 0$$

De laatste ongelijkheid (3.4.40) volgt uit de overweging dat alle x_k toegelaten punten zijn terwijl $\hat{x}$ het toegelaten punt is met de minimale objectfunctiewaarde en uit de overweging dat de termen $r_k\Phi(x_k)$ steeds groter zijn dan 0. □

3.4.11. De ongelijkheidsrelaties (3.4.37) (3.4.39) en (3.4.40) vormen de basis voor het bewijs van de volgende aan Stelling 3.4.9 analoge convergentiestelling voor parametrische barrièrefunctiemethoden.

Stelling 3.4.11 Als $\{r_k\}$ een monotoon naar 0 convergerende rij van parameters is en $\{x_k\}$de rij van daarmee corresponderende oplossingen van het onbeperkte minimaliseringsprobleem (3.4.36) dan geldt dat ieder verdichtingspunt van de rij $\{x_k\}$ een oplossing is van het originele beperkte minimaliseringsprobleem (3.4.9)

Bewijs : Het bewijs verloop vrijwel analoog aan het van Stelling 3.4.9. Als $\{x_k\}_{k\in K}$ een deelrij is die convergeert naar $\bar{x}$ dan geldt op grond van de continuïteit van f dat

$$\lim_{k\in K} f(x_k) = f(\bar{x})$$

De rij $\{B(x_k,r_k)\}_{k\in K}$ is een naar onder begrensde niet-stijgende rij die (dus) een limiet heeft waarvoor geldt

$$\lim_{k\in K} B(x_k,r_k) = \bar{B} \geq f(\bar{x})$$

Stel dat $f(\bar{x}) > f(\hat{x})$ dan moet er een $\tilde{x} \in S^0$ bestaan zo dat

$$f(\bar{x}) > f(\tilde{x}) > f(\hat{x})$$

Er geldt dan dat

$$B(\tilde{x},r_k) = f(\tilde{x}) + r_k\Phi(\tilde{x}) \geq f(x_k) + r_k\Phi(x_k) = B(x_k,r_k)$$

hetgeen in de limiet resulteert in

$$\lim_{k\in K} B(\tilde{x},r_k) = f(\tilde{x}) \geq \lim_{k\in K} B(x_k,r_k) = \bar{B} \geq f(\bar{x})$$

hetgeen in tegenspraak is met de eerdere veronderstelling. De conclusie is daarom dat

$$\lim_{k\in K} f(x_k) = f(\bar{x}) = f(\hat{x})$$

hetgeen te bewijzen was. □

Opgemerkt kan worden dat dezelfde argumentatie die werd toegepast om aan te tonen in het bewijs dat $f(\bar{x}) = f(\hat{x})$ kan worden gebruikt om aan te tonen dat ook

$$\lim_{k\in K} B(x_k,r_k) = \bar{B} = f(\bar{x}) = f(\hat{x}) \qquad (3.4.41)$$

hetgeen in het bijzonder impliceert dat

$$\lim_{k\in K} r_k \Phi(x_k) = 0 \qquad (3.4.42)$$

Een analoog resultaat had ook kunnen worden bewezen voor de boetefunctiemethoden waar, mutatis mutandis,

$$\lim_{k\in K} \frac{1}{r_k} \Psi(x_k) = 0 \qquad (3.4.43)$$

Afgeleiden van boete- en barrièrefuncties

3.4.12. Bij de meeste methoden voor het oplossen van onbeperkte minimaliseringsproblemen spelen de eerste en tweede afgeleiden van de objectfunctie (d.i. de gradiënt en de Hessiaan) zowel vanuit theoretisch als vanuit practisch standpunt een belangrijke rol.
Omdat onbeperkte minimalisering het hoofdbestanddeel vormt van boete- en barrière-functiemethoden geldt dat in het bijzonder ook voor de gradiënt en de Hessiaan van boete- en barrièrefuncties. Vandaar dat hieronder nader wordt ingegaan op de afgeleiden van de hiervoor besproken boete- en barrièrefuncties. Ter vermijding van duplicaties worden daarbij naast elkaar alleen de eigenschappen behandeld van de boetefuncties P(x,r) voor uitsluitend ongelijkheidbeperkingen (vgl(3.4.12)) en de daarmee corresponderende barrièrefuncties B(x,r) (vgl(3.4.19)).
De eigenschappen van de afgeleiden van de boete functies P(x,r) voor gelijkheidsbeperkingen (vgl(3.4.6)) als ook die van de afgeleiden van de gemengde boetefuncties M(x,r)(vgl(3.4.26)) kunnen daar onmiddellijk van worden afgeleid.

3.4.13. De gradiënt van de boetefunctie P(x,r) voor ongelijkheidsbeperkingen (3.4.12) wordt formeel gegeven door de uitdrukking

$$\nabla^T P(x,r) := \nabla^T f(x) + \frac{1}{r} \nabla^T \Psi^-(x) \tag{3.4.44}$$

waarin met (3.4.13)

$$\nabla^T \Psi^-(x) := \sum_{i=1}^{m} \frac{d\psi_i}{dt} (g_i(x)) \nabla^T g_i^-(x) \tag{3.4.45}$$

Een probleem in deze uitdrukking is de gradiënt van de functie $g_i^-(x)$ in die punten waar $g_i(x) = 0$ en wel omdat de gradiënt daar niet zonder meer bestaat. Met de definitie

$$\begin{aligned} \nabla g_i^-(x) &:= 0 && \text{als } g_i(x) \geq 0 \\ &:= \nabla g_i(x) && \text{" } g_i(x) < 0 \end{aligned} \tag{3.4.46}$$

kan dit probleem echter eenvoudig worden opgelost, zij het dat er een discontinuïteit optreedt daar waar $g_i(x) = 0$. In het algemeen zal deze discontinuïteit resulteren in een discontinuïteit in de gradiënt van de boetefunctie tenzij het een boetefunctie betreft die volgens de definitie van Lootsma (vgl. pt. 3.4.5) van een orde λ is met λ groter dan nul. In dat geval immers geldt voor de afgeleiden van de functies $\psi_i(t)$, die de bouwstenen vormen van de verliesfunctie $\Psi^-(x)$ (vgl(3.4.13)), dat deze een nulpunt hebben (van orde λ) in het punt $t=0$ zodat de termen

$$\frac{d\psi_i}{dt} (g_i(x)) \nabla^T g_i^-(x)$$

continu zijn in de punten waar $g_i(x) = 0$. De gradiënt van de verliesfunctie $\Psi^-(x)$ wordt daarmee een continue funktie in x die (ook) gegeven wordt door

$$\nabla^T \Psi^-(x) := \sum_{i \in I_V(x)} \frac{d\psi_i}{dt} (g_i(x)) \nabla^T g_i(x) \tag{3.4.47}$$

waarin $I_V(x)$ de verzameling voorstelt van de indices van de overschreden beperkingen (3.1.22a)

$$I_V(x) := \{i \in \mathbb{N} \mid g_i(x) < 0,\ i=1,\ldots m\} \tag{3.4.48}$$

Uitgewerkt voor de kwadratische (orde 1) verliesfunctie (3.4.7) resulteert dit in

$$\nabla^T\psi^-(x) := \sum_{i\in I_V(x)} k_i g_i(x)\nabla^T g_i(x) \tag{3.4.49a}$$

of equivalent in

$$\nabla^T\psi^-(x) := \sum_{i=1}^{m} k_i g_i^-(x)\nabla^T g_i(x) \tag{3.4.49b}$$

waarmee op zijn beurt de gradiënt van de corresponderende boetefunctie (na transpositie) gelijk wordt aan

$$\begin{aligned}\nabla P(x,r) &:= \nabla f(x) + \sum_{i\in I_V(x)} \frac{k_i g_i(x)}{r}\nabla g_i(x)\\ &:= \nabla f(x) + \sum_{i=1}^{m} \frac{k_i g_i^-(x)}{r}\nabla g_i(x)\end{aligned} \tag{3.4.50}$$

Een analoge uitwerking voor de modulus-verliesfunctie (3.4.8) levert problemen omdat de functies $\psi_i(t)$ niet continu differentieerbaar zijn in het punt t=0.

3.4.14. Voor de gradiënt van de barrièrefunctie B(x,r) (3.4.19) geldt analoog aan (3.4.44)

$$\nabla^T B(x,r) := \nabla^T f(x) + r\,\nabla^T\Phi(x) \tag{3.4.51}$$

waarin met de impliciete veronderstelling dat alle beperkingen vertegenwoordigd in de barrièrefunctie passief zijn en Φ gegeven wordt door (3.4.17)

$$\nabla^T\Phi(x) := \sum_{i=1}^{m} \frac{d\varphi_i}{dt}(g_i(x))\nabla^T g_i(x) \tag{3.4.52a}$$

ofwel

$$\nabla^T\Phi(x) := \sum_{i\in I_p(x)} \frac{d\varphi_i}{dt}(g_i(x))\nabla^T g_i(x) \tag{3.4.52b}$$

waar $I_p(x)$ de verzameling voorstelt van alle beperkingen die passief zijn in x (vgl(3.1.22c))

$$I_p(x) := \{i \in \mathbb{N} \mid g_i(x) > 0,\ i=1,\ldots,m\} \tag{3.4.53}$$

Uitwerking voor de drie besproken barrièrefuncties geeft onmiddellijk het volgende resultaat:

a) voor de logarithmische barrièrefunctie (3.4.22)

$$\nabla^T\Phi(x) := \sum_{i\in I_p(x)} -\frac{k_i}{g_i(x)} \nabla^T g_i(x) \tag{3.4.54}$$

en

$$\nabla B(x,r) := \nabla f(x) - \sum_{i\in I_p(x)} \frac{k_i r}{g_i(x)} \nabla g_i(x) \tag{3.4.55}$$

b) voor de inverse barrièrefunctie (3.4.20)

$$\nabla^T\Phi(x) := \sum_{i\in I_p(x)} -\frac{k_i}{(g_i(x))^2} \nabla^T g_i(x) \tag{3.4.56}$$

en

$$\nabla B(x,r) := \nabla f(x) - \sum_{i\in I_p(x)} \frac{k_i r}{(g_i(x))^2} \nabla g_i(x) \tag{3.4.57}$$

c) voor kwadratische inverse barrièrefunctie (3.4.21)

$$\nabla^T\Phi(x) := \sum_{i\in I_p(x)} -2\frac{k_i}{(g_i(x))^3} \nabla^T g_i(x) \tag{3.4.58}$$

en

$$\nabla B(x,r) := \nabla f(x) - \sum_{i\in I_p(x)} 2\frac{k_i r}{(g_i(x))^3} \nabla g_i(x) \tag{3.4.59}$$

Relatie met Lagrange multiplicatoren

3.4.15. In het minimum x_k van de boetefunctie corresponderend met r_k

$$P(x,r_k) := f(x) + \frac{1}{r_k} \sum_{i=1}^{m} \psi_i(g_i^-(x)) \tag{3.4.60}$$

geldt dat de gradiënt gelijk wordt aan nul

$$\nabla P(x_k,r_k) := \nabla f(x_k) + \frac{1}{r_k} \sum_{i \in I_V(x_k)} \frac{d\psi_i}{dt}(g_i(x_k))\, \nabla g_i(x_k) = 0 \tag{3.4.61}$$

Dit resultaat blijkt goed te kunnen worden vergeleken met de noodzakelijke voorwaarde voor het minimum x_k van het beperkte minimaliseringsprobleem

$$\min \{f(x) \mid g_i(x) \geq g_i^-(x_k),\ i=1,\ldots,m\} \tag{3.4.62}$$

welke luidt

$$\nabla f(x_k) - \sum_{i=1}^{m} \lambda_{k,i}\, \nabla g_i(x_k) = 0 \tag{3.4.63}$$

waar $\lambda_{k,i}$ de i-de component is van de Lagrange-vector waarvoor geldt

$$\begin{aligned} \lambda_{k,i} &\geq 0 \quad \text{als } g_i(x_k) = g_i^-(x_k) \\ &= 0 \quad \text{"} \quad g_i(x_k) > 0 \end{aligned} \tag{3.4.64}$$

Vergelijking van beide uitdrukkingen geeft aanleiding tot de definitie

$$\begin{aligned} \lambda_i(x,r) &:= -\frac{1}{r} \frac{d\psi_i}{dt}(g_i(x)) \qquad \text{als } g_i(x) \leq 0 \\ &:= 0 \qquad\qquad\qquad\quad \text{"} \quad g_i(x) > 0 \end{aligned} \tag{3.4.65}$$

welke voor de kwadratische boetefunctie resulteert in

$$\begin{aligned} \lambda_i(x,r) &= -\frac{k_i}{r} g_i(x) \qquad \text{als } g_i(x) \leq 0 \\ &:= 0 \qquad\qquad\quad \text{"} \quad g_i(x) > 0 \end{aligned} \tag{3.4.66}$$

Voor de aldus gedefinieerde functie $\lambda_i(x,r)$ geldt als x_k het minimum is van de boetefunctie $P(x,r_k)$ (3.4.60) dat

$$\lambda_i(x_k,r_k) = \lambda_{k,i} \tag{3.4.67}$$

waar $\lambda_{k,i}$ gelijk is aan de i-de component van de boven gedefinieerde Lagrange parameter. Met de definitie (3.4.65) van de functies $\lambda_i(x,r)$ kan de gradiënt van de boetefunctie (3.4.47) worden weergegeven door

$$\nabla P(x,r) := \nabla f(x) - \sum_{i=1}^{m} \lambda_i(x,r)\, \nabla g_i(x) \tag{3.4.68}$$

3.4.16. Analoog aan het voorgaande geldt in het minimum x_k van de barrièrefunctie

$$B(x,r_k) := f(x) + r_k \sum_{i=1}^{m} \varphi_i(g_i(x)) \tag{3.4.69}$$

dat de gradiënt gelijk wordt aan nul

$$\nabla B(x_k,r_k) := \nabla f(x_k) + r_k \sum_{i=1}^{m} \frac{d\varphi_i}{dt}(g_i(x_k))\nabla g_i(x_k) = 0 \tag{3.4.70}$$

Dit resultaat is op zijn beurt vergelijkbaar met de noodzakelijke voorwaarde voor het minimum in x_k van het beperkte minimaliseringsprobleem

$$\min \{f(x) \mid g_i(x) \geq g_i(x_k) > 0 \ , \ i=1,\ldots m\} \tag{3.4.71}$$

welke luidt

$$\nabla f(x_k) - \sum_{i=1}^{m} \lambda_{k,i}\, \nabla g_i(x_k) = 0 \tag{3.4.72}$$

waar $\lambda_{k,i}$ de i-de component is van de Lagrange multiplicator waarvoor geldt

$$\lambda_{k,i} \geq 0 \qquad i=1,\ldots,m \tag{3.4.73}$$

Vergelijking van de uitdrukkingen (3.4.70) en (3.4.72) leidt in dit geval tot de aan (3.4.65) analoge definitie van de functie

$$\lambda_i(x,r) := - r \frac{d\varphi_i}{dt} (g_i(x)) \tag{3.4.74}$$

welke definitie voor de logarithmische barrièrefunctie resulteert in

$$\lambda_i(x,r) := \frac{k_i r}{g_i(x)} \tag{3.4.75}$$

voor de inverse barrièrefunctie in

$$\lambda_i(x,r) := \frac{k_i r}{(g_i(x))^2} \tag{3.4.76}$$

en voor de kwadratische inverse functie in

$$\lambda_i(x,r) := \frac{k_i r}{2(g_i(x))^3} \tag{3.4.77}$$

Juist als in het voorgaande geldt dat als x_k het minimum is van de barrièrefunctie $B(x,r_k)$ dan geldt dat

$$\lambda_i(x_k,r_k) = \lambda_{k,i} \tag{3.4.78}$$

waar $\lambda_{k,i}$ gelijk is aan de i-de component van de boven gedefinieerde Lagrange vector (3.4.73). Voor de gradiënt van de barrièrefunctie kan met de definitie van de functie $\lambda_i(x,r)$ worden geschreven

$$\nabla B(x,r) := \nabla f(x) - \sum_{i=1}^{m} \lambda_i(x,r) \nabla g_i(x) \tag{3.4.79}$$

3.4.17. Het originele probleem met uitsluitend ongelijkheidsvoorwaarden (3.4.9)

$$\min \{f(x) \mid g_i(x) \geq 0 \quad , i=1,\ldots,m, \; x \in \mathbb{R}^n\}$$

kan ook worden opgevat als de primale formulering van het zadelpunts-probleem

$$\min_{x} \{ \max_{\lambda_i} \{f(x) - \sum_{i=1}^{m} \lambda_i g_i(x) \mid \lambda_i \geq 0,\ i=1,\ldots,m\} \mid x \in \mathbb{R}^n\} \qquad (3.4.80)$$

De corresponderende duale formulering van dit probleem luidt:

$$\max_{\lambda_i} \{\min_{x} \{ f(x) - \sum_{i=1}^{m} \lambda_i g_i(x) \mid x \in \mathbb{R}^n\} \mid \lambda_i \geq 0,\ i=1,\ldots,m\} \qquad (3.4.81)$$

of, equivalent in het geval zowel $f(x)$ als $-g_i(x)$, $i=1,\ldots m$ convex zijn, (d.i. in het geval (3.4.9) een convex programmeringsprobleem is)

$$\max_{\lambda_i} \{f(x) - \sum_{i=1}^{m} \lambda_i g_i(x) \mid \nabla f(x) - \sum_{i=1}^{m} \lambda_i \nabla g_i(x) = 0, \lambda_i \geq 0, i=1,\ldots m\} \qquad (3.4.82)$$

Die paren van variabelen (x,λ), waar λ de m-vector van Lagrangeparameters λ_i is, die voldoen aan de nevenvoorwaarden van dit laatste, duale probleem worden <u>duaal toegelaten oplossingen</u> genoemd. Vergelijking van de voorwaarde in (3.4.82) met de voorwaarden (3.4.61) en (3.4.70) voor het voor gegeven r_k door het paar (x_k,r_k) gekarakteriseerde minimum van de boetefunctie $P(x,r)$, respectievelijk de barrièrefunctie $B(x,r)$, leert dat deze paren (x_k,λ_k) juist duaal toegelaten oplossingen zijn van het originele probleem. Als de parameter r_k nu naar 0 gaat, convergeren volgens Stelling 3.4.9 en Stelling 3.4.11 de minima x_k van de boetefuncties $P(x,r_k)$, resp. barrièrefunctie $B(x,r_k)$ naar de oplossing $\hat{x}$ van het originele probleem. In het geval deze oplossing $\hat{x}$ nu een regulier punt is en bovendien wordt aan de tweede orde voldoende voorwaarde (3.1.51) voor een minimum dan convergeren ook de Lagrange-vectoren $\lambda_k := \lambda(x_k,r_k)$ naar de Lagrange vector van het originele probleem. Dit resultaat is weergegeven in de volgende stelling :

<u>Stelling 3.4.17:</u> Wordt een boete-resp. barrièrefunctiemethode gebruikt voor het oplossen van het minimaliseringsprobleem (3.4.9) met een boete, resp. barrièrefunctie gedefinieerd door (3.4.12), resp.(3.4.19) dan geldt als de met de naar 0 convergerende rij $\{r_k\}$ corresponderende

rij $\{x_k\}$ convergeert naar de oplossing $\hat{x}$ van het originele probleem en deze oplossing een regulier punt is waar voldaan wordt aan de tweede orde voldoende voorwaarden voor een minimum (3.1.51), dat ook de corresponderende rij $\{\lambda(x_k,r_k)\}$ convergeert naar de Lagrangevector $\hat{\lambda}$ die behoort bij de oplossing van het originele probleem d.i. in de terminologie van de Stellingen 3.4.9 en 3.4.11

$$\lim_{k\in K} \lambda(x_k,r_k) = \lim_{k\in K} \lambda_k = \hat{\lambda} \tag{3.4.83}$$

waar $\hat{\lambda} = [\hat{\lambda}_1,\hat{\lambda}_2 \ldots \hat{\lambda}_m]^T$ voldoet aan

$$\nabla f(\hat{x}) - \sum_{i=1}^{m} \hat{\lambda}_i \nabla g_i(\hat{x}) = 0 \tag{3.4.84}$$

Bewijs: Het bewijs kan worden geleverd met de observatie dat onder de gestelde voorwaarden de vectoren λ_k de Lagrangevectoren zijn van de beperkte minimaliseringsproblemen (3.4.62) resp. (3.4.71): Met de veronderstelde continuïteit van de gradiënten $\nabla g_i(x)$ volgt het resultaat daarmee onmiddellijk. □

Een soort gelijk resultaat kan ook worden bewezen voor het algemenere minimaliseringsprobleem (3.4.25) van het type GNLI waarin naast ongelijkheden ook gelijkheden voorkomen en waarbij gebruik moet worden gemaakt van de gemengde boetefunctie (3.4.26).

