NONLINMIN, een procedure voor het minimaliseren van niet-lineaire functies onder niet-lineaire nevenvoorwaarden

5. NONLINMIN, EEN PROCEDURE VOOR HET MINIMALISEREN VAN NIET-LINEAIRE FUNCTIES ONDER NIET-LINEAIRE NEVENVOORWAARDEN

door J.L. de Jong
(Technische Hogeschool Eindhoven)

5.1. Inleiding

De bestaande methoden voor het numeriek oplossen van niet-lineaire optimaliseringsproblemen met nevenvoorwaarden kunnen ruwweg worden ingedeeld in drie categorieën (vgl. LUENBERGER [1973]), t.w. boetefunctiemethoden, primale methoden en duale of Lagrangemethoden. De eerste categorie bestaat uit die methoden, waarbij geprobeerd wordt aan de beperkingen te voldoen door aan het overschrijden van de beperkingen "