Hessiaan van boete- en barrièrefuncties

3.4.18. Van de boetefunctie (3.4.12)

$$P(x,r) = f(x) + \frac{1}{r} \sum_{i\in I_V(x)} \psi_i(g_i(x))$$

met als gradiënt ((3.4.61), (3.4.68))

$$\nabla P(x,r) = \nabla f(x) + \frac{1}{r} \sum_{i\in I_V(x)} \frac{d\psi_i}{dt}(g_i(x)) \nabla g_i(x)$$

$$= \nabla f(x) - \sum_{i=1}^{m} \lambda_i(x,r) \nabla g_i(x)$$

wordt de Hessiaan gegeven door

$$\nabla^2 P(x,r) = G(x) - \sum_{i=1}^{m} \lambda_i(x,r) G_i(x)$$

$$+ \frac{1}{r} \sum_{i \in I_V(x)} \frac{d^2\psi_i}{dt^2}(g_i(x)) \nabla g_i(x) \nabla^T g_i(x) \qquad (3.4.85)$$

De eerste twee matrixtermen in deze uitdrukking vormen een benadering voor de Hessiaan van de Lagrangefunctie (vgl(3.1.10)) van het originele minimaliseringsprobleem (3.4.9). De laatste matrixterm is de som van een aantal rang-één-matrices met coëfficiënten die groter worden naarmate r kleiner wordt. Bijvoorbeeld, in het geval van de kwadratische verliesfunctie (3.4.7) geldt dat deze coëfficiënten gelijk zijn aan

$$\frac{1}{r} \frac{d^2\psi_i}{dt^2} = \frac{k_i}{r} \qquad (3.4.86)$$

waarmee

$$\nabla^2 P(x,r) := \nabla^2 L(x,\lambda(x,r)) + \sum_{i=1}^{m} \frac{k_i}{r} \nabla g_i(x) \nabla g_i^T(x) \qquad (3.4.87)$$

De rang één matrices $\nabla g_i(x) \nabla g_i^T(x)$ hebben als nulruimte het orthogonale complement van de deelruimte opgespannen door de vectoren $\nabla g_i(x), i \in I_V(x)$. Wordt een boetefunctiemethode gebruikt voor de oplossing van het originele probleem met behulp van de hier beschouwde algemene boetefunctie dan zullen de opvolgende eigenwaarden van de Hessianen van de boetefuncties in de minima x_k van de naar 0 convergerende parameters r_k deels convergeren naar de eigenwaarden bij de Hessiaan van de Lagrangefunctie van het originele probleem (met eigenvectoren in het orthogonale complement M(x) (3.1.23) van de deelruimte opgespannen door de actieve normalen in het optimale punt) en deels convergeren naar oneindig (met eigenvectoren in de deelruimte opgespannen door de actieve normalen in het optimale punt). De voor de convergentie van onbeperkte minimaliseringsproblemen belangrijke conditie van deze Hessianen, d.i. de verhouding van grootste tot kleinste eigenwaarde neemt toe naarmate de parameters r_k dichter naar 0 naderen. De met de parameters r_k corresponderende onbeperkte minimaliseringsproblemen worden daarmee

toenemend slechter geconditioneerd naarmate r_k kleiner wordt. Het resultaat is een toenemend slechtere convergentie van alle op het gebruik van de gradiënt gebaseerde minimaliserings algorithmen.

3.4.19. Voor de Hessiaan van de barrièrefunctie (3.4.19)

$$B(x,r) = f(x) + r \sum_{i\in I_p(x)} \varphi_i(g_i(x))$$

geldt nagenoeg hetzelfde resultaat. De gradiënt in dit geval (vgl(3.4.52), (3.4.79)

$$\nabla B(x,r) = \nabla f(x) + r \sum_{i\in I_p(x)} \left.\frac{d\varphi_i}{dt}\right|_{t=g_i(x)} \nabla g_i(x)$$

$$= \nabla f(x) - \sum_{i\in I_p(x)} \lambda_i(x,r)\, \nabla g_i(x)$$

geeft bij differentiatie als Hessiaan

$$\nabla^2 B(x,r) = G(x) - \sum_{i\in I_p(x)} \lambda_i(x,r)\, G_i(x) + r \sum_{i\in I_p(x)} \left.\frac{d^2\varphi_i}{dt^2}\right|_{t=g_i(x)} \nabla g_i(x)\nabla^T g_i(x) \qquad (3.4.88)$$

Ook hier vormen de eerste twee matrixtermen weer een benadering voor de Hessiaan van de Lagrangefunctie van het originele minimaliseringsprobleem. De laatste matrixterm is weer een som van rang-één-matrices met coëfficiënten die afhankelijk zijn van r en $g_i(x)$. Voor de drie besproken barrièrefunctietermen (3.4.22), (3.4.20) en (3.4.21) geldt, respectievelijk

$$r\frac{d^2\varphi_i}{dt^2} = r\frac{k_i}{t^2} = \frac{k_i r}{g_i^2(x)} \qquad (3.4.89a)$$

$$r\frac{d^2\varphi_i}{dt^2} = 2r\frac{k_i}{t^3} = \frac{2k_i r}{g_i^3(x)} \qquad (3.4.89b)$$

en

$$r \frac{d^2\varphi_i}{dt^2} = 6r \frac{k_i}{t^4} = \frac{6k_i r}{g_i^4(x)} \qquad (3.4.89c)$$

Evaluatie van deze coëfficiënten in de minima x_k van de barrièrefuncties corresponderend met de monotoon naar 0 convergerende parameters r_k resulteert als de betreffende i-de beperking actief is in de oplossing $\hat{x}$, in een naar boven onbegrensde rij van coëfficiënt waarden, ofwel, anders gezegd

$$\lim_{k \in K} r_k \frac{d^2\varphi_i}{dt^2} (g_i(x_k)) = \infty \qquad (3.4.90)$$

Toepassing van een barrièrefunctiemethode voor de oplossing van het originele probleem (3.4.9) met een van de besproken barrièrefuncties resulteert juist als de toepassing van boetefunctie methoden in onbeperkte minimaliseringsproblemen waarbij de Hessiaan van de objectfunctie in toenemende mate slecht geconditioneerd wordt met als gevolg een toenemend slechtere convergentie van de eventueel op de gradiënten gebaseerde minimaliseringsalgorithmen.

Extrapolatie

3.4.20. Een niet ongebruikelijke procedure bij de practische toepassing van boete- en barrièrefunctie methoden is dat de successievelijke onbeperkte minimaliseringsproblemen die corresponderen met afnemende waarde van de parameter r_k steeds uitgaan van het in het voorgaande optimaliseringsprobleem (corresponderend met parameter r_{k-1}) bepaalde minimum x_{k-1} als startpunt. Nodig is dit niet en betere beginschattingen zijn o.a. mogelijk door gebruik te maken van extrapolatie. Het idee daarbij is dat de minima van boete- en barrièrefuncties corresponderend met gegeven r-waarden worden opgevat als een (vector-)functie van de parameter r. Voor ieder van de n-componenten van de minima wordt daartoe extrapolatie-polynoom opgesteld

$$x_i(r) = A_{0,i} + A_{1,i} r + A_{2,i} r^2 + \ldots\ldots \qquad (3.4.91)$$

Evaluatie van een aantal minima corresponderend met een aantal r-waarden biedt de mogelijkheid deze extrapolatiepolynomen (of beter de coëfficiënt-vectoren ervan) te bepalen. Een beter startpunt, resp. een betere schatting voor het minimum van het originele probleem kan met deze polynomen worden

bepaald door substitutie van resp. $r := r_{k+1}$ en $r := 0$. In de practijk (vgl. [3.4.18]) blijkt deze extrapolatie een belangrijke versnelling van het convergentieproces van boete- en barrièrefunctiemethoden teweeg te kunnen brengen.

Niet-parametrische boete- en barrièrefunctiemethoden

3.4.21. Niet-parametrische boete- en barrièrefuncties verschillen van de hiervoor besproken parametrische boete- en barrièrefuncties daarin dat de opvolgende onbeperkte minimaliseringsproblemen die in de plaats komen van het originele beperkte minimaliseringsprobleem niet afhankelijk zijn van een in principe vooraf vastgelegde naar 0 convergerende rij van parameters. In plaats daarvan, en in tegenstelling tot wat de naam doet vermoeden, worden de opvolgende onbeperkte minimaliseringsproblemen bepaald door een rij van (evengoed als parameters op te vatten) truncatieniveau's (eng: truncation levels), waarvan de waarde in ieder nieuw onbeperkt minimaliseringsprobleem wordt afgeleid uit het resultaat van het voorgaande onbeperkt minimaliseringsprobleem. Uitgangspunt voor de niet-parametrische boetefunctiemethoden (de niet-parametrische barrièrefunctiemethoden worden besproken in par 3.4.24) is de vervanging van het beperkt minimaliseringsprobleem (3.4.9)

$$\min \{f(x) \mid g_i(x) \geq 0 \quad , i=1,\ldots m\}$$

door een rij onbeperkte minimaliserings problemen van de gedaante

$$\min \{P^*(x,t_k) \mid x \in \mathbf{R}^n\} \tag{3.4.92}$$

waarin in de veronderstelling dat

$$t_k \leq f(\hat{x}) \tag{3.4.93}$$

de niet-parametrische boetefunctie $P^*(x,t)$ wordt gegeven door

$$P^*(x,t) := \psi_0((f(x)-t)) + \sum_{i=1}^{m} \psi_i(g_i^-(x)) \tag{3.4.94}$$

waarin $\psi_0(y)$ een reëelwaardige functie is van één variabele met dezelfde eigenschappen als de eerder gedefinieerde (vgl(3.4.5)) functies $\psi_i(y)$.
In het geval van de kwadratische verliesfunctie (3.4.7) resulteert bijvoorbeeld

$$P^*(x,t) := \tfrac{1}{2}k_0(f(x)-t)^2 + \tfrac{1}{2}\sum_{i=1}^{m} k_i(g_i^-(x))^2 \tag{3.4.95}$$

Voor de opvolgende minima $x_k := x(t_k)$ van deze laatste niet-parametrische boetefunctie (3.4.95) gelden analoog aan de situatie bij de parametrische boetefuncties (vgl. Lemma 3.4.8) enkele ongelijkheden die een rol spelen bij de convergentie van de niet-parametrische boete functiemethoden. Deze zijn gegeven in het volgende lemma:

Lemma 3.4.21 (vgl. [3.4.22]): Als $x_k := x(t_k)$ het minimum is van het onbeperkt minimaliseringsprobleem

$$\min \{\tfrac{1}{2}k_0(f(x)-t_k)^2 + \frac{1}{2}\sum_{i=1}^{m} k_i(g_i^-(x))^2 \mid x \in \mathbb{R}^n\} \tag{3.4.96}$$

en $\hat{x}$ de oplossing is van het corresponderende originele probleem (3.4.9) dan geldt

$$\text{als } t_k \leq f(\hat{x}) \quad : \quad f(x_k) \leq f(\hat{x}) \text{ en} \tag{3.4.97}$$

$$\text{als } t_k < t_{k+1} \quad : \quad f(x_k) \leq f(x_{k+1}) \tag{3.4.98}$$

Bovendien geldt als het originele probleem (3.4.9) een convex minimaliseringsprobleem is dat

$$t_k \leq f(x_k) \leq f(\hat{x})$$

Bewijs: Als x_k het minimum is van (3.4.96) dan geldt

$$\tfrac{1}{2}k_0(f(x_k)-t_k)^2 + \frac{1}{2}\sum_{i=1}^{m} k_i(g_i^-(x_k))^2 \leq \tfrac{1}{2}k_0(f(\hat{x})-t_k)^2 + 0$$

waaruit onmiddellijk volgt dat

$$(f(x_k)-t_k)^2 \leq (f(\hat{x}) - t_k)^2$$

zodat of

$$f(x_k) - t_k \leq f(\hat{x}) - t_k \qquad \text{als } f(x_k) \geq t_k$$

of

$$t_k - f(x_k) \leq f(\hat{x}) - t_k \qquad \text{als } f(x_k) \leq t_k$$

In het eerste geval volgt de ongelijkheid (3.4.97) onmiddellijk in het tweede geval volgt hetzelfde resultaat uit

$$f(x_k) \leq t_k \text{ en } t_k \leq f(\hat{x})$$

Als x_k en x_{k+1} respectievelijk de minima zijn van de boetefunctie corresponderend met de truncatieniveau's t_k en t_{k+1} dan geldt

$$\tfrac{1}{2}k_0(f(x_k)-t_k)^2 + \tfrac{1}{2}\sum_{i=1}^{m} k_i(\bar{g}_i(x_k))^2 \le \tfrac{1}{2}k_0(f(x_{k+1})-t_k)^2 + \tfrac{1}{2}\sum_{i=1}^{m} k_i(\bar{g}_i(x_{k+1}))^2$$

$$\tfrac{1}{2}k_0(f(x_{k+1})-t_{k+1})^2 + \tfrac{1}{2}\sum_{i=1}^{m} k_i(\bar{g}_i(x_{k+1}))^2 \le \tfrac{1}{2}k_0(f(x_k)-t_{k+1})^2 + \tfrac{1}{2}\sum_{i=1}^{m} k_i(\bar{g}_i(x_k))^2$$

waaruit na optelling volgt

$$t_k f(x_k) + t_{k+1}f(x_{k+1}) \ge t_k f(x_{k+1}) + t_{k+1}f(x_k)$$

of

$$(t_{k+1}-t_k)f(x_{k+1}) \ge (t_{k+1}-t_k)f(x_k)$$

zodat

$$f(x_{k+1}) \ge f(x_k)$$

Als $f(x)$ en $-g_i(x)$, $i=1,\ldots m$, convexe functies zijn dan volgt uit

$$\tfrac{1}{2}k_0(f(x_k)-t_k)^2 + \tfrac{1}{2}\sum_{i=1}^{m} (\bar{g}_i(x_k))^2 \le \tfrac{1}{2}k_0(f(\hat{x})-t_k)^2$$

dat

$$f(x_k) \ge t_k$$

Stel namelijk het tegenovergestelde dan gold dat er op de verbindingslijn $x(\alpha) := x_k + \alpha(\hat{x} - x_k)$ voor $0 < \alpha < 1$ een punt x' bevond waar

$$f(x') - t_k = 0$$

Als $\bar{g}_i(x_k) \neq 0$ dan geldt in het punt x' op grond van de concaviteit van $\bar{g}_i(x)$ dat

$$\bar{g}_i(x_k) \le \bar{g}_i(x') \le \bar{g}_i(\hat{x}) = 0$$

hetgeen zou impliceren dat

$$P^*(x', t_k) \le P^*(x_k, t_k)$$

Dit is tegenspraak met de definitie van x_k. De veronderstelling dat $f(x_k) < t_k$ is dus onjuist. □

.22. Op grond van de in Lemma 3.4.21 gegeven ongelijkheden ligt het voor de hand dat, indien telkens voor het volgende truncatieniveau waarden worden gekozen of gegenereerd die voldoen aan de voorwaarde

$$t_k < t_{k+1} \leq \hat{f} = f(\hat{x}) \tag{3.4.99}$$

in dat geval juist als bij de parametrische boetefuncties een rij oplossingen $\{x_k\}$ van het onbeperkte minimaliseringsprobleem wordt gegenereerd die convergeert naar de oplossing van het originele probleem (3.4.9)

$$\lim_{k\to\infty} x_k = \hat{x} \tag{3.4.100}$$

In het geval het **originele** probleem een convex programmeringsprobleem is zal omdat in ieder minimum x_k geldt

$$k_0(f(x_k)-t_k)\nabla f(x_k) + \sum_{i\in I_V(x_k)} k_i g_i(x_k)\nabla g_i(x_k) = 0 \tag{3.4.101}$$

onder de gebruikelijke regulariteits condities ook gelden dat(vgl. Stelling 3.4.17)

$$\lim_{k\to\infty} - \frac{k_i g_i(x_k)}{k_0(f(x_k)-t_k)} = \hat{\lambda}_i \tag{3.4.102}$$

Voor het bewijs van deze en soortgelijke uitspraken kan o.a. worden verwezen naar [3.4.20].
Opgemerkt kan worden dat het convergentieproces van niet-parametrische boetefunctiemethoden in meerdere opzichten equivalent is aan het convergentieproces van de corresponderende parametrische boetefunctiemethoden met parameterwaarden (vgl(3.4.66)) gelijk aan

$$r_k := k_0(f(x_k)-t_k) \tag{3.4.103}$$

Een discussie over deze equivalentie kan worden gevonden in [3.4.11].

.4.23. Essentieel voor de convergentie van niet-parametrische boetefuncties is de manier waarop het volgende truncatie niveau t_{k+1} wordt bepaald uit de resultaten gevonden bij het voorgaande (k-de) onbeperkt minimaliseringsprobleem. Een aantal generatie formules zijn hiervoor gesuggereerd.
Tot de bekendste behoren (vgl[3.4.20])

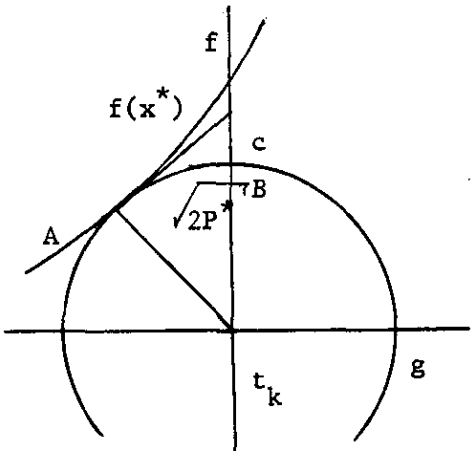
i) $$t_{k+1} := f(x_k) \tag{3.4.104}$$

ii) $$t_{k+1} := t_k + \sqrt{2P^*(x_k, t_k)} \tag{3.4.105}$$

iii) $$t_{k+1} := t_k + \frac{2P^*(x_k, t_k)}{f(x_k)-t_k} \tag{3.4.106}$$

en iv) $$t_{k+1} := t_k + \frac{P^*(x_k, t_k)}{f(x_k)-t_k} \tag{3.4.107}$$

De eerste van deze generatieformules werd o.a. gebruikt door Lootsma (vgl [3.4.17]) die daarmee minder gunstige convergentieeigenschappen constateerde. De tweede en derde generatieformule werden gesuggereerd door respectievelijk Morrison [3.4.22] en Wolfe [3.4.15]. De relatie tussen deze twee onderling als ook hun relatie met de eerste generatieformule kunnen worden geïllustreerd aan de hand van een schets voor het geval van één nevenvoorwaarde (zie Figuur 3.4.22). In deze schets is langs de verticale as aangegeven de functiewaarde f(x) en langs de horizontale as de functiewaarde g(x) voor dezelfde x. De oplossing van het kleinste kwadratenprobleem (3.4.96) is het punt A. De waarde voor t_{k+1} gesuggereerd door Morrison (3.4.105) is het punt B en de waarde voor t_{k+1} gesuggereerd door Wolfe (3.4.106) is het punt C. Deze laatste parameters worden in de literatuur ook wel aangeduid als de "tangent" parameters. De generatieformule (iv):(3.4.107) werd gepropageerd door Staha [3.4.20], die met de daarvan voor algemenere toepassingen afgeleide generatieformule



Figuur 3.4.22: Geometrie van de generatieformules

$$t_{k+1} := t_k + \max\left\{(f(x_k)-t_k), \frac{P^*(x_k, t_k)}{f(x_k)-t_k}\right\} \tag{3.4.108}$$

goede convergentieresultaten kon rapporteren.

3.4.24. Niet-parametrische barrièrefunctie-methoden geven aanleiding tot dezelfde of analoge opmerkingen als hiervoor gemaakt voor de niet-parametrische boetefunctiemethoden. Daarnaast gelden nog een aantal speciale eigenschappen die specifiek zijn voor het barrièrefunctiekarakter: Juist als bij de parametrische barrièrefunctiemethoden geldt dat deze alleen kunnen worden toegepast bij minimaliseringsproblemen met uitsluitend ongelijkheidsbeperkingen (3.4.9)

$$\min\{ f(x) \mid g_i(x) \geq 0,\ i=1,\ldots,m\}$$

en een toegelaten gebied met een niet-leeg inwendige (vgl(3.4.15))

$$S^0 := \{x \mid g_i(x) > 0,\ i=1,\ldots,m\} \neq \emptyset$$

Is daaraan voldaan dan kan het beperkte minimaliseringsprobleem worden vervangen door een rij van de truncatieniveau's t_k afhankelijke onbeperkte minimaliseringsproblemen van de vorm

$$\min \{B^*(x,t_k) \mid x \in S^0\} \tag{3.4.109}$$

waarin

$$B^*(x,t) := k_0\varphi_0(t - f(x)) + \sum_{i=1}^{m} k_i\varphi_i(g_i(x)) \tag{3.4.110}$$

waarin $\varphi_0(y)$ een reëel waardige functie van één variabele is met dezelfde eigenschappen als de eerder gedefinieerde (vgl(3.4.18)) functies $\varphi_i(y)$. In tegenstelling tot de situatie bij niet-parametrische boetefuncties (vgl(3.4.93)) moet het truncatieniveau hier zo gekozen worden dat

$$t_k \geq f(\hat{x}) \tag{3.4.111}$$

Wordt voor de functies $\varphi_i(y)$, $i=0,\ldots,m$, de logarithmische (orde 1) functie (3.4.22) gekozen, dan krijgt de functie $B^*(x,t)$ (3.4.110) de gedaante

$$B^*(x,t) = - k_0 \ln(t-f(x)) - \sum_{i=1}^{m} k_i \ln(g_i(x)) \tag{3.4.112}$$

Wordt voor $\varphi_i(y)$, $i=0,\ldots,m$ de inverse (orde 2) functie (3.4.20) gekozen, dan wordt $B^*(x,t)$ (3.4.110) gelijk aan

$$B^*(x,t) = \frac{k_0}{t-f(x)} + \sum_{i=1}^{m} \frac{k_i}{g_i(x)} \tag{3.4.113}$$

De eerste van deze barrièrefuncties werd o.a. bestudeerd door Lootsma [3.4.17], de tweede o.a. door Fiacco en McCormick [3.4.11], in beide gevallen met de generatieformule voor het truncatieniveau

$$t_{k+1} := f(x_k) \tag{3.4.114}$$

3.4.25. Analoog aan de situatie in het geval van de parametrische boetefuncties kan worden aangetoond dat de minima x_k van de met afnemende t_k-waarden corresponderende onbeperkte minimaliserings problemen (3.4.109) convergeren naar de oplossing $\hat{x}$ van het originele beperkte probleem (3.4.9)

$$\lim_{k\to\infty} x_k = \hat{x} \tag{3.4.115}$$

Bovendien geldt in het geval van convexe programmeringsproblemen corresponderend met de niet-parametrische barrièrefuncties (3.4.112) en (3.4.113) respectievelijk

$$\lim_{k\to\infty} \left(\frac{k_i}{k_0}\right) \frac{t_k - f(x_k)}{g_i(x_k)} = \hat{\lambda}_i \tag{3.4.116}$$

en

$$\lim_{k\to\infty} \frac{k_i}{k_0} \frac{(t_k - f(x_k))^2}{g_i(x_k)} = \hat{\lambda}_i \tag{3.4.117}$$

Het convergentiegedrag van de niet-parametrische barrièrefunctiemethoden is weer equivalent aan het convergentiegedrag van de corresponderende parametrische barrièrefunctiemethoden met opvolgende parameterwaarden gelijk aan respectievelijk (vgl(3.4.103))

$$r_k := (t_k - f(x_k))/k_0 \tag{3.4.118}$$

en

$$r_k := (t_k - f(x_k))^2/k_0 \tag{3.4.119}$$

Voor verdere details m.b.t. de convergentie van deze methoden zij in het bijzonder verwezen naar de publicaties van Lootsma [3.4.17] en Fiacco en Mc Cormick [3.4.11]. Opgemerkt kan worden dat geen van deze auteurs zich positief uitspreken over de practische convergentieeigenschappen van de niet-parametrische boete- en barrièrefunctiemethoden.

Methode van Huard

3.4.26. Een speciale niet-parametrische barrièrefunctiemethode welke als eerste gepubliceerde methode van deze categorie veel bekendheid heeft gekregen in de literatuur is de "Centra methode" (eng: Method of Centers) van Huard [3.4.13]. Het idee achter deze methode is dat het onbeperkte minimaliseringsprobleem (3.4.9)

$$\min \{f(x) \mid g_i(x) \geq 0,\ i=1,\ldots,m\}$$

met een toegelaten gebied met een niet-leeg inwendige (3.4.15) wordt vervangen door een rij (onbeperkte) maximaliseringsproblemen van de vorm

$$\max \{\theta(x,S_k) \mid x \in S_k^0 \subset \mathbb{R}^n\} \tag{3.4.120}$$

waarin $\theta(x,S_k)$ een afstandsfunctie is die gedefinieerd is op het toegelaten gebied

$$S_k := \{x \in \mathbb{R}^n \mid f(x) \leq f(x_k),\ g_i(x) \geq 0,\ i=1,\ldots,m\} \tag{3.4.121}$$

en wel zo dat

$$\begin{aligned} \theta(x,S_k) &> 0 \quad \text{als } x \in S_k^0 \\ &= 0 \quad \text{als } x \in S_k \backslash S_k^0 (=\text{rand van } S_k) \end{aligned} \tag{3.4.122}$$

De maxima x_k van het maximaliseringsprobleem (3.4.120) worden door Huard "centra" genoemd omdat zij kunnen worden opgevat als de punten met de grootste "afstand" tot de rand. Aangezien deze centra in het inwendige liggen zijn de maximaliseringsproblemen (3.4.120) in de practijk op te vatten als onbeperkte maximaliseringsproblemen. Voor de afstandsfunctie $\theta(x,S_k)$ suggereerde Huard twee keuzen

$$\theta(x,S_k) := \min [f(x_k)-f(x),\ g_i(x),\ i=1,\ldots,m] \tag{3.4.123}$$

en

$$\theta(x,S_k) := (f(x_k)-f(x)) \prod_{i=1}^{m} g_i(x) \tag{3.4.124}$$

De eerste van deze functies is niet-differentieerbaar, hetgeen o.a. impliceert dat de meest gebruikelijke onbeperkte minimaliseringsalgorithmen niet kunnen worden toegepast. In plaats daarvan suggereerde Huard dan ook

het gebruik van locale linearisatie en de oplossing van het gelineariseerde probleem met lineaire programmeringstechnieken. Een dergelijke speciale behandeling is niet nodig voor de maximalisering van de tweede functie (3.4.124), die kan worden behandeld op dezelfde manier als alle voorgaande niet-parametrische boete- en barrièrefunctiemethoden.

(Onbeperkte) minimalisering van boete- en barrièrefuncties

3.4.27. Gezien het speciale karakter van de in pt. 3.4.18 en 3.4.19 besproken Hessiaan van boete- en barrièrefuncties waarvan een aantal (gelijk aan het aantal actieve beperkingen) eigenwaarden naar ∞ gaan als de parameters r_k naar 0 gaan (vgl. (3.4.87) en (3.4.88)) ligt het vanuit theoretisch standpunt voor de hand dat voor de minimalisering van boete- en barrièrefuncties het best gebruik gemaakt kan worden van algorithmen waarvan de convergentie snelheid niet afhankelijk is van de conditie van de Hessiaan. Een dergelijk algorithme is de in paragraaf 2.5 besproken methode van Newton die kwadratische convergentieeigenschappen bezit, onafhankelijk van de conditie van de Hessiaan. Dit theoretisch resultaat werd zowel door Fiacco & Mc Cormick [3.4.7] als Lootsma [3.4.17] (en vele anderen) experimenteel geverifieerd. Daarnaast bleken ook met de nauw aan de methode van Newton gerelateerde quasi-Newton met methoden met in het bijzonder de BFS-algorithme (vgl pt 2.8.11) goede convergentie resultaten te kunnen worden bereikt. In het door hem ontwikkelde ALGOL-60 programma MINIFUN (dat ook op het THE-Rekencentrum aanwezig is(zie [3.4.24]) en [3.4.18] waarin niet-lineaire minimaliseringsproblemen met nevenvoorwaarden worden opgelost met behulp van een (gemengde) boete- en barrièrefunctie aanpak geeft Lootsma de gebruiker de keuze tussen de Newton algorithme (met de modificatie van Fiacco en Mc Cormick (pt 2.5.20)) en de BFS-algorithme afhankelijk van het eenvoudig beschikbaar zijn van tweede orde informatie van de betreffende probleemfuncties. Naast deze algemene algorithmen zijn door diverse auteurs verschillende suggesties gedaan en methoden ontwikkeld speciaal bestemd voor het minimaliseren van boetefuncties. Deze methoden maken vrijwel allemaal gebruik van de kennis van de speciale structuur van de Hessiaan. Voor een aantal interessante suggesties in dit verband kan in het bijzonder worden verwezen naar Luenberger [3.4.1].

Referenties

3.4.28. Voor meer details over de in deze paragraaf besproken boete- en barrière-functie methoden kan worden verwezen naar de volgende publikaties

[3.4.1] : Zie [1.1.1] Luenberger (1973)

[3.4.2] : Zie [1.1.2] Jacoby, Kowalik & Pizzo (1972

[3.4.3] : Zie [1.1.3] Murray (1972)

[3.4.4] : Zie [1.1.4] Gill & Murray (1974)

[3.4.5] : Zie [2.1.3] Zangwill (1969)

[3.4.6] : Zie [2.2.7] Kowalik & Osborne (1968)

[3.4.7] : Zie [2.5.10] Fiacco & Mc Cormick (1968)

[3.4.8] : Zie [2.5.14] Lootsma (1972)

[3.4.9] : Zie [2.10.14] Powell (1972)

[3.4.10]: Courant, R: Variational methods for the solution of problems of equilibrium and vibrations, Bull.Am.Math.Soc.49 (1943) pp. 1-23

[3.4.11]: Fiacco, A.V. and Mc Cormick, G.P.: The sequential unconstrained minimization technique (SUMT) without parameters. Operat.Res.15(1967) pp.820-829

[3.4.12]: Fletcher, R and McCann, A.P.: Acceleration Techniques for nonlinear programming, in "Optimization" (R. Fletcher, ed), Academic Press, London (1969)

[3.4.13]: Huard, P: Resolution of mathematical programming by the method of centers, in:"Nonlinear Programming", (J.Abadie, ed.), North-Holland, Publ.Co., Amsterdam (1967)

[3.4.14]: Künzi, H.P. und Oettli, W: Nichtlineare Optimierung: Neuere Verfahren, Bibliographie, Lecture Notes in Operations Research and Mathematical Systems nr. 16, Springer Verlag, Berlin (1969)

[3.4.15]: Kowalik, J, Osborne, M.R. and D.M. Ryan: A new method for constrained optimization problems, Operat.Res, 17 (1969) pp. 973-983

[3.4.16] : Lootsma, F.A.: Hessian matrices of penalty functions for solving constrained-optimization problems Philips Res.Repts 24 (1969) pp 322-330

[3.4.17] : Lootsma, F.A.: Boundary properties of penalty functions for constrained optimization, Proefschrift TH Eindhoven,(mei 1970)

[3.4.18] : Lootsma, F.A.: The ALGOL-60-procedure MINIFUN for solving non-linear optimization problems Philips Nat.Lab.Report nr 4761, Eindhoven (1972)

[3.4.19] : Lootsma, F.A.: A survey of methods for solving constrained optimization problems via unconstrained minimization, in "Numerical methods for nonlinear optimization", (F.A. Lootsma, Ed) Academic Press, London (1972)

[3.4.20] : Lootsma, F.A;: Convergence rates of quadratic exterior penalty function methods for solving constrained-minimization problems Philips Res.Rept., 29 (1974) pp.1-12

[3.4.21] : McCormick G.P.: Penalty function versus non-penalty function methods for constrained nonlinear programming problems, Math.Progr, 1(1971) pp.217-238

[3.4.22] : Morrison, D.D.: Optimization by least squares, SIAM J. Numer Anal 5(1968) pp 83-88

[3.4.23] : Ryan, D.M.: Penalty and barrier functions, Ch 6 of [3,4.4] (1974)

[3.4.24] : - : BEATHE-procedure MINIFUN for solving nonlinear optimization problems, THE-RC- Informatie nr 57, Eindhoven (1974)

§ 3.5 Duale methoden

3.5.1. Naast de hiervoor in § 3.2 en § 3.3 besproken primale methoden en de in § 3.4 besproken boete- en barrièrefunctiemethoden is er in de laatste paar jaar (na 1969) een geheel nieuwe categorie van numerieke methoden ontwikkeld die bijzonder bruikbaar is gebleken voor de efficiënte oplossing van minimaliseringsproblemen met niet-lineaire nevenvoorwaarden. Deze categorie van methoden, die in deze syllabus worden aangeduid met de naam duale methoden, wordt gekenmerkt door de omstandigheid dat in meerdere of mindere mate expliciet gebruik wordt gemaakt van de duale probleemformulering van het originele probleem. In het bijzonder wordt bij deze methoden gebruik gemaakt van (schattingen van) de Lagrange multiplicatoren van de beperkte minimaliseringsproblemen. De twee belangrijkste klassen van methoden die tot deze categorie behoren zijn de hieronder na elkaar te bespreken aangevulde-Lagrangefunctie- of multiplicatorenmethoden (eng: augmented Lagrangian-or multipliermethods) en de exacte-boetefunctie-methoden (eng; exact penaltyfunction methods). Vooral aan de eerste van deze beide klassen van methoden wordt in de huidige numerieke niet-lineaire-programmeringsliteratuur veel aandacht besteed.

3.5.2. Het eenvoudigste uitgangspunt voor de bespreking van de theorie van de duale methoden is het standaard minimaliseringsprobleem GNLE (vgl.pt.3.1.4) met uitsluitend (niet-lineaire) gelijkheidsvoorwaarden, d.i. het Lagrangeprobleem (vgl. (3.1.2))

$$\min \{f(x) \mid h_j(x) = 0, \quad j = 1,\ldots,m\} \tag{3.5.1}$$

Uitgangspunt voor de theorie is de veronderstelling dat het optimale punt $\hat{x}$ van dit probleem een regulier punt is waar voldaan wordt aan de voldoende voorwaarden voor het optreden van een minimum (vgl. Stelling 3.1.16), d.w.z. verondersteld wordt dat er een Lagrange(-multiplicatoren-)vector $\hat{\lambda} \in \mathbb{R}^m$ bestaat zodanig dat voldaan wordt aan de relaties

$$\begin{aligned} \nabla_x \mathcal{L}(\hat{x},\hat{\lambda}) &= \nabla_x f(\hat{x}) - N(\hat{x})\hat{\lambda} \\ &= \nabla_x f(\hat{x}) - \sum_{j=1}^{m} \hat{\lambda}_j \nabla h_j(\hat{x}) = 0 \end{aligned} \tag{3.5.2}$$

en

$$\nabla_\lambda \mathcal{L}(\hat{x},\hat{\lambda}) = h(\hat{x}) = 0 \tag{3.5.3}$$

en

$$\forall_{z \in \mathbb{R}^n} : z \neq 0 \wedge N(\hat{x})^T z = 0 \Rightarrow z^T \nabla_{xx} \mathcal{L}(\hat{x}, \hat{\lambda}) z > 0 \qquad (3.5.4)$$

in welke uitdrukkingen

$$\mathcal{L}(x, \lambda) = f(x) - \lambda^T h(x) \qquad (3.5.5)$$

de Lagrange functie is die correspondeert met het originele probleem (3.5.1)

3.5.3. Een belangrijke rol in de theorie van de duale methoden wordt gespeeld door de zg <u>perturbatie</u> of <u>primale functie</u> (zie [3.5.2], [3.5.4] of [3.5.8]) die wordt gedefinieerd door de uitdrukking

$$p(s) := \min_x \{ f(x) \mid h(x) = s,\ x \in \mathbb{R}^n,\ s \in \mathbb{R}^m \} \qquad (3.5.6)$$

De primale functie is het resultaat van de minimalisering van de functie $f(x)$ voor verschillende waarden $s \in \mathbb{R}^m$ voor de nevenvoorwaarden $h(x) = s$. Onder de gebruikelijke (in het voorgaande punt gedeeltelijk genoemde) veronderstellingen kan worden aangetoond dat de primale functie bestaat en continu differentieerbaar is in de directe omgeving van het optimale punt $\hat{x}$ Als $x(s)$ het punt is waarvoor gegeven s het minimum $p(s)$ van het probleem (3.5.6) wordt aangenomen dan geldt in het bijzonder dat

$$p(s) := f(x(s)) \qquad (3.5.7)$$

Het orginele of <u>primale probleem</u> (3.5.1) komt in deze terminologie overeen met het probleem van het bepalen van de waarde van de perturbatie functie in het punt $s = 0$, d.i. de functiewaarde

$$p(0) := f(x(0)) := f(\hat{x}) \qquad (3.5.8)$$

3.5.4. De bijzondere eigenschap van de primale functie die in het navolgende een grote rol speelt is de omstandigheid dat de gradiënt ervan (met betrekking tot de vectorvariabele s) gelijk is aan de Lagrange multiplicatorenvector die correspondeert met het voor de betreffende waarde van s gespecificeerde minimaliseringsprobleem (3.5.6). In formule vorm

$$\nabla p(s) = \lambda(s) \qquad (3.5.9)$$

Deze Lagrange (multiplicatoren)vector $\lambda(s) \in \mathbb{R}^m$ bestaat op grond van de noodzakelijke voorwaarden voor een minimum van het geperturbeerde probleem in het punt x(s) en wordt daardoor gekenmerkt dat voldaan wordt aan de relatie

$$\nabla f(x(s)) - \nabla h(x(s))\ \lambda(s) = 0 \qquad (3.5.10)$$

in welke uitdrukking de notatie $\nabla h(x(s))$ is gebruikt voor de $n \times m$-matrix die de gradiënten van de gelijkheidsbeperkingen als kolommen heeft, d.w.z. de matrix

$$\nabla h(x(s)) := [\nabla h_1(x(s)),\ \nabla h_2(x(s)) \ldots\ \nabla h_m(x(s))] =: N(x(s)) \qquad (3.5.11)$$

3.5.5. De afleiding van de eigenschap (3.5.9) kan worden gebaseerd op de observatie dat het punt $(x(s), \lambda(s), s)$ een stationair punt is met betrekking tot zowel het eerste argument x als het tweede argument λ van de Lagrangefunctie van het geperturbeerde probleem

$$\phi(x,\lambda,s) := f(x) - \lambda^T(h(x)-s) \qquad (3.5.12)$$

Deze functie is gerelateerd met de primale functie p(s) volgens de relatie (vgl. (3.5.7))

$$p(s) = \phi(x(s),\ \lambda(s), s) \qquad (3.5.13)$$

Differentiatie hiervan geeft

$$\nabla p(s) = \nabla_x \phi\left[\frac{dx(s)}{ds}\right] + \nabla_\lambda \phi\ \left[\frac{d\lambda(s)}{ds}\right] + \nabla_s \phi \qquad (3.5.14)$$

welke uitdrukking met

$$\nabla_x \phi = \nabla_\lambda \phi = 0 \qquad \nabla_s \phi = \lambda(s) \qquad (3.5.15)$$

onmiddellijk leidt tot het genoemde resultaat (3.5.9). Een andere, zelfs directere afleiding van hetzelfde volgt bij differentiatie van (3.5.7) onder gebruikmaking van (3.5.10). Dit resulteert in

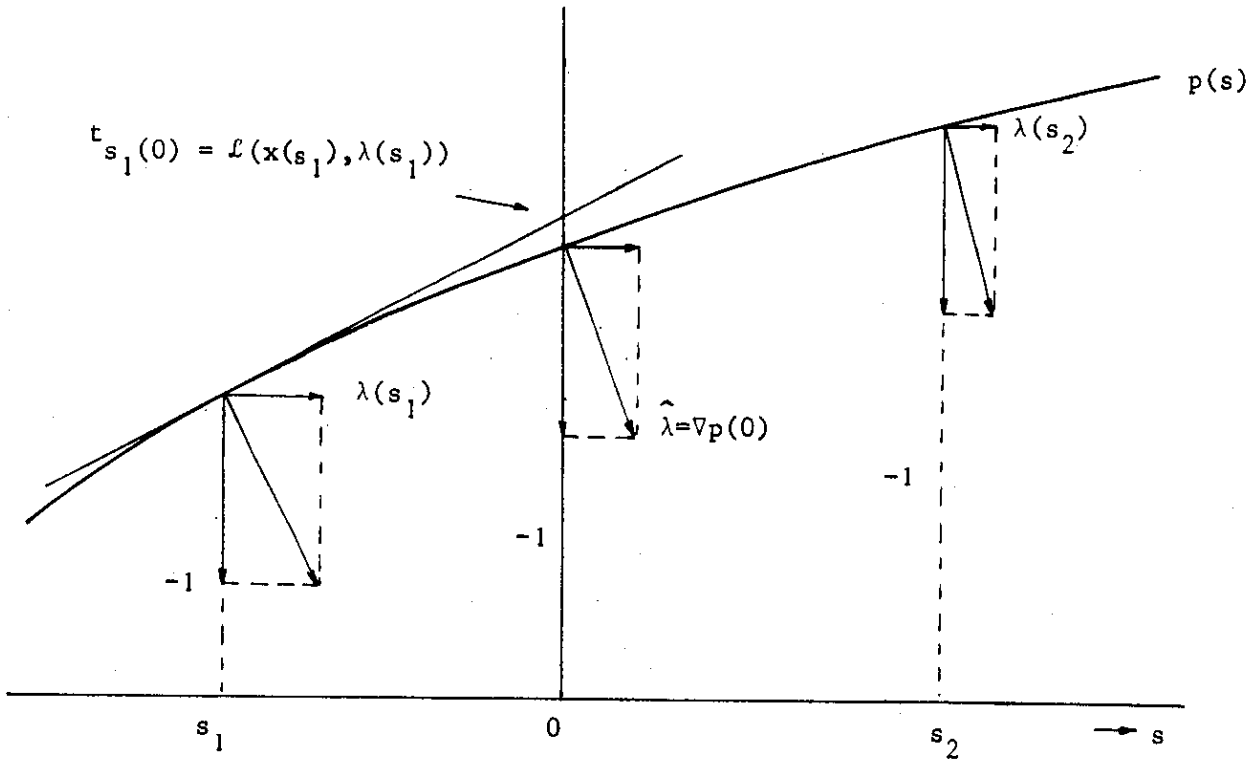
$$\nabla^T p(s) = \nabla^T f(x(s))\ \left[\frac{dx(s)}{ds}\right] = \lambda^T(s)\nabla^T h(x(s))\ \left[\frac{dx(s)}{ds}\right]$$

$$= \lambda^T(s)\ \left[\frac{ds}{ds}\right] = \lambda^T(s)$$

Opgemerkt kan worden dat in het gezochte snijpunt van de primale functie met de s=0-as in het bijzonder geldt dat

$$\nabla p(0) = \lambda(0) := \hat{\lambda} \tag{3.5.16}$$

Een geometrische interpretatie van de primale functie is weergegeven in Figuur 3.5.5



Figuur 3.5.5: De primale functie p(s) en zijn gradiënt

3.5.6. Met de uitdrukking (3.5.9) van de primale functie kan het raakvlak aan (de grafiek van) de primale functie in een punt $(p(\bar{s}),\bar{s})$ worden weergegeven door de uitdrukking

$$\begin{aligned} t_{\bar{s}}(s) &= p(\bar{s}) + \nabla^T p(\bar{s})(s-\bar{s}) \\ &= p(\bar{s}) + \lambda^T(\bar{s})(s-\bar{s}) \\ &= f(x(\bar{s})) - \lambda^T h(x(\bar{s})) + \lambda^T(\bar{s})s \end{aligned} \tag{3.5.17}$$

Voor het punt waar dit raakvlak de lijn s=0 (d.i. de verticale as in Figuur 3.5.5) snijdt volgt daarmee

$$\begin{aligned} t_{\bar{s}}(0) &= f(x(\bar{s})) - \lambda^T(\bar{s})h(x(\bar{s})) \\ &= \phi(x(\bar{s}),\lambda(\bar{s}),0) \end{aligned} \tag{3.5.18}$$

ofwel, en dat is het resultaat waaromheen een groot deel van de hiernavolgende theorie draait

$$t_{\bar{s}}(0) = \mathcal{L}(x(\bar{s}),\ \lambda(\bar{s})) \tag{3.5.19}$$

In woorden betekent dit dat het stuk dat door het raakvlak aan de perturbatiefunctie p(s) wordt afgesneden van de verticale as(en dat uiteraard kan worden gebruikt als een schatting voor de waarden van de oplossing p(0) van het orginele probleem)juist gelijk is aan de waarde van de Lagrangefunctie van het orginele probleem geëvalueerd voor de waarden $x(\bar{s})$ en $\lambda(\bar{s})$ van x en λ die corresponderen met de oplossing van het geperturbeerde probleem in het raakpunt $(p(\bar{s}),\bar{s})$ (Zie Figuur 3.5.5).

3.5.7. De directe toepassing van het resultaat (3.5.19) als een mogelijkheid voor het bepalen van een benadering van de optimale oplossing p(0) biedt weinig voordelen: Voor het evalueren van de waarde van de Lagrangefunctie in (3.5.19) is het namelijk noodzakelijk de oplossing te bepalen van het geperturbeerde probleem

$$\min \{f(x) \mid h(x) = \bar{s},\ x \in \mathbb{R}^n\} \tag{3.5.20}$$

en het oplossen daarvan heeft uiteraard dezelfde moeilijkheidsgraad als de oplossing van het orginele probleem (3.5.1). Het gebruik dat desondanks wordt gemaakt van het resultaat (3.5.19) is dan ook gebaseerd op een andere aanpak van het probleem en wel een aanpak die uitgaat van de observatie dat in de optimale punten x(s) die de waarden van de primale functie p(s) (vgl.(3.5.7)) bepalen steeds voldaan wordt aan de relatie (3.5.10)

$$\nabla f(x(s)) - \nabla h(x(s))\lambda(s) = 0$$

Deze relatie kan worden opgevat als een stelsel van n vergelijkingen met n+m onbekenden. In het geval de Hessiaan (m.b.t. de variabele x) van de Lagrangefunctie (3.5.5)

$$\nabla_{xx}\mathcal{L}(x,\lambda) := G(x) - \sum_{j=1}^{m} \lambda_j H_j(x) \tag{3.5.21}$$

niet-singulier is in een omgeving van de optimale waarde $\hat{x}$ en $\hat{\lambda}$ van het originele probleem dan geldt wegens de impliciete-functiestelling (vgl[3.5.1:App.A]) dat voor iedere waarde van λ in een omgeving van de optimale waarde $\hat{\lambda}$ een vector $x(\lambda)$ kan worden bepaald zodanig dat voldaan wordt aan het stelsel (3.5.10), in een van λ afhankelijke formulering

$$\nabla f(x(\lambda)) - \nabla h(x(\lambda))\lambda = 0 \tag{3.5.22}$$

Voor de aldus te bepalen vectorfunctie $x(\lambda)$ geldt dan bovendien in het bijzonder nog dat

$$x(\hat{\lambda}) = \hat{x} \tag{3.5.23}$$

In de practijk betekent dit door de impliciete-functiestelling gegarandeerde resultaat dat het, gegeven een willekeurige λ in een omgeving van $\hat{\lambda}$, in principe mogelijk is de met die λ corresponderende vectoren $x(\lambda)$ en

$$s(\lambda) = h(x(\lambda)) \tag{3.5.24}$$

te bepalen en daarmee ook het corresponderende punt $(p(s(\lambda)),s(\lambda))$ van de grafiek van de primale functie. Het snijpunt van de verticale as $(s=0)$ met het raakvlak aan (de grafiek van) de primale functie in dat punt kan daarna onmiddellijk worden bepaald op de manier zoals hiervoor uiteengezet (vgl(3.5.19)). De waarde van het stuk dat van de verticale as wordt afgesneden kan op die wijze worden beschouwd als een functie van λ (vgl(3.5.19)), het geen leidt tot de definitie van de functie

$$d(\lambda) := t_{s(\lambda)}(0) = \mathcal{L}(x(\lambda),\lambda) = f(x(\lambda)) - \lambda^T h(x(\lambda)) \tag{3.5.25}$$

Deze aldus in principe te genereren functie $d(\lambda)$ staat in de literatuur bekend als de <u>duale functie</u> (bv.[3.5.3]) die correspondeert met het orginele probleem (3.5.1). Het verband tussen deze duale functie en de eerder gedefinieerde primale of perturbatiefunctie wordt op grond van de besproken geometrische interpretatie (zie Figuur 3.5.5) gegeven door

$$d(\lambda) = p(s(\lambda)) - \lambda^T s(\lambda) \tag{3.5.26a}$$

welke uitdrukking ook volgt uit de overweging dat (vgl.(3.5.13), (3.5.18))

$$d(\lambda(s)) = \varphi(x(s),\lambda(s),0) = p(s) - \lambda^T(s)s. \tag{3.5.26b}$$

Welke uitdrukking ook volgt uit de overweging dat (vgl.(3.5.13), (3.5.18)) duale functie is dat in het bijzonder geldt dat (vgl(3.5.24))

$$d(\hat{\lambda}) = \mathcal{L}(x(\hat{\lambda}),\hat{\lambda}) = f(\hat{x}) = p(0) \tag{3.5.27}$$

Bij toepassingen van de duale functie wordt in plaats van (3.5.25) veelal gebruik gemaakt van de equivalente definitie

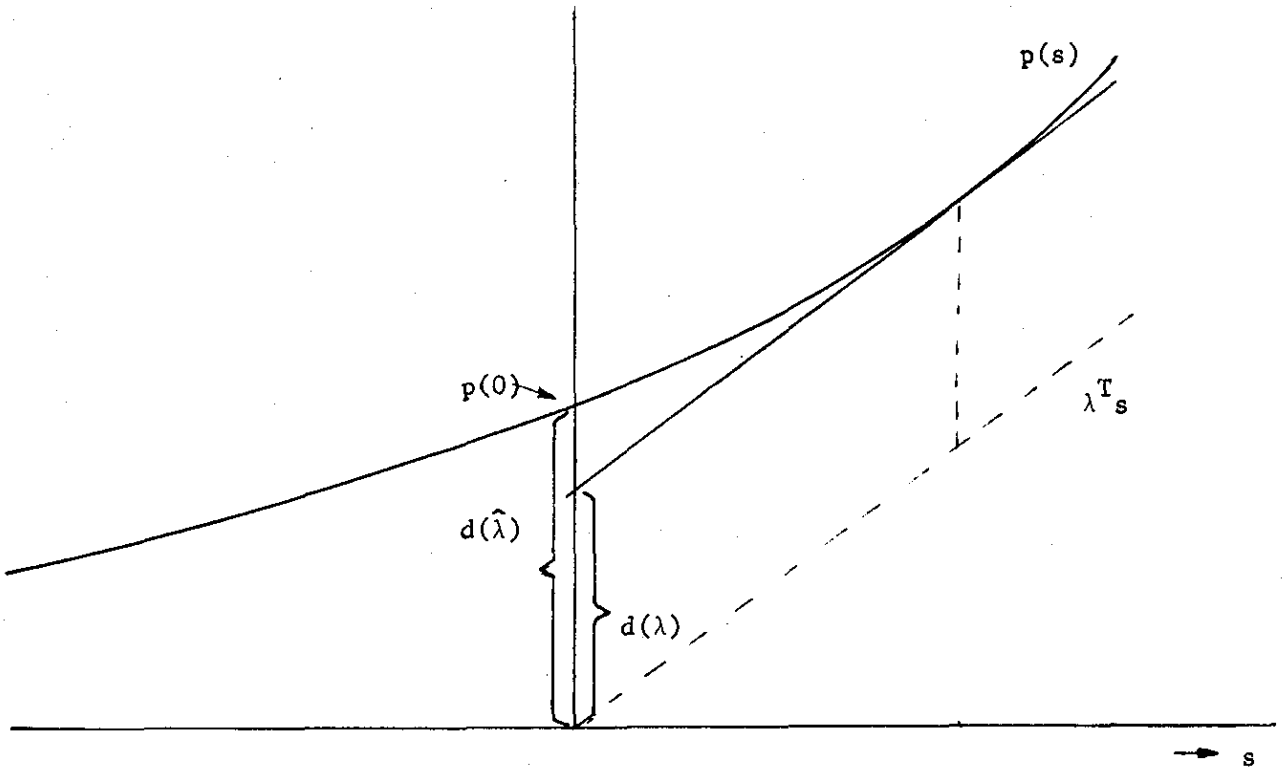
$$\begin{aligned} d(\lambda) &:= \{f(x) - \lambda^T h(x) \mid \nabla f(x) - \nabla h(x)\lambda = 0\} \\ &:= \{\mathcal{L}(x,\lambda) \mid \nabla_x \mathcal{L}(x,\lambda) = 0\} \end{aligned} \tag{3.5.28}$$

Duale formuleringen bij convexe Lagrange functies

3.5.8. In het geval de Lagrange functie (3.5.5) een convexe functie is kan de duale functie ook worden gedefinieerd door de aan de definitie van de primale functie (3.5.6) analoge definitie (vgl(3.5.28))

$$d(\lambda) = \min \{f(x) - \lambda^T h(x) \mid x \in \mathbb{R}^n\} \tag{3.5.29}$$

Omdat (vgl(3.5.25)) $d(\lambda)$ de lengte voorstelt die het raakvlak in het punt $(p(s),s)$ aan de primale functie afsnijdt van de verticale as volgt op geometrische gronden dat in het geval van een convexe Lagrangefunctie



Figuur 3.5.8. De primale functie p(s) en de duale functie $d(\lambda)$ in het geval van een convexe Lagrange functie

het originele probleem,(d.i. (vgl pt 3.5.3) het probleem van de bepaling van de waarde $p(0) = f(x(\hat{\lambda}))$),equivalent is met het maximaliseringsprobleem

$$\begin{aligned} &\max \{d(\lambda) \mid \lambda \in \mathbb{R}^m\} \\ &= \max_{\lambda} \{ \min_{x} \{f(x) - \lambda^T h(x) \mid x \in \mathbb{R}^n\} \mid \lambda \in \mathbb{R}^m\} \end{aligned} \tag{3.5.30}$$

Dit maximaliseringsprobleem staat in de literatuur bekend als het duale probleem (corresponderend met het probleem (3.5.1)) ofwel als de duale formulering van het probleem (3.5.1). Voor de oplossing ervan geldt (vgl(3.5.27))

$$d(\hat{\lambda}) = f(x(\hat{\lambda})) = f(\hat{x}) \tag{3.5.31}$$

Afgeleiden van de duale functie

3.5.9. De hierboven op geometrische gronden gevonden equivalentie (in het geval van convexe Lagrange functies) van het primale probleem (3.5.1) en het duale probleem (3.5.30) kan ook met analytische argumenten worden aangetoond. Een rol daarbij spelen de afgeleiden van de duale functie $d(\lambda)$. Uitgangspunt voor de bepaling daarvan is de overweging dat voor de duale functie geldt

$$d(\lambda) := f(x(\lambda)) - \lambda^T h(x(\lambda)) := \mathcal{L}(x(\lambda),\lambda) \tag{3.5.32}$$

waar $x(\lambda)$ voldoet aan de relatie (3.5.22)

$$\nabla f(x(\lambda)) - \nabla h(x(\lambda))\lambda = \nabla_x \mathcal{L}(x(\lambda),\lambda) = 0$$

Voor de eerste afgeleide of gradiënt van de duale functie volgt daarmee direct dat

$$\nabla d^T(\lambda) = [\nabla f(x(\lambda)) - \nabla h(x(\lambda))\lambda]^T \left[\frac{dx(\lambda)}{d\lambda}\right] - h^T(x(\lambda)) \tag{3.5.33}$$

welke uitdrukking met (3.5.22) overgaat in de bijzonder eenvoudig te hanteren uitdrukking

$$\nabla d(\lambda) := -h(x(\lambda)) \tag{3.5.34}$$

3.5.10. Differentiatie van (3.5.34) geeft

$$\nabla^2_{\lambda\lambda} d(\lambda) := -\nabla^T h(x(\lambda))\left[\frac{dx(\lambda)}{d\lambda}\right] \tag{3.5.35}$$

in welke uitdrukking de $n \times m$-matrix $\left[\frac{dx(\lambda)}{d\lambda}\right]$ voorkomt die kan worden bepaald met behulp van het resultaat van differentiatie van (3.5.22)

$$\nabla^2_{xx} \mathcal{L}(x(\lambda),\lambda)\left[\frac{dx(\lambda)}{d\lambda}\right] - \nabla h(x(\lambda)) = 0 \tag{3.5.36}$$

In de veronderstelling dat de Hessiaan van de Lagrange functie niet-singulier is (hetgeen o.a. het geval zal zijn voor (x,λ)-waarden in een omgeving van de optimale punt $(\hat{x},\hat{\lambda})$ indien de Lagrange functie aldaar strikt convex is) volgt hieruit

$$[\frac{dx(\lambda)}{d\lambda}] = [\nabla^2_{xx} \mathcal{L}(x(\lambda),\lambda)]^{-1} \nabla h(x(\lambda)) \qquad (3.5.37)$$

Voor de <u>Hessiaan van de duale functie</u> volgt daarmee dan

$$\nabla^2_{\lambda\lambda} d(\lambda) := - \nabla^T h(x(\lambda)) [\nabla^2_{xx} \mathcal{L}(x(\lambda),\lambda)]^{-1} \nabla h(x(\lambda)) \qquad (3.5.38)$$

3.5.11. Uit (3.5.38) blijkt dat als de Lagrangefunctie strikt convex is in de omgeving van het optimale punt $(\hat{x},\hat{\lambda})$ dat dan de duale functie in dezelfde omgeving strikt concaaf is. Aangezien in het optimale punt $(\hat{x},\hat{\lambda})$ ook geldt dat

$$\nabla d(\hat{\lambda}) = - h(x(\hat{\lambda})) = - h(\hat{x}) = 0 \qquad (3.5.39)$$

volgt onmiddellijk dat de duale functie $d(\lambda)$ voor $\lambda = \hat{\lambda}$ een (lokaal) maximum heeft met als maximale waarde (vgl(3.5.31))

$$d(\hat{\lambda}) = d(x(\hat{\lambda})) = f(\hat{x}).$$

De equivalentie tussen de duale en de primale formulering van het orginele probleem in het geval van een strikt convexe Lagrangefunctie is daarmee ook analytisch aangetoond. Deze equivalentie impliceert dat de oplossing van het orginele probleem in dit geval dus ook kan worden bepaald door de oplossing van het duale probleem. In een aantal gevallen biedt deze duale probleemaanpak duidelijk voordelen boven de gewone primale aanpak.

<u>Numerieke oplossing van het duale probleem</u>

3.5.12. Voor de numerieke oplossing van het duale probléem (3.5.30) kunnen in principe alle in het voorgaande hoofdstuk besproken onbeperkte minimaliseringsalgorithmen worden toegepast. Gezien de eenvoud waarmee de gradiënt (3.5.34) van de duale functie bepaald kan worden en gezien de omstandigheid dat voor iedere functieevaluatie een (onbeperkt) minimaliseringsprobleem moet worden opgelost genieten gradiënt- en hoger orde methoden daarbij veruit de voorkeur boven eventuele "direct search"-methoden.

Bij de gradiëntmethoden wordt daarbij dan uitgegaan van de standaard-iteratieformule (vgl.pt. 2.1.5 en pt. 2.4.1)

$$\begin{aligned} \lambda^{(k+1)} &:= \lambda^{(k)} - \alpha^{(k)} \nabla d(\lambda^{(k)}) \\ &:= \lambda^{(k)} + \alpha^{(k)} h(x(\lambda^{(k)})) \end{aligned} \tag{3.5.40}$$

waarin $\alpha^{(k)}$ een nader (door bijvoorbeeld lijnminimalisering) te bepalen stapgrootte voorstelt. Eventuele hoger-orde methoden hebben de Newton-formule (2.5.3) tot uitgangspunt. In dit geval krijgt deze de vorm

$$\begin{aligned} \lambda^{(k+1)} &:= \lambda^{(k)} - [\nabla^2_{\lambda\lambda} d(\lambda^{(k)})]^{-1} \nabla d(\lambda^{(k)}) \\ &:= \lambda^{(k)} - [N^{(k)T} \mathcal{L}^{(k)-1}_{xx} N^{(k)}]^{-1} h(x(\lambda^{(k)})) \end{aligned}$$

in welke uitdrukking gebruik werd gemaakt van de verkorte notaties (vgl(3.5.11))

$$N^{(k)} := \nabla h(x(\lambda^{(k)}) \tag{3.5.42}$$

en

$$\mathcal{L}^{(k)}_{xx} := \nabla^2_{xx} \mathcal{L}(x(\lambda^{(k)}), \lambda^{(k)}) \tag{3.5.43}$$

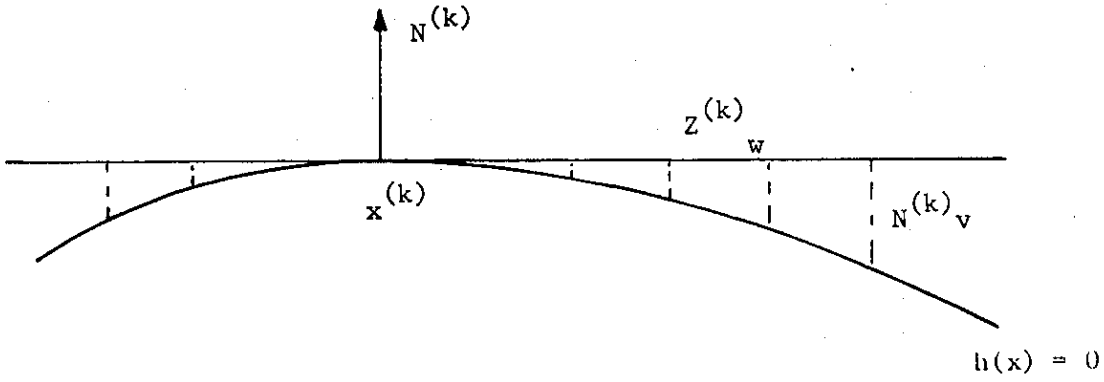
Opgemerkt kan worden dat de convergentiesnelheid van gradiënt- en daarvan afgeleide methoden in dit geval afhangt (vgl pt 2.4.5-2.4.7) van de conditie (d.i. de verhouding van de grootste tot de kleinste eigenwaarde) van de Hessiaan van de duale functie die in de notatie van (3.5.41) de vorm heeft

$$N^{(k)T} \mathcal{L}^{(k)-1}_{xx} N^{(k)} \tag{3.5.44}$$

3.5.13. Het hierboven genoemde resultaat voor de convergentiesnelheid van duale methoden contrasteert op interessante wijze met een (door o.a. Luenberger [3.5.1] besproken) analoog resultaat voor de toepassing van primale methoden op problemen met niet lineaire beperkingen. De convergentiesnelheid daarbij blijkt namelijk direct afhankelijk van de conditie van de matrix

$$Z^{(k)T} \mathcal{L}^{(k)}_{xx} Z^{(k)} \tag{3.5.45}$$

Hierin is $Z^{(k)}$ de matrix van orthonormale basisvectoren van het orthogonale complement van de deelminute opgespannen door de kolommen van de matrix $N^{(k)}$. Dit convergentie resultaat hangt nauw samen met de omstandigheid dat (vgl pt 3.3.27) toepassingen van primale methoden op minimaliseringsproblemen met niet-lineaire nevenvoorwaarden kunnen worden opgevat als toepassingen van gradiëntmethoden in het (n-q)-dimensionale raakvlak aan de beperkingen mits men de objectfunctie daar fictief de waarden toekent van de objectfunctie ter plaatse van de werkelijke beperkingen. In de practijk kan dit worden opgevat als een projectie op het raakvlak zoals geschetst in Figuur 3.5.12



Figuur 3.5.12. Projectie van de functiewaarden f(x) t.p.v. de beperkingen $h(x) = 0$ op het raakvlak $N^{(k)T}(x-x^{(k)}) = 0$

In tweede-orde-ontwikkeling leidt deze "projectie" zoals besproken in pt 3.3.27 tot het probleem

$$\min \{f(x^{(k)}) + \nabla^T f(x^{(k)}) \; (Z^{(k)}w + N^{(k)}v) + \tfrac{1}{2} \; w^T \; Z^{(k)T} G^{(k)} Z^{(k)} w \;\Big|$$

$$\Big| \; (N^{(k)T} N^{(k)} v)_i = - \tfrac{1}{2} \; w^T Z^{(k)T} H_i^{(k)} Z^{(k)} w, \; i = 1,\ldots,q\}$$

dat met

$$\lambda^{(k)} := (N^{(k)T} N^{(k)})^{-1} \; N^{(k)T} \; \nabla f(x^{(k)}) \qquad (3.5.47)$$

ook geformuleerd kan worden als

$$\min \{f(x^{(k)}) + \nabla^T f(x^{(k)}) \; Z^{(k)}w + \tfrac{1}{2} w^T Z^{(k)T} \; \nabla^2_{xx} \, \mathcal{L}(x^{(k)}, \lambda^{(k)}) \; Z^{(k)}_w \;\Big|$$

$$\Big| \; w \in \mathbb{R}^{n-q}\} \qquad (3.5.48)$$

Voor een meer rigoreuze afleiding van dit conceptueel bijzonder interessante resultaat zij verwezen naar Luenberger ([3.5.1]).

Aangevulde Lagrangefuncties

3.5.14. Bij de in de voorgaande punten besproken duale aanpak van optimaliseringsproblemen met nevenvoorwaarden werd een expliciet gebruik gemaakt van de veronderstelling dat de Lagrangefunctie van het originele probleem een (strikt) convexe functie was in de variabele $x \in \mathbb{R}^n$. Zonder deze convexiteit kan niet worden gegarandeerd dat het stationaire punt van de Lagrangefunctie gevonden kan worden met behulp van een onbeperkte minimaliseringsalgorithme. (vgl. pt 3.5.8). Bij de meeste minimaliseringsproblemen met nevenvoorwaarden geldt dat in het optimale punt voldaan wordt aan de voldoende voorwaarden voor een minimum (Stelling 3.1.16) hetgeen o.a. impliceert dat de Lagrangefunctie convex is in alle richtingen in het raakvlak aan de beperkingen in het optimale punt.
Er geldt immers (vgl(3.1.51))

$$\forall_{Z \in \mathbb{R}^n} : \hat{N}^T Z = 0 \Rightarrow Z^T \mathcal{L}_{xx}(\hat{x},\hat{\lambda}) Z > 0 \qquad (3.5.49)$$

Anders ligt dit voor de richtingen loodrecht op dit raakvlak, of anders gezegd, voor de richtingen in de deelruimte opgespannen door de normalen op de beperkingen. Aan de convexiteit van de Lagrangefunctie in deze richtingen wordt door de optimaliteitsvoorwaarden geen enkele eis gesteld. Bij vele practische problemen blijkt (zie b.v. [3.5.9] voor een discussie daarover) er van convexiteit in die richtingen geen sprake te zijn. Om juist in die gevallen toch gebruik te kunnen maken van de hiervoor besproken duale aanpak is het concept van de aangevulde Lagrangefunctie (eng: Augmented Lagrangian (function) geïntroduceerd. Hieronder verstaat men een functie van 3 variabelen $x \in \mathbb{R}^n$, $\lambda \in \mathbb{R}^m$ en $\rho \in \mathbb{R}^1_+$ die bestaat uit de som van de Lagrangefunctie plus een door de factor ρ geschaalde aanvulling in de trant van een boetefunctie. Deze aanvulling geeft de functie in de richting van vectoren in de deelruimte opgespannen door de normalen op de beperkingen een meer convex karakter. In formulevorm krijgt de aangevulde Lagrangefunctie dan de gedaante

$$\begin{aligned} Q(x,\lambda,\rho) &:= \mathcal{L}(x,\lambda) + \rho\, \Psi(x) \\ &:= f(x) - \lambda^T h(x) + \rho\, \Psi(x) \\ &:= f(x) + \Omega(x,\lambda,\rho) \end{aligned} \qquad (3.5.50)$$

waarin, respectievelijk, naar analogie met de boetefuncties (vgl (3.4.2)) geldt

$$\Psi(x) := \sum_{i=1}^{m} \psi_i\ (h_i(x)) \tag{3.5.51}$$

en

$$\begin{aligned} \Omega(x,\lambda,\rho) &:= \sum_{i=1}^{m} \omega_i\ (h_i(x),\ \lambda_i,\rho) \\ &:= \sum_{i=1}^{m} (-\lambda_i h_i(x) + \rho\ \psi_i(h_i(x))) \end{aligned} \tag{3.5.52}$$

en waarin de functies $\psi_i : \mathbb{R}^1 \to \mathbb{R}^1_+$ tweemaal continu differentieerbare functies zijn met de eigenschap (vgl(3.4.5)) dat

$$\begin{aligned} \psi_i(t) &= 0 \quad \text{als} \quad t = 0 \\ &> 0 \quad \text{als} \quad t \neq 0 \end{aligned} \tag{3.5.53}$$

en

$$\psi_i''(0) = 1 \tag{3.5.54}$$

De meest gebruikte vorm van de functies $\psi_i(t)$ is ook hier weer (vgl (3.4.7)) de kwadratische functie

$$\psi_i(t) = \tfrac{1}{2}t^2 \tag{3.5.55}$$

waarmee de meest gebruikte vorm van de aangevulde Lagrangefunctie wordt

$$Q(x,\lambda,\rho) := f(x) - \lambda^T h(x) + \tfrac{1}{2}\rho\ h^T(x)h(x) \tag{3.5.56}$$

Veel van de hierna volgende theorie is in het bijzonder op deze laatste, speciale vorm aan de aangevulde Lagrangefunctie geënt.

Afgeleiden van de aangevulde Lagrangefunctie

3.5.15. Eenmaal differentieren naar x van de aangevulde Lagrangefunctie geeft onmiddellijk als algemene vorm voor de gradiënt naar x

$$\begin{aligned} \nabla_x Q(x,\lambda,\rho) &:= \nabla f(x) - \sum_{i=1}^{m} \lambda_i \nabla h_i(x) + \rho \sum_{i=1}^{m} \psi_i'(h_i(x))\ \nabla h_i(x) \\ &:= \nabla f(x) - \sum_{i=1}^{m} (\lambda_i - \rho\psi_i'(h_i(x)))\nabla h_i(x) \end{aligned} \tag{3.5.57}$$

welke vorm in het speciale geval van de kwadratische aanvulling (met (3.5.55)) overgaat in

$$\nabla_x Q(x,\lambda,\rho) := \nabla f(x) - \nabla h(x)\ (\lambda - \rho h(x)) \tag{3.5.58}$$

Substitutie van de optimale waarden $\hat{x}$ en $\hat{\lambda}$ behorend bij het optimale punt in deze uitdrukkingen levert in beide gevallen als gradiënt van de aangevulde Lagrangefunctie (die correspondeert met de optimale waarde van $\hat{\lambda}$) in het optimale punt $\hat{x}$

$$\nabla_x Q(\hat{x},\hat{\lambda},\rho) := \nabla f(\hat{x}) - \nabla h(\hat{x})\ \hat{\lambda} = \nabla_x \mathcal{L}(\hat{x},\hat{\lambda}) \tag{3.5.59}$$

Aangezien in het optimale punt (vgl(3.5.2))

$$\nabla_x \mathcal{L}(\hat{x},\hat{\lambda}) = 0$$

volgt dat ook

$$\nabla_x Q(\hat{x},\hat{\lambda},\rho) = 0 \tag{3.5.60}$$

Dit resultaat impliceert dat de betreffende aangevulde Lagrangefunctie een stationair punt heeft m.b.t. x in het optimale punt van het orginele probleem.

3.5.16. Differentiatie naar x van de uitdrukkingen (3.5.57) en (3.5.58) voor de gradiënt van aangevulde Lagrangefuncties levert op zijn beurt als uitdrukkingen voor de Hessiaan van diezelfde functies respectievelijk

$$\nabla^2_{xx} Q(x,\lambda,\rho) := G(x) - \sum_{i=1}^{m} (\lambda_i - \rho\psi_i'(h_i(x))\ H_i(x) + \rho \sum_{i=1}^{m} \psi_i''(h_i(x))\nabla h_i(x)\nabla h_i^T(x) \tag{3.5.61}$$

en

$$\nabla^2_{xx} Q(x,\lambda,\rho) := G(x) - \sum_{i=1}^{m} (\lambda_i - \rho\ h_i(x))H_i(x) + \rho\ \nabla h(x)\ \nabla h^T(x) \tag{3.5.62}$$

Met de optimale waarden $\hat{x}$ en $\hat{\lambda}$ van het orginele probleem gaan deze uitdrukkingen beide over in

$$\nabla^2_{xx} Q(\hat{x},\hat{\lambda},\rho) := G(\hat{x}) - \sum_{i=1}^{m} \hat{\lambda}_i H_i(\hat{x}) + \rho\, \nabla h(\hat{x}) \nabla h^T(\hat{x})$$

$$:= \nabla^2_{xx} \mathcal{L}(\hat{x},\hat{\lambda}) + \rho\, \nabla h(\hat{x}) \nabla h^T(\hat{x}) \qquad (3.5.63)$$

Uit dit laatste resultaat kan eenvoudig worden afgeleid dat in het geval dat in het optimale punt van het orginele probleem voldaan is aan de voldoende voorwaarde voor optimaliteit (3.5.4) dat dan steeds een voldoende grote positieve waarde $\hat{\rho} > 0$ voor ρ gevonden kan worden zodanig dat de Hessiaan van de aangevulde Lagrangefunctie $Q(x,\hat{\lambda},\hat{\rho})$ strikt positief definiet is in het optimale punt $\hat{x}$.
In dat geval geldt zowel

$$\nabla_x Q(\hat{x},\hat{\lambda},\hat{\rho}) = 0 \qquad (3.5.64)$$

als

$$\forall_{y \in \mathbb{R}^n} : y^T \nabla^2_{xx} Q(\hat{x},\hat{\lambda},\hat{\rho}) y > 0 \qquad (3.5.65)$$

hetgeen op zijn beurt impliceert dat de aangevulde Lagrangefunctie $Q(x,\hat{\lambda},\hat{\rho})$ juist als de gewone Lagrangefunctie bij convexe minimaliseringsproblemen een onbeperkt minimum heeft in het optimale punt van het orginele probleem. Dit resultaat dat op zich zelf een belangrijke illustratie is van de analogie die bestaat tussen het gebruik van de Lagrangefunctie bij convexe problemen en het gebruik van de aangevulde Lagrangefunctie bij niet-convexe problemen, vormt de theoretische basis voor het gebruik van de hieronder te bespreken aangevulde-Lagrange- en de exacte-boetefunctie methoden voor het oplossen van niet-lineaire minimaliseringsproblemen met niet-lineaire nevenvoorwaarden.

3.5.17. De aangevulde Lagrangefunctie kan ook worden opgevat als de gewone Lagrangefunctie van een van het orginele probleem afgeleide "aangevuld" probleem

$$\min \{\tilde{f}(x) \mid h(x) = 0\} \qquad (3.5.66)$$

waarin de (ook hieronder steeds met een ~-symbool aangegeven) "aangevulde functie" $\tilde{f}(x)$ de vorm heeft

$$\tilde{f}(x) := f(x) + \rho \sum_{i=1}^{m} \psi_i(h_i(x)) \qquad (3.5.67)$$

, ofwel, in het speciale geval van de gebruikelijke kwadratische aanvul-

ling, de vorm

$$\tilde{f}(x) := f(x) + \tfrac{1}{2}\,\rho\, h^T(x)h(x) \tag{3.5.68}$$

Verondersteld wordt daarbij dat ρ telkens een zo groot gekozen schaalfactor is dat de Hessiaan van de corresponderende aangevulde Lagrangefunctie positief definitief is in een omgeving van de optimale waarden voor x en λ (vgl pt(3.5.15). Zoals eenvoudig kan worden aangetoond heeft het aangevulde probleem (3.5.66) een beperkt minimum met dezelfde optimale waarde

$$\tilde{f}(\hat{x}) = f(\hat{x}) \tag{3.5.69}$$

voor dezelfde (optimale) waarden $\hat{x}$ en $\hat{\lambda}$ als het orginele probleem. In het bijzonder geldt ook dat

$$\begin{aligned}\nabla\tilde{\mathcal{L}}(\hat{x},\hat{\lambda}) &= \nabla f(\hat{x}) + \sum_{i=1}^{m} \psi_i(h_i(\hat{x}))\,\nabla h_i(\hat{x}) - \sum_{i=1}^{m} \hat{\lambda}_i \nabla h_i(\hat{x}) \\ &= \nabla f(\hat{x}) - \nabla h(\hat{x})\hat{\lambda} = \\ &= \nabla\mathcal{L}(\hat{x},\hat{\lambda}) = 0\end{aligned} \tag{3.5.70}$$

Het verschil tussen het aangevulde probleem en het orginele probleem ligt namelijk in de gegarandeerde convexiteit van Lagrangefunctie in het optimale punt. Dit laatste maakt het mogelijk toch gebruik te maken van een duale probleemaanpak bij een in wezen niet-convex probleem. Hieronder zal deze aanpak waarbij de aangevulde Lagrangefunctie de plaats inneemt van de gewone Lagrangefunctie bij convexe problemen ten overvloede nogmaals in het kort worden beschreven. Ter vereenvoudiging wordt de discussie daarbij beperkt tot het geval van de kwadratische aanvullling (3.5.55). Voor andere aangevulde Lagrangefuncties verloopt de discussie analoog.

Duale probleemformulering met aangevulde Lagrangefuncties

3.5.18. Uitgangspunt voor de duale probleemformulering (vgl pt 3.5.8) van het aangevulde probleem (3.5.68)

$$\min \{f(x) + \tfrac{1}{2}\rho h^T(x)h(x) \mid h(x) = 0\}$$

is (vgl pt 3.5.3) is de aangevulde primale functie die wordt gegeven door

$$\tilde{p}(s) := \min \{f(x) + \tfrac{1}{2}\rho h^T(x)h(x) \mid h(x) = s\} \tag{3.5.71}$$

en waarvoor geldt (vgl(3.5.13))

$$\tilde{p}(s) := f(\tilde{x}(s)) + \tfrac{1}{2}\rho h^T(\tilde{x}(s))h(\tilde{x}(s)) - \tilde{\lambda}^T(s)(h(\tilde{x}(s))-s)$$

$$:= \tilde{\varphi}(\tilde{x}(s),\ \tilde{\lambda}(s),s) \qquad (3.5.72)$$

waarin $\tilde{x}(s)$ de waarde van x is waarom het minimum in (3.5.71) wordt aangenomen en waarin $\tilde{\lambda}(s)$ de waarde van λ is waarvoor geldt (vgl(3.5.10))

$$\nabla f(\tilde{x}(s)) + \rho\nabla h(\tilde{x}(s))h(\tilde{x}(s)) - \nabla h(\tilde{x}(s))\tilde{\lambda}(s) = 0 \qquad (3.5.73)$$

Voor de gradiënt van de aangevulde primale functie volgt naar analogie van (3.5.9), (3.5.14) en (3.5.15) dat

$$\nabla\tilde{p}(s) = \nabla_s\tilde{\phi}(\tilde{x}(s),\tilde{\lambda}(s),s) = \tilde{\lambda}(s) \qquad (3.5.74)$$

en het raakvlak aan de aangevulde primale functie in het punt $(\tilde{p}(\bar{s}),\bar{s})$ kan daarmee worden weergegeven door (vgl (3.5.17))

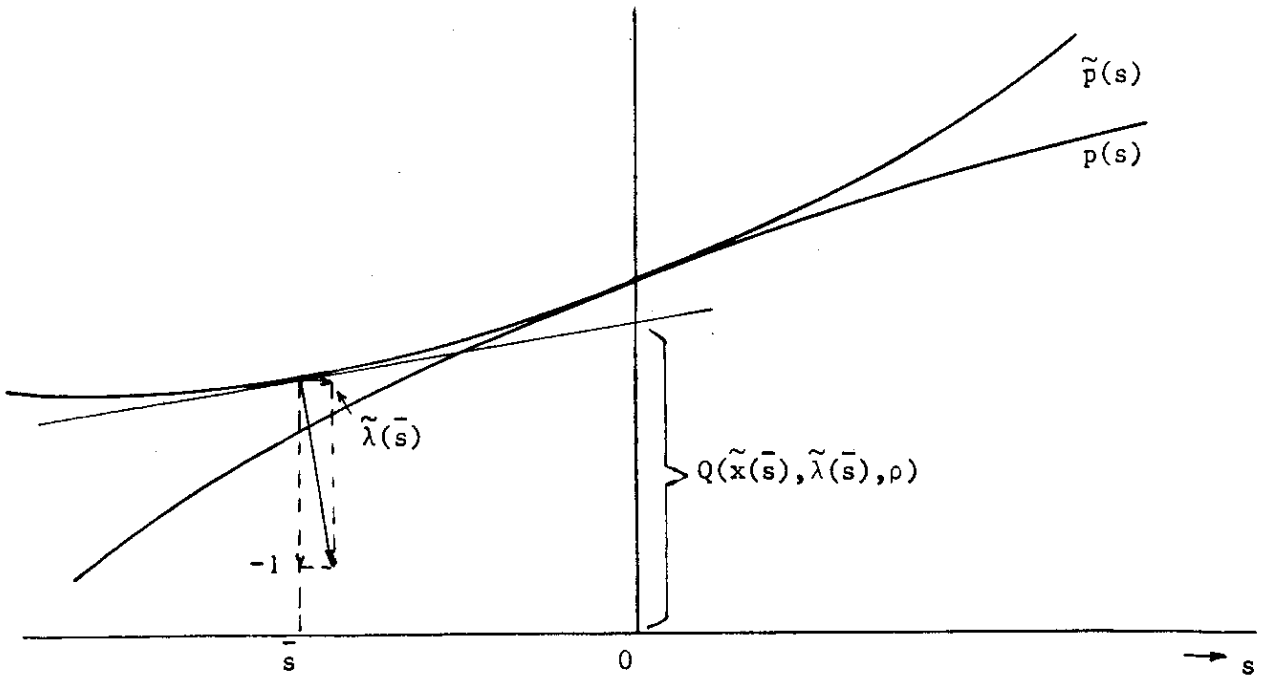
$$\underset{\bar{s}}{\tilde{t}}(s) := \tilde{p}(\bar{s}) + \nabla^T\tilde{p}(\bar{s})(s-\bar{s})$$

$$= f(\tilde{x}(\bar{s})) + \tfrac{1}{2}\rho h^T(\tilde{x}(\bar{s}))h(\tilde{x}(\bar{s})) + \tilde{\lambda}^T(\bar{s})(s-\bar{s})$$

Voor het snijpunt van dit raakvlak (zie figuur 3.5.18) met de verticale as s = 0 volgt daaruit

$$\underset{\bar{s}}{\tilde{t}}(0) := \tilde{\varphi}(\tilde{x}(\bar{s}),\tilde{\lambda}(\bar{s}),0) := Q(\tilde{x}(\bar{s}),\tilde{\lambda}(\bar{s}),\rho) \qquad (3.5.76)$$

waarin $Q(x,\lambda,\rho)$ de eerder gedefinieerde aangevulde Lagrangefunctie (3.5.56) is



Figuur 3.5.18: Aangevulde en gewone primale functie

Omdat de aangevulde primale functie (vgl(3.5.71))

$$\tilde{p}(s) := f(\tilde{x}(s)) + \tfrac{1}{2}\rho s^T s := p(s) + \tfrac{1}{2}\rho s^T s \tag{3.5.77}$$

voor de gekozen voldoende grote waarde van ρ strikt convex is geldt dat het snijpunt van het raakvlak met de verticale as ook kan worden opgevat als een functie van de normaal $(\tilde{\lambda}(s),-1)^T$ op dit raakvlak. Dit leidt dan tot de definitie van de <u>aangevulde duale functie</u> naar analogie van (3.5.25)

$$\begin{aligned}\tilde{d}(\lambda) &:= Q(\tilde{x}(\lambda),\lambda,\rho) \\ &:= f(\tilde{x}(\lambda)) + \tfrac{1}{2}\rho h^T(\tilde{x}(\lambda))h(x(\lambda)) - \lambda^T h(\tilde{x}(\lambda))\end{aligned} \tag{3.5.78}$$

waar $\tilde{x}(\lambda)$ en λ zijn gekoppeld aan elkaar via de voorwaarde (vgl(3.5.22) en (3.5.73))

$$\nabla f(\tilde{x}(\lambda)) - \nabla h(\tilde{x}(\lambda))\,[\lambda - \rho\, h(\tilde{x}(\lambda))] = 0 \tag{3.5.79}$$

Deze aangevulde duale functie kan, omdat $Q(x,\lambda,\rho)$ convex is (vgl(3.5.65)), ook worden weergegeven door de gebruikelijker formulering (vgl(3.5.29))

$$\tilde{d}(\lambda) := \min\{f(x) + \tfrac{1}{2}\rho h^T(x)h(x) - \lambda^T h(x) \mid x \in \mathbb{R}^n\} \tag{3.5.80}$$

Voor de <u>gradiënt</u> van $\tilde{d}(\lambda)$ volgt op analoge wijze als voor de gradiënt van de niet-aangevulde Lagrangefunctie (vgl(3.5.34))

$$\nabla\tilde{d}(\lambda) := -h(\tilde{x}(\lambda)) \tag{3.5.81}$$

en voor de <u>Hessiaan</u> eveneens analoog (vgl(3.5.38))

$$\nabla^2_{\lambda\lambda}\,\tilde{d}(\lambda) := -\nabla^T h(\tilde{x}(\lambda))\,[\nabla^2_{xx}\,Q(\tilde{x}(\lambda),\lambda\rho)]^{-1}\,\nabla h(\tilde{x}(\lambda) \tag{3.5.82}$$

Omdat de Hessiaan van de aangevulde Lagrangefunctie strikt positief definiet is (vgl pt 3.5.11) geldt dat de Hessiaan van de aangevulde duale functie strikt negatief definiet is en de aangevulde duale functie dus strikt concaaf met een (lokaal) maximum voor de optimale waarde $\hat{\lambda}$ van λ. Het orginele probleem (3.5.1) is daar equivalent aan het probleem

$$\begin{aligned}&\max_{\lambda}\{\min_{x}\{Q(x,\lambda,\rho) \mid x \in \mathbb{R}^n\} \mid \lambda \in \mathbb{R}^m\} \\ &= \max_{\lambda}\{\min_{x}\{f(x)-\lambda^T h(x)+\tfrac{1}{2}\rho h^T(x)h(x) \mid x \in \mathbb{R}^n\} \mid \lambda \in \mathbb{R}^m\}\end{aligned} \tag{3.5.83}$$

Dit resultaat impliceert de eerder aangekondigde mogelijkheid voor de toepassing van de duale probleemaanpak voor niet-convexe problemen.

Multiplicatoren - of aangevulde -Lagrangefunctiemethoden

3.5.19. De hiervoor besproken duale probleemformulering van niet-convexe problemen met behulp van de aangevulde-Lagrangefunctie vormt de theoretische basis voor een klasse van numerieke methoden voor de oplossing van niet-lineaire minimaliseringsproblemen met niet-lineaire nevenvoorwaarden die bekend staan als de multiplicatoren methoden ofwel aangevulde-Lagrangefunctiemethoden. Het basisidee achter deze methoden werd in 1969 onafhankelijk van elkaar gesuggereerd door Hestenes [3.5.16] en Powell [3.5.23] en een jaar later nog eens door Haarhoff en Buys [3.5.15]. Direct daarna hebben een groot aantal onderzoekers zich met de methoden beziggehouden en bijgedragen tot de snelle ontwikkeling ervan. Bijzondere vermelding in dit verband verdient zeker het werk van Rockafellar [3.5.24][3.5.25], Miele en medewerkers [3.5.21] en Bertsekas [3.5.5],[3.5.6]. Deze laatste schreef ook een duidelijk survey artikel [3.5.7] dat de stand van zaken weergeeft aan het einde van het jaar 1975.

3.5.20. De procedure bij de toepassing van de multiplicatorenmethoden is vergelijkbaar met de procedure bij het gebruik van boetefunctiemethoden. De algorithme heeft een iteratief karakter en het hoofdbestanddeel van iedere iteratie wordt gevormd door de (onbeperkte) minimalisering van de aangevulde Lagrangefunctie met telkens aangepaste schattingen voor de (Lagrange)-multiplicatorenvector λ en de schaalfactor ρ. In de k-de iteratie wordt op deze manier het (onbeperkte) minimum gezocht van de functie van x

$$\begin{aligned} Q(x,\lambda^{(k)},\rho^{(k)}) &:= f(x) - \lambda^{(k)T}h(x) + \rho^{(k)}\,\Psi(h(x)) \\ &:= f(x) - \lambda^{(k)T}h(x) + \rho^{(k)}\sum_{i=1}^{m}\psi_i(h(x)) \end{aligned} \qquad (3.5.84)$$

of, in het speciale geval van een kwadratische aanvulling, van de functie

$$Q(x,\lambda^{(k)},\rho^{(k)}) := f(x) - \lambda^{(k)T}h(x) + \tfrac{1}{2}\rho^{(k)}h^{T}(x)h(x) \qquad (3.5.85)$$

Zodra een (goede) benadering $x^{(k)}$ voor de oplossing $\tilde{x}(\lambda^{(k)})$ (vgl (3.5.79)) van het minimaliseringsprobleem in de k-de iteratie

$$\min\{Q(x,\lambda^{(k)},\rho^{(k)}) \mid x \in \mathbb{R}^n\} \qquad (3.5.86)$$

gevonden is, worden nieuwe schattingen voor λ en ρ bepaald. Voor de verbetering van de schatting voor λ wordt daarbij bij nagenoeg alle methoden gebruik gemaakt van de multiplicatoren-iteratieformule (van Hestenes [3.5.16])

$$\lambda^{(k+1)} := \lambda^{(k)} - \rho^{(k)} \psi'(h(x^{(k)})) \tag{3.5.87}$$

in welke uitdrukking $\psi'(h(x))$ een kolomvector voorstelt met als elementen de afgeleiden

$$\psi_i'(h(x)) := \frac{d\psi_i}{dt}(h(x)) \tag{3.5.88}$$

In het geval van de kwadratische aanvulling gaat deze multiplicatoren-iteratieformule dan over in zijn meest gebruikte vorm

$$\lambda^{(k+1)} := \lambda^{(k)} - \rho^{(k)} h(x^{(k)}) \tag{3.5.89}$$

Tegelijk met deze multiplicatoren aanpassing, die hieronder in meer detail zal worden besproken, wordt meestal ook (en dit wordt door de meeste auteurs op grond van numerieke ervaringen aanbevolen) de schaalfactor vergroot met een vaste factor $\nu > 1$, d.i.

$$\rho^{(k+1)} := \nu\rho^{(k)} \tag{3.5.90}$$

Integenstelling tot de situatie bij de boetefunctiemethoden behoeft deze schaalfactor ρ niet naar oneindig te gaan voor convergentie, reden waarom de factor ν meestal niet veel groter is dan 1 (meestal geldt $1 < \nu \leq 10$). Nodig is dan wel dat begonnen wordt met een redelijke beginschatting $\rho^{(1)}$ voor de schaalfactor. Voor de beginschatting voor de multiplicator $\lambda^{(1)}$ kan gekozen worden

$$\lambda^{(1)} := 0 \tag{3.5.91}$$

Met deze waarde komt de eerste iteratie van de multiplicatorenmethode overeen met de eerste iteratie van een boetefunctiemethode.

3.5.21. De multiplicatoreniteratieformule dankt zijn speciale vorm (3.5.87) aan de omstandigheid dat in het optimale punt $x^{(k)}$ van de minimaliseringsprobleem in de k-de iteratie geldt

$$\nabla_x Q(x^{(k)},\lambda^{(k)},\rho^{(k)}) = 0 = \nabla f(x^{(k)}) - \nabla h(x^{(k)})(\lambda^{(k)} - \rho^{(k)}\psi'(h(x^{(k)}))) \tag{3.5.92}$$

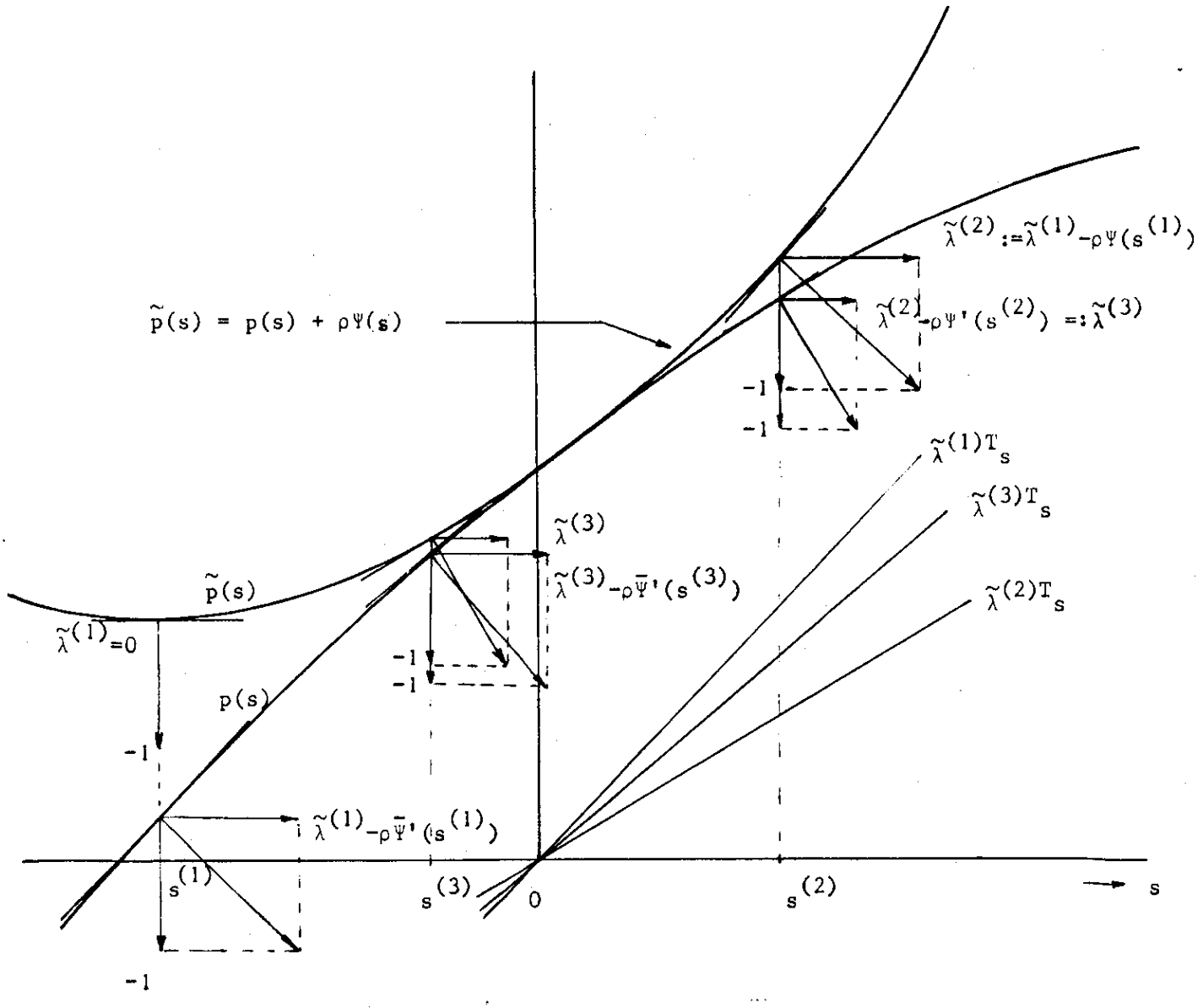
terwijl in het optimale punt van het orginele probleem moet gelden (vgl(3.5.2))

$$0 = \nabla f(\hat{x}) - \nabla h(\hat{x})\ \hat{\lambda}$$

Vergelijking van deze uitdrukkingen doet vermoeden dat $\lambda^{(k+1)}$ gegeven door de uitdrukking (3.5.87)

$$\lambda^{(k+1)} := \lambda^{(k)} - \rho^{(k)}\ \Psi'(h(x^{(k)}))$$

inderdaad een betere schatting is voor de optimale Lagrangemultiplicatoren-vector $\hat{\lambda}$. Een illustratieve geometrische interpretatie van dit resultaat gegeven in Figuur 3.5.21 die in samenhang kan worden beschouwd met figuur 3.5.18. Uit de figuur blijkt duidelijk de versnelling van de convergentie die veroorzaakt wordt door het



Figuur 3.5.21: Geometrische interpretatie van de multiplicatoren iteratie-formule (3.5.87) bij constante waarde ρ van de schaalfactor.

medenemen van de Lagrange term $-\lambda^T h(x)$ in vergelijking met de gewone boetefunctiemethoden (i.e. het geval met $\lambda = \lambda^{(1)} \equiv 0$).

3.5.22. De multiplicatoreniteratieformule (3.5.87) kan ook worden opgevat als een toepassing van de gradiëntmethode met een vaste stapgrootte voor de oplossing van het aangevulde duale probleem (vgl (3.5.83)). Om deze opvatting te verifiëren moet de iteratieformule (3.5.87) worden herschreven in de vorm

$$\lambda^{(k+1)} := \lambda^{(k)} - \rho^{(k)} K_\Psi^{(k)} h(x^{(k)}) \tag{3.5.92}$$

ofwel (vgl (3.5.81))

$$\lambda^{(k+1)} := \lambda^{(k)} + \rho^{(k)} K_\Psi^{(k)} \nabla\tilde{d}(\lambda^{(k)}) \tag{3.5.93}$$

in welke uitdrukkingen $K_\Psi^{(k)}$ een diagonaalmatrix voorstelt met als diagonaal elementen de integraaluitdrukkingen

$$\int_0^1 \Psi_i'' (\tau h(x^{(k)}))d\tau \tag{3.5.94}$$

In het gebruikelijke geval van een kwadratische aanvulling geldt dat deze integralen gelijk zijn aan 1 zodat $K_\Psi^{(k)}$ gelijk is aan de eenheidsmatrix. In dat geval geldt dan dat de multiplicatoren-iteratieformule exact gelijk is aan

$$\lambda^{(k+1)} := \lambda^{(k)} + \rho^{(k)} \nabla\tilde{d}(\lambda^{(k)}) \tag{3.5.95}$$

De vaste stapgrootte $\rho^{(k)}$ kan als $\rho^{(k)} >> 1$ worden uitgelegd als een benadering van een stap volgens de methode van Newton. In het bijzonder in het geval van een kwadratische aanvulling kan dit aannemelijk worden gemaakt. In het beschouwde geval dat $\rho >> 1$ geldt dan bij benadering namelijk voor de Hessiaan van de (aangevulde) duale functie (vgl(3.5.83))

$$\begin{aligned}\nabla^2_{\lambda\lambda} d(\lambda) &:= -\nabla^T h(\tilde{x}(\lambda)) \,[\nabla^2_{xx} Q(\tilde{x}(\lambda),\lambda,\rho)]^{-1} \nabla h(\tilde{x}(\lambda)) \\ &:= -\nabla^T h(\tilde{x}(\lambda)) \,[\nabla^2_{xx} \mathcal{L}(\tilde{x}(\lambda),\lambda) + \rho \nabla h(\tilde{x}(\lambda)) \nabla^T h(\tilde{x}(\lambda)]^{-1} \nabla h(\tilde{x}(\lambda)) \\ &\approx -\frac{1}{\rho} I_{m\times m}\end{aligned} \tag{3.5.96}$$

Aangetoond kan worden (vgl [3.5.7]) dat in het geval de schaalfactoren $\rho^{(k)}$ naar ∞ gaan, de convergentie van de Lagrange multiplicatoren superlineair verloopt. De convergentiesnelheid van de multiplicatorenmethode

als duale gradiëntmethode hangt nauw samen (vgl pt 3.5.13) met het conditiegetal (d.i. verhouding van de grootste tot de kleinste eigenwaarde) van de Hessiaan van de (aangevulde) duale functie en dit conditiegetal nadert naar 1 naarmate ρ toeneemt. De convergentiesnelheid zal in het algemeen groter worden naarmate ρ toeneemt. Het bepalen van de waarde van de (aangevulde) duale functie door minimalisering van de aangevulde Lagrangefunctie wordt tegelijkertijd moeilijker omdat de conditie van dat probleem juist slechter wordt met toenemende waarde van ρ (juist als bij de boetefunctiemethoden). Het uiteindelijke resultaat ten aanzien van de convergentie van multiplicatoren (of aangevulde Lagrangefunctiemethoden) blijkt in de practijk zeer positief: In het algemeen is de convergentiesnelheid van multiplicatorenmethoden aanmerkelijker groter dan de convergentiesnelheid van de vergelijkbare boete- en barrièrefunctiemethoden.

3.5.23. De formulering van de aangevulde Lagrangefunctiemethoden is niet steeds dezelfde. Een vermeldenswaardige andere formulering die tevens een iets ander licht werpt op de methode als zodanig is de fomulering van Powell [3.5.23]. Deze definieerde een aangevulde Lagrangefunctie met behulp van de uitdrukking

$$Q(x,\theta,S) := f(x) + \tfrac{1}{2}(h(x) - \theta)^T S(h(x) - \theta) \qquad (3.5.97)$$

waarin S een $m \times m$ diagonaalmatrix met positieve (diagonaal)elementen σ_i voorstelt en θ een m-dimensionale parametervector. Uitschrijven van deze uitdrukking geeft

$$Q(x,\theta,S) := f(x) - \theta^T Sh(x) + \tfrac{1}{2}h^T(x) + Sh(x) + \tfrac{1}{2}\theta^T S\theta \qquad (3.5.98)$$

ofwel indien gebruik gemaakt wordt van de definitie

$$\lambda := S\theta \qquad (3.5.99)$$

de met (3.5.50) vergelijkbare uitdrukking

$$Q(x,\theta,S) = f(x) - \lambda^T h(x) + \tfrac{1}{2}\, h^T(x)Sh(x) + \tfrac{1}{2}\theta^T S\theta \qquad (3.5.100)$$

In deze formulering gelden θ en S als iteratieparameters (in plaats van λ en ρ) en voor deze suggereerde Powell in [3.5.23] de aanpassingsformules

$$\theta^{(k+1)} := \theta^{(k)} + h(x^{(k)}) \qquad (3.5.101)$$

en

$$S^{(k+1)} := \nu\, S^{(k)} \qquad \text{als } \|h(x^{(k+1)})\|_\infty > 0{,}25\|h(x^{(k)})\|_\infty$$
$$:= S^{(k)} \qquad \text{anders} \tag{3.5.102}$$

waarin

$$\|h(x^{(k+1)})\|_\infty = \max_i \left| h_i(x^{(k+1)}) \right| \tag{3.5.103}$$

Het is eenvoudig in te zien dat indien voor S geldt

$$S^{(k+1)} = S^{(k)} := \rho^{(k)} I_{m\times m} \tag{3.5.104}$$

dat de aanpassingsformule (3.5.101) van Powell equivalent is met de eerder genoemde multiplicatoreniteratieformule (3.5.87) van Hestenes.

Aangevulde Lagrangefuncties voor ongelijkheden

3.5.24. Voor het geval van minimaliseringsproblemen met ongelijkheden, d.w.z. problemen van het type

$$\min \{f(x) \mid g_i(x) \geq 0,\ i=1,\ldots,m\} \tag{3.5.105}$$

werd door Rockafellar [3.5.24] een modificatie van de hiervoor besproken aangevulde Lagrangefuncties gesuggereerd en wel in de vorm van de volgende definitie van de aangevulde Lagrangefunctie voor ongelijkheden

$$Q(x,\lambda,\rho) := f(x) + \bar{\Omega}(x,\lambda,\rho)$$
$$:= f(x) + \sum_{i=1}^{m} \bar{\omega}_i\,(g_i(x),\lambda_i,\rho) \tag{3.5.106}$$

waarin

$$\bar{\omega}_i(t,\lambda,\rho) := -\lambda_i t + \rho\psi_i(t) \qquad \text{als } t < \tau_i$$
$$:= -\lambda_i\tau_i + \rho\psi_i(\tau_i) \qquad \text{als } t \geq \tau_i \tag{3.5.107}$$

als τ_i het getal is waarvoor geldt

$$\frac{d\bar{\omega}_i}{dt} = -\lambda_i + \rho\psi_i'(\tau_i) = 0 \tag{3.5.108}$$

In het (meest gebruikte) geval dat (vgl(3.5.55))

$$\psi_i(t) = \tfrac{1}{2}t^2$$

gaat deze definitie voor $\bar{\omega}_i(t,\lambda,\rho)$ over in

$$\begin{aligned} \bar{\omega}_i(t,\lambda,\rho) &= -\lambda_i t + \tfrac{1}{2}\rho t^2 && \text{als } t < \frac{\lambda_i}{\rho} \\ &= -\tfrac{1}{2}\lambda_i^2/\rho && \text{als } t \geq \frac{\lambda_i}{\rho} \end{aligned} \qquad (3.5.109a)$$

waarmee

$$\begin{aligned} \bar{\omega}_i(g_i(x),\lambda,\rho) &:= -\lambda_i g_i(x) + \tfrac{1}{2}\rho\, g_i^2(x) && \text{als } g_i(x) < \frac{\lambda_i}{\rho} \\ &:= -\tfrac{1}{2}\lambda_i^2/\rho && \text{als } g_i(x) \geq \frac{\lambda_i}{\rho} \end{aligned}$$

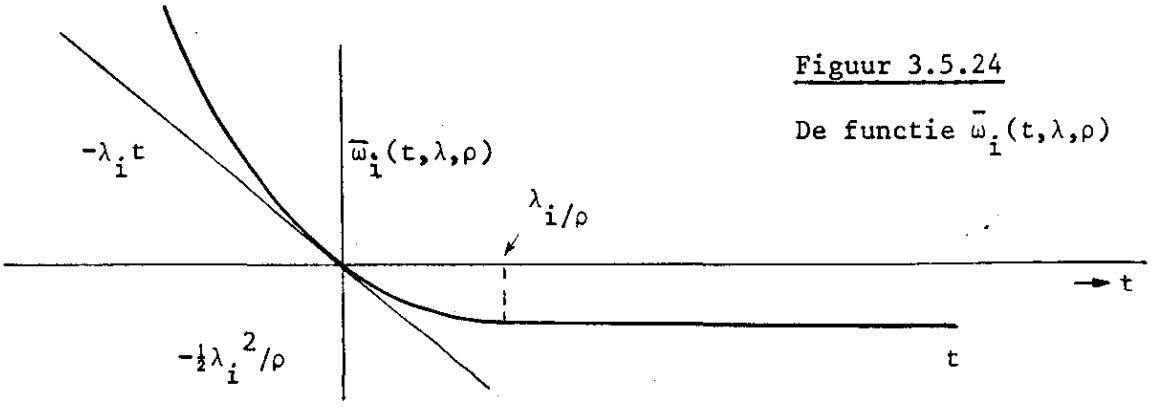
ofwel

$$\bar{\omega}_i(g_i(x),\lambda,\rho) := -\lambda_i \min\left[g_i(x),\frac{\lambda_i}{\rho}\right] + \tfrac{1}{2}\rho\{\min[g_i(x), \frac{\lambda_i}{\rho}]\}^2 \qquad (3.5.110)$$

of ook

$$\bar{\omega}_i(g_i(x),\lambda,\rho) := -\tfrac{1}{2}\lambda_i^2/\rho + \tfrac{1}{2}\rho\,\{\min[(g_i(x) - \frac{\lambda_i}{\rho}), 0]\}^2 \qquad (3.5.111)$$

Deze laatste formulering sluit duidelijk aan bij de formulering van Powell (vgl pt. 3.5.23). Een geometrische interpretatie van de functie $\omega_i(t,\lambda,\rho)$ is gegeven in Figuur 3.5.24



Figuur 3.5.24

De functie $\bar{\omega}_i(t,\lambda,\rho)$

De modificatie van Rockafellar bestaat dus daaruit dat in het niet-toegelaten gebied en bovendien in een klein deel van het toegelaten gebied in de directe omgeving van de ongelijkheidsbeperking gebruik wordt gemaakt van dezelfde aangevulde Lagrangefunctie als in het geval van uitsluitend gelijkheidsbeperkingen. In het toegelaten gebied op enige afstand verwijderd van de beperkingen is de gemodificeerde aangevulde Lagrangefunctie gelijk

aan de orginele objectfunctie minus een constante.

3.5.25. Opgemerkt kan worden dat de door Rockafellar gesuggereerde aangevulde Lagrangefunctie (3.5.106) voor ongelijkheden ook resulteert indien men de ongelijkheidsbeperkingen door invoering van nieuwe (positieve) slack-variabelen transformeert tot gelijkheidsbeperkingen, vervolgens de in het voorgaande besproken aangevulde Lagrangefunctie (3.5.50) vormt en daarna het minimum bepaalt van deze aangevulde Lagrangefunctie als functie van de slackvariabelen. Immers, maakt men gebruik van de transformatie van de ongelijkheden naar gelijkheden met behulp van de vervanging

$$g_i(x) \geq 0 \rightarrow g_i(x) - z_i^2 = 0 \tag{3.5.112}$$

en vormt men de aangevulde Lagrangefunctie

$$Q(x,z,\lambda,\rho) = f(x) + \sum_{i=1}^{m} [-\lambda_i(g_i(x) - z_i^2) + \rho\psi_i(g_i(x)-z_i^2)] \tag{3.5.113}$$

dan volgen onmiddellijk als voorwaarden voor een minimum van deze laatste functie beschouwd als een functie van de slack-variabelen z_i de vergelijkingen

$$[-\lambda_i + \rho\psi_i'(g_i(x)-z_i^2)] \quad (-2z_i) = 0 \qquad i = 1...m \tag{3.5.114}$$

Aan deze vergelijkingen wordt voldaan indien voor z_i gekozen worden de waarden

$$z_i = 0 \tag{3.5.115a}$$

of

$$z_i = \pm \sqrt{g_i(x) - \tau_i} \tag{3.5.115b}$$

waarbij de laatste waarde uiteraard alleen mogelijk is indien

$$g_i(x) \geq \tau_i \tag{3.5.116}$$

In het eerste geval resulteert voor de met de index i corresponderende aanvullingsterm in de aangevulde Lagrangefunctie

$$\bar{\omega}_i(g_i(x),\lambda_i,\rho) = -\lambda_i g_i(x) + \rho\psi_i(g_i(x)) \tag{3.5.117}$$

en in het tweede geval

$$\bar{\omega}_i(g_i(x),\lambda_i,\rho) = -\lambda_i\tau_i + \rho\psi_i(\tau_i) \qquad (3.5.118)$$

Aangezien het tweede een lagere waarde oplevert dan het eerste geval verdient dit tweede geval de voorkeur indien aan de voorwaarde (3.5.116) voldaan wordt. Wordt daar niet aan voldaan dan geeft de eerste uitdrukking de laagste waarde. Dit resultaat impliceert juist de in (3.5.107) gespecificeerde aangevulde Lagrangefunctie.

3.5.26. De in de voorgaande punten besproken gemodificeerde aangevulde Lagrangefunctie voor ongelijkheden kan op vrijwel dezelfde wijze worden toegepast in een algorithme van een aangevulde-Lagrangefunctiemethode als de eerder besproken aangevulde Lagrangefuncties voor uitsluitend gelijkheden. Het voornaamste onderdeel van een dergelijke algorithme is ook hier de onbeperkte minimalisering van de aangevulde Lagrangefunctie voor gegeven schattingen voor de iteratieparameters $\lambda^{(k)}$ en $\rho^{(k)}$, d.w.z. de oplossing van het probleem (vgl (3.5.86))

$$\min \{f(x) + \Omega(x,\lambda^{(k)},\rho^{(k)}) \mid x \in \mathbb{R}^n\} \qquad (3.5.119)$$

Als beginschatting kunnen voor de componenten van de Lagrangeparameter vector $\lambda^{(1)}$ gekozen worden de waarden (vgl(3.5.91))

$$\lambda_i^{(1)} := 0 \qquad (3.5.120)$$

terwijl voor $\rho^{(1)}$ gekozen kan worden een redelijk groot getal

$$\rho^{(1)} >> 1 \qquad (3.5.121)$$

Na de oplossing van het onbeperkte minimaliseringsprobleem (3.5.119) in de k-de iteratie kan de schattingen voor de Lagrange multiplicatoren $\lambda_i^{(k+1)}$ worden verbeterd met dezelfde multiplicatoren formules als voor het geval van uitsluitend gelijkheden (vgl(3.5.87),(3.5.89)), met dien verstande dat de schattingen voor de multiplicatoren nooit negatief mogen worden. Dit laatste kan worden bereikt door gebruik te maken van de aangepaste multiplicatoren formule

$$\lambda_i^{(k+1)} := \max \{0,(\lambda_i^{(k)} - \rho\psi_i'(g_i(x^{(k)}))\} \qquad (3.5.122)$$

ofwel in het geval van een kwadratische aanvulling (3.5.56)

$$\lambda_i^{(k+1)} := \max \{0,\ \lambda_i^{(k)} - \rho g_i(x^{(k)})\} \tag{3.5.123}$$

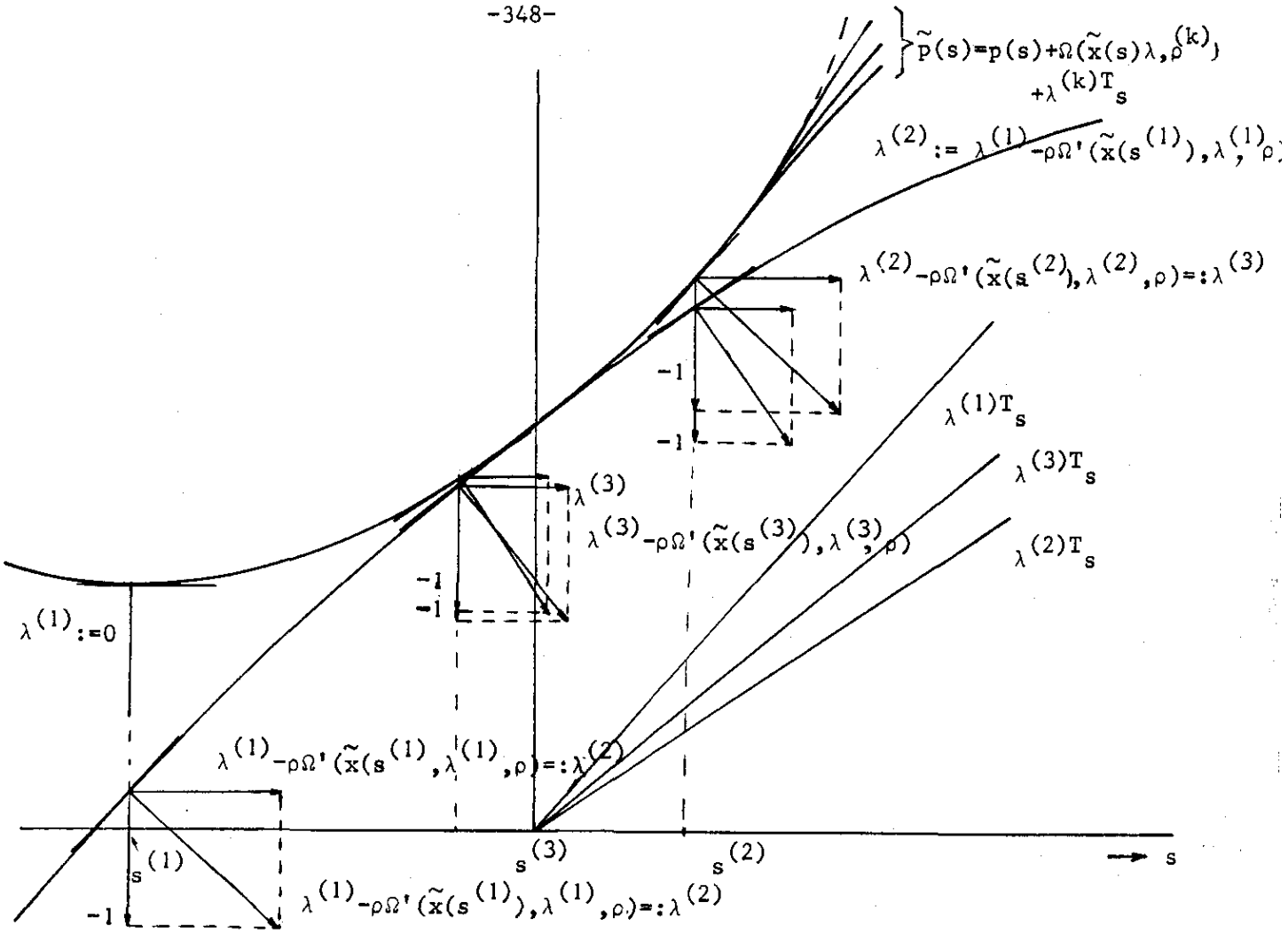
Voor de aanpassingsformule voor de schatting van de schaalfactor $\rho^{(k)}$ kan gebruik gemaakt worden van dezelfde aanpassingsformule als in het geval uitsluitend gelijkheden, d.w.z. de formule (3.5.90)

$$\rho^{(k+1)} := \nu\rho^{(k)}$$

3.5.27. Voor een geometrische interpretatie in de trant van Figuur 3.5.21 van de in het voorgaande punt besproken aangevulde-Lagrangefunctiemethode voor ongelijkheden is het nuttig de gemodificeerde aangevulde Lagrangefunctie (3.5.106) op te vatten als de Lagrangefunctie van een "gemodificeerde aangevulde objectfunctie die naar analogie van (3.5.67) de vorm krijgt

$$\begin{aligned} \bar{f}(x) &:= f(x) + \sum_{i=1}^{m} \psi_i(g_i(x)) && g_i(x) \le \tau_i \\ &:= f(x) + \sum_{i=1}^{m} [\lambda_i g_i(x) - \tfrac{1}{2}\frac{\lambda_i^2}{\rho}] && g_i(x) > \tau_i \end{aligned} \tag{3.5.124}$$

Vooral de tweede van deze uitdrukking en verdient enige aandacht omdat de aanvulling een lineaire functie is van $g_i(x)$ en geen kwadratische of anderszins convexe functie. De geometrische interpretatie van de aangevulde-Lagrangefunctiemethode met behulp van de met de primale functie die correspondeert met deze gemodificeerde aangevulde functie is weergegeven in Figuur 3.5.27. Het daar weergegeven resultaat verschilt slechts weinig van de eerder in Figuur 3.5.21 weergegeven geometrische interpretatie van de aangevulde-Lagrangefunctiemethode voor uitsluitend gelijkheden



Figuur 3.5.27: Geometrische interpretatie van de aangevulde-Lagrangefunctie-methode in het geval van ongelijkheden.

Exacte-boetefunctiemethoden

3.5.28. Een nadeel van de hiervoor besproken aangevulde-Lagrangefunctiemethoden isdat voor de oplossing van het minimaliseringsprobleem met beperkingen steeds een serie onbeperkte minimaliseringsproblemen moeten worden opgelost. Vandaar dat pogingen zijn gedaan om boetefuncties te bedenken waarvan het minimum samen zou vallen met het minimum van het beperkte minimaliseringsprobleem (zoals het geval is bij de aangevulde-Lagrangefuncties met de correcte waarde voor de Lagrangemultiplicator). De ontwikkeling van een dergelijke functie werd bestudeerd door een aantal auteurs waaronder in het bijzonder ook Fletcher ([3.5.9] t/m [3.5.14]) en Mårtenson [3.5.20]. Tot de eisen waaraan een dergelijke ideale boetefunctie moet voldoen behoort zeker de eis dat de resulterende functie een continue functie is van x, een aantal malen differentieerbaar, een afgeleide naar x heeft die gelijk is aan nul in het optimale punt $\hat{x}$, d.i.

$$\nabla\phi(\hat{x}) = 0 \qquad (3.5.125)$$

en die zo mogelijk een Hessiaan heeft die positief definiet is in het optimale punt, d.i.

$$\forall_{z \in \mathbb{R}^m} : z^T \nabla^2 \phi(\hat{x}) z > 0 \qquad (3.5.126)$$

Een functie die in het geval van een probleem van de gedaante (3.5.1)

$$\min \{ f(x) \mid h_j(x) = 0,\ j=1,\ldots,m \}$$

aan een aantal van deze eisen voldoet is de met het probleem corresponderende Lagrangefunctie met daarin de Lagrangemultiplicatorenvector vervangen door een vectorfunctie $\lambda(x)$, d.i. de functie

$$\phi(x) := f(x) - \lambda^T(x) h(x) \qquad (3.5.127)$$

Wordt gebruik gemaakt van de notatie $\nabla\lambda(x)$ voor de matrix met als kolommen de gradienten van de componenten $\lambda_j(x)$ van de vectorfunctie $\lambda(x)$,

$$\nabla\lambda(x) := [\nabla\lambda_1(x)\ \nabla\lambda_2(x),\ldots,\nabla\lambda_m(x)] \qquad (3.5.128)$$

dan volgt voor de gradiënt van de functie $\phi(x)$

$$\nabla\phi(x) := \nabla f(x) - \nabla\lambda(x) h(x) - N(x)\lambda(x) \qquad (3.5.129)$$

waarin dan (vgl(3.5.11)) $N(x)$ de matrix is van normalen

$$N(x) = [n_1(x),\ldots n_m(x)] := \nabla h(x) := [\nabla h_1(x),\ldots\ h_m(x)]$$

Wordt verder geëist dat de functie $\lambda(x)$ zodanig is dat in het optimale punt geldt

$$\lambda(\hat{x}) = \hat{\lambda} \qquad (3.5.130)$$

dan volgt onmiddellijk (met (3.5.2) en (3.5.129)) dat de gradiënt van de functie $\phi(x)$ inderdaad nul wordt in het optimale punt.
Differentiatie van de gradiënt (3.5.129) van de functie $\phi(x)$ herschreven in de vorm

$$\nabla\phi(x) = \nabla f(x) - \sum_{j=1}^{m} \nabla\lambda_j(x) h_j(x) - \sum_{j=1}^{m} n_j(x)\lambda_j(x) \qquad (3.5.131)$$

geeft als uitdrukking voor de Hessiaan van de functie $\phi(x)$

$$\nabla^2_{xx}\phi(x) = G(x) - \sum_{j=1}^{m} \nabla^2_{xx} \lambda_j(x)h_j(x) - \sum_{j=1}^{m} \nabla\lambda_j(x)n_j^T(x)$$

$$- \sum_{j=1}^{m} H_j(x)\lambda_j(x) - \sum_{j=1}^{m} n_j(x)\nabla\lambda_j^T(x) \qquad (3.5.132)$$

$$= \nabla^2_{xx}\mathcal{L}(x,\lambda) - \nabla\lambda(x)N^T(x) - N(x)\,\nabla\lambda^T(x)$$

$$- \sum_{j=1}^{m} \nabla^2_{xx}\lambda_j(x)h_j(x)$$

en in het optimale punt

$$\nabla^2_{xx}\phi(\hat{x}) = \nabla^2\mathcal{L}(\hat{x},\hat{\lambda}) - \nabla\lambda(\hat{x})N^T(\hat{x}) - N(\hat{x})\nabla\lambda^T(\hat{x}) \qquad (3.5.133)$$

3.5.29. Voor de verdere uitwerking van de uitdrukking voor de Hessiaan van $\phi(x)$ is het nodig eerst een keus te doen voor de vectorfunctie $\lambda(x)$. De overweging dat $\hat{\lambda}$ een oplossing is van het stelsel (3.5.2)

$$N(\hat{x})\hat{\lambda} - \nabla f(\hat{x}) = 0$$

leidt tot de suggestie om voor (x) te nemen de kleinste-kwadratenoplossing van het (meestal) overgedetermineerde stelsel

$$N(x)\lambda - \nabla f(x) = 0 \qquad (3.5.134)$$

of, equivalent voor iedere x de oplossing van het probleem

$$\min \{\|N(x)\lambda - \nabla f(x)\|^2 \mid \lambda \in \mathbb{R}^m\} \qquad (3.5.135)$$

Deze oplossing kan met de eerder besproken (pt 2.10.7) pseudo- of gegeneraliseerde inverse (vgl(3.1.62))

$$N^+(x) = \begin{pmatrix} (n_1^+(x))^T \\ (n_2^+(x))^T \\ (n_m^+(x))^T \end{pmatrix} \qquad (3.5.136)$$

worden weergegeven door (vgl(3.1.41))

$$\lambda(x) = N^+(x)\nabla f(x) \qquad (3.5.137a$$

of, compomentsgewijs,

$$\lambda_j(x) = (n^+(x))^T \nabla f(x) \tag{3.5.137b}$$

waarin, als, zoals meestal het geval, in het beschouwde punt de matrix $N(x)$ maximum rang heeft (en $m \leq n$),de pseudo-inverse $N^+(x)$ gegeven wordt door

$$N^+(x) = (N^T(x)N(x))^{-1}N^T(x) \tag{3.5.138}$$

Voor de gradiënt van de aldus gedefinieerde Lagrange-multiplicatorfunctie volgt met (3.5.137b)

$$\nabla\lambda_j(x) = [\frac{\partial}{\partial x} n_j^+(x)]^T \nabla f(x) + G(x)n_j^+(x) \tag{3.5.139}$$

in welke uitdrukking voor de matrix $[\frac{\partial}{\partial x} n_j^+(x)]$ op grond van (differentiatie van) de relatie (vgl(3.1.64)

$$\begin{aligned}(n_j^+(x))^T n_i(x) &= 0 \qquad i \neq j \\ &= 1 \qquad i = j\end{aligned} \tag{3.5.140}$$

geldt

$$n_i^T(x) [\frac{\partial}{\partial x} n_j^+(x)] = -(n_j^+(x))^T H_i(x) \tag{3.5.141}$$

De gradiënt van de functie $\lambda_j(x)$ in het optimale punt waar

$$\nabla f(\hat{x}) = \sum_{i=1}^{m} \hat{\lambda}_i n_i(\hat{x}) \tag{3.5.142}$$

wordt daarmee dan gelijk aan

$$\begin{aligned}\nabla\lambda_j(\hat{x}) &= \sum_{i=1}^{m} \hat{\lambda}_i [\frac{\partial}{\partial x} n_j^+(\hat{x})]^T n_i(\hat{x}) + G(\hat{x})n_j^+(\hat{x}) \\ &= -\sum_{i=1}^{m} \hat{\lambda}_i H_i(\hat{x})n_j^+(\hat{x}) + G(\hat{x})n_j^+(\hat{x}) \\ &= \nabla_{xx}^2 \mathcal{L}(\hat{x},\hat{\lambda})n_j^+(\hat{x})\end{aligned} \tag{3.5.143}$$

3.5.30. Substitutie van de keuze van $\lambda(x)$ met de hierboven afgeleide consequenties geeft in het bijzonder voor de functie $\phi(x)$ in het algemeen de uitdrukking

$$\begin{aligned}\phi(x) &:= f(x) - \nabla^T f(x)(N^+(x))^T h(x) \\ &= f(x) - h^T(x)N^+(x)\nabla f(x)\end{aligned} \tag{3.5.144}$$

welke in het optimale punt leidt tot

$$\phi(\hat{x}) = f(\hat{x}) - \lambda^T(\hat{x})h(\hat{x}) = f(\hat{x}) \tag{3.5.145}$$

Voor de gradiënt in het optimale punt (vgl(3.5.131) volgt

$$\nabla\phi(\hat{x}) = \nabla f(\hat{x}) - \nabla h(\hat{x})\lambda(\hat{x}) - \nabla\lambda(\hat{x})h(\hat{x}) = 0 \tag{3.5.146}$$

en idem voor de Hessiaan in het optimale punt (vgl(3.5.132))

$$\begin{aligned}\nabla^2_{xx}\phi(\hat{x}) &= \nabla^2_{xx}\mathcal{L}(\hat{x},\hat{\lambda}) - \sum_{j=1}^{m} \nabla^2_{xx}\mathcal{L}(\hat{x},\hat{\lambda})n_j^+(x)n_j(x) \\ &\quad - \sum_{j=1}^{m} n_j(x)(n_j^+(x))^T \nabla^2_{xx}\mathcal{L}(\hat{x},\hat{\lambda}) \\ &= \nabla^2_{xx}\mathcal{L}(\hat{x},\hat{\lambda}) - \nabla^2_{xx}\mathcal{L}(\hat{x},\hat{\lambda})(N^+(\hat{x}))^T N^T(\hat{x}) \\ &\quad - N(\hat{x})\,N^+(\hat{x})\,\nabla^2_{xx}\mathcal{L}(\hat{x},\hat{\lambda})\end{aligned} \tag{3.5.147}$$

Met gebruikmaking van de definitie van de projectie operator $P(x)$ op de deelruimte opgespannen door de normalen op de beperkingen

$$P(x) = N(x)\,N^+(x) \tag{3.5.148}$$

en de corresponderende definitie van de projectie operator $\bar{P}(x)$ op het orthogonale complement van de deelruimte opgespannen door de normalen (vgl(3.2.20))

$$\bar{P}(x) = I - N(x)N^+(x) \tag{3.5.149}$$

kan deze uitdrukking voor de Hessiaan van de functie $\phi(x)$ in het optimale punt nog worden herschreven in de conceptueel interessante vorm

$$\nabla^2_{xx}\phi(\hat{x}) = \bar{P}(\hat{x})\ \nabla^2_{xx}\mathcal{L}(\hat{x},\hat{\lambda})\bar{P}(\hat{x}) - P(\hat{x})\ \nabla^2_{xx}\mathcal{L}(\hat{x},\hat{\lambda})P(\hat{x}) \tag{3.5.150}$$

3.5.31. Dit resultaat illustreert duidelijk dat de gesuggereerde functie $\phi(x)$ (3.5.144) zelfs in het geval dat de Hessiaan van de Lagrangefunctie strikt positief definiet is, niet voldoet aan de eis van een positief definiete Hessiaan. Het probleem daarbij vormen de richtingen in de deelruimte opgespannen door de normalen. Om dit tekort van de functie $\phi(x)$ te corrigeren is het noodzakelijk de functie aan te vullen met een of meerdere termen die de convexiteit van de resulterende functie in de richting van de normalen vergroot. Een mogelijkheid daartoe biedt een aanvulling in de vorm van een kwadratische term van de gedaante

$$\Delta\phi(x) = \tfrac{1}{2}\rho h^T(x)\ S\ h(x) \tag{3.5.151}$$

waarin S een (symmetrische) positief definiete matrix voorstelt. De gradiënt van deze aanvullingsterm wordt gegeven door

$$\nabla(\Delta\phi(x)) = \rho\ N(x)\ S\ h(x) = \rho \sum_{j=1}^{m} n_j(x)\ (Sh(x))_j \tag{3.5.152}$$

en de Hessiaan door

$$\nabla^2(\Delta\phi(x)) = \rho \sum_{j=1}^{m} H_j(x)(Sh(x))_j + \rho\ N(x)S\ N^T(x) \tag{3.5.153}$$

De gradient en Hessiaan van de aangevulde functie

$$\begin{aligned}\tilde{\phi}(x) &= \phi(x) + \Delta\phi(x) \\ &= f(x) - h^T(x)N^+(x)\nabla f(x) + \tfrac{1}{2}\rho h^T(x)Sh(x)\end{aligned} \tag{3.5.154}$$

in het optimale punt $\hat{x}$ worden daarmee gelijk aan respectievelijk (vgl(3.5.148))

$$\nabla\tilde{\phi}(\hat{x}) = 0 \tag{3.5.155}$$

en (vgl(3.5.150))

$$\nabla^2\tilde{\phi}(\hat{x}) = \bar{P}(\hat{x})\nabla^2_{xx}\mathcal{L}(\hat{x},\hat{\lambda})\bar{P}(\hat{x}) - P(\hat{x})\nabla^2_{xx}\mathcal{L}(\hat{x},\hat{\lambda})P(\hat{x}) + \rho\ N(\hat{x})S\ N^T(\hat{x}) \tag{3.5.156}$$

Kiest men voor de matrix S een matrix van de vorm

$$S := N^{+}(\hat{x})(N^{+}(\hat{x}))^{T} \tag{3.5.157}$$

of, algemener een van x afhankelijke matrix

$$S(x) = N^{+}(x)(N^{+}(x))^{T} \tag{3.5.158}$$

dan volgt dat

$$\nabla^2\tilde{\phi}(\hat{x}) := \bar{P}(\hat{x})\ \nabla^2_{xx}\mathcal{L}(\hat{x},\hat{\lambda})\ \bar{P}(\bar{x}) \qquad P(\hat{x})\ (\rho I - \nabla^2\mathcal{L}(\hat{x},\hat{\lambda}))P(\hat{x}) \tag{3.5.159}$$

welke matrix door de schaalfactor ρ maar groot genoeg te kiezen steeds positief definiet gemaakt kan worden. De resulterende functie

$$\tilde{\phi}(x) = f(x) - h^{T}(x)\ N^{+}(x)\nabla f(x) - \tfrac{1}{2}\rho\ h^{T}(x)N^{+}(x)(N^{+}(x))^{T}h(x) \tag{3.5.160}$$

wordt in dat geval een (boete)functie die een onbeperkt minimum heeft in hetzelfde punt $\hat{x}$ waar het orginele probleem (5.3.1) een beperkt minimum heeft. Eenmaal toepassen van een onbeperkt minimaliseringsalgorithme (zoals besproken in het voorgaande hoofdstuk) is dan voldoende voor de oplossing van het minimaliseringsprobleem met beperkingen. Boetefuncties zoals $\tilde{\phi}(x)$ (3.5.160) (zie [3.5.14]) die deze laatste eigenschap hebben aangeduid als <u>exacte boetefuncties</u> . Methoden die gebaseerd zijn op het gebruik van dergelijke functies voor de oplossing van onbeperkte minimaliseringsproblemen worden overeenkomstig aangeduid als <u>exacte-boete-functiemethoden</u>.

3.5.32. Ondanks hun theoretische aantrekkelijkheid hebben de exacte boetefunctiemethoden minder toepassing gevonden in de practijk dan op theoretische gronden had mogen worden verwacht. Dit is onder meer het gevolg van een aantal practische bezwaren, waarvan de voornaamste zijn de moeilijkheid waarmee ongelijkheidsbeperkingen kunnen worden ingepast in de frame werk van de geschetste methode en in de tweede plaats de omstandigheid dat voor het evalueren van de functie zoals blijkt uit de formulering (3.5.160) eerste-orde informatie (d.i. de afgeleiden van de objectfunctie en beperkingen) nodig is en voor het evalueren van de gradiënt ervan ((3.5.131), (3.5.143) en (3.5.152)) tweede orde informatie (d.i. Hessiaan van objectfunctie en beperkingen) nodig is. In het geval dat deze eerste en tweede informatie gemakkelijk te verkrijgen is, is dit laatste bezwaar minder belangrijk, zij het dat in het geval van de besproken multiplicatorenmethoden tweede-

orde-methoden mogelijk zijn in het geval dat bij de exacte-boetefunctie methoden slechts eerste-orde methoden kunnen worden gebruikt. Overigens kan daarbij wel worden opgemerkt dat, omdat de Hessiaan van de exacte-boetefunctie in het optimale punt (vgl(3.5.159) alleen afhankelijk is van tweede orde informatie, het in de laatste geval wel mogelijk is om snel convergerende methoden toe te passen die gebruik maken van benaderingen van de tweede-orde afgeleiden van de exacte boete-functiemethoden. (zie b.v. [3.5.2][3.5.26]). In dit laatste geval blijft wel het bezwaar bestaat dat het voor iedere functieevaluatie de pseudo inverse $N^t(x)$ (3.5.136) van de matrix van normalen van de beperkingen moet worden bepaald hetgeen telkens een niet te verwaarlozen hoeveelheid rekenwerk met zich meebrengt.

Andere methoden die gebruik maken van Lagrange-functies

3.5.33. Naast de hiervoor geschetste multiplicatoren- (of aangevulde-Lagrangefunctie-) methoden en exacte-boetefunctiemethoden zijn er in de laatste paar jaren in de literatuur een groot aantal andere methoden gesuggereerd die gebruik maken van Lagrangefuncties die corresponderen met de beperkte minimaliseringsproblemen. Voor een deel zijn dit variaties en uitbreidingen van de besproken methoden (b.v.[3.5.21],[3.5.22][3.5.26],[3.5.27]) voor een ander deel ook methoden die gebruik maken van anders gedefinieerde Lagrangefuncties (b.v. [3.5.19]).
Daarnaast blijkt uit de literatuur ([3.5.13]) dat er ook grote belangstelling bestaat voor methoden die de minimaliseringsproblemen met niet-lineaire nevenvoorwaarden vervangen door een reeks kwadratische minimaliseringsproblemen met lineaire nevenvoorwaarden. In dat geval wordt bij de definitie van de kwadratische benadering veelvuldig gebruik gemaakt van een kwadratische benadering van de Lagrangefunctie in plaats van de objectfunctie. Uitwerking van diverse algorithmen voor deze kwadratische minimaliseringsproblemen geeft aanleiding tot procedures en uitdrukkingen die vergelijkbaar zijn of verwant blijken met de hiervoor in deze paragraaf besproken procedures en uitdrukkingen. Voor de details daarvan wordt verwezen naar de desbetreffende literatuur.

Literatuur

3.5.34. Meer informatie over de in deze paragraaf besproken methoden en technieken kan o.a. worden gevonden in de volgende publicaties.

[3.5.1]: Zie [1.1.1] Luenberger (1973)

[3.5.2]: Zie [1.1.4] Gill & Murray (1974)

[3.5.3]: Zie [1.1.7] Adby & Dempster (1974)

[3.5.4]: Zie [2.4.7] Luenberger (1969)

[3.5.5]: Bertsekas, D.P.: Combined primal-dual and penalty methods for constrained minimization SIAM J. Control, 13, (1975), pp 521-544

[3.5.6]: Bertsekas, D.P.: On penalty and multiplier methods for constrained minimization, in "Nonlinear Programming 2", (O.L. Mangasarian, R.R. Meyer and S.M. Robinson (eds)), Academic Press, New York (1975), pp 165-191

[3.5.7]: Bertsekas, D.P.: Multiplier methods, a survey, Automatica 12 (1976), pp 133-145

[3.5.8]:Buys, J.D.: Dual algorithms for constrained optimization problems, Doctoral Thesis University of Leiden, The Netherlands, 1972

[3.5.9]: Fletcher, R: A class of methods for nonlineair programming with termination and convergence properties, in: "Integer and nonlineair programming". (J. Abadie, (Ed)) North-Holland Publ. Co., Amsterdam (1970), pp 157-175

[3.5.10]: Fletcher, R and Lill, S.A: A class of methods for nonlinear programming II: Computational experience, in "Nonlinear Programming" (S. Rosen, O.L. Mangasarian and K. Ritter (eds)) Academic Press, New York (1971) pp. 67-92

[3.5.11]: Fletcher, R: A class of methods for nonlinear programming III: Rates of convergence, in "Numerical methods for nonlinear optimization" (F.A. Lootsma (Ed)), Academic Press, London, (1972) pp 371-381

[3.5.12]: Fletcher, R: An exact penaltyfunction for nonlinear programming with inequalities, Math. Progr. 5 (1973) pp 129-150

[3.5.13]: Fletcher, R: Methods related to Lagrangian functions, Ch VIII of [3.5.2] (1974) pp 219-239

[3.5.14]: Fletcher, R: An ideal penalty function for constrained optimization in "Nonlinear Programming 2" (O.L. Mangasarian, R.R. Meyer and S.M. Robinson (Eds)) Academic-Press, New York (1975) pp 121-163

[3.5.15]: Haarhoff, P.C. and Buys, J.D.: A new method for the optimization of a nonlinear function subject to nonlinear constraints, The Computer J. 13 (1970) pp 178-184.

[3.5.16]: Hestenes, M.R.: Multiplier and gradient methods, J. Opt. Theory & Appl. 4 (1969) pp 303-320.

[3.5.17]: Kort, B.W.: Rate of convergence of the method of multipliers with inexact minimization, in "Nonlinear Programming 2" (O.L. Mangasarian, R.R. Meyer and S.M. Robinson(Eds)) Academic Press, New York, (1975) pp 193-214.

[3.5.18]: Lill, S.A.: Generalization of an exact method for solving equality constrained problems to deal with inequalities, in "Numerical Methods for nonlinear optimization" (F.A. Lootsma (Ed)), Acadamic Press, London (1972), pp 383-393.

[3.5.19]: Mangasarian, O.L. Unconstrained Lagrangians in nonlinear programming, SIAM J. on Control 13 (1975) pp 772-791.

[3.5.20]: Mårtensson, K: A new approach to constrained function optimization, J. Opt. Theory & Appl., 12 (1973) pp 531-554.

[3.5.21]: Miele, A, Moseley, P.E., Levy A.V. and Coggins, G.M.: On the method of multipliers for mathematical programming problems, J. Opt. Theory & Appl 10 (1972) pp 1-33.

[3.5.22]: Pierre, D.A., and Lowe, M.J.: Mathematical programming via augmented Lagrangians. An introduction with computer programs, Addison-Wesley Publ. Co, Reading Mass. (1975).

[3.5.23]: Powell, M.J.D.: A method for nonlinear constraints in minimization problems, in: "Optimization" (R. Fletcher (Ed)) Academic Press, New York (1969) pp 283-298.

[3.5.24]: Rockafellar, R.T.: A dual approch to solving nonlinear programming problems by unconstrained minimization, Math Progr. 5 (1973) pp 354-373.

[3.5.25]: Rockafellar, R.T.: Augmented Lagrange multiplier functions and duality in nonconvex programming, SIAM J Control, 12 (1974) pp 268-285.

[3.5.26]: Rupp, R.D.: On the combination of the multiplier method of Hestenes and Powell with Newton's method, J. Opt. Theory & Appl, 15 (1975) pp 167-187.

[3.5.27]: Tripathi, S.S. and Narendra, K.S.: Constrained optimization problems using multiplier methods, J. Opt. Theory & Appl 9 (1972) pp 59-70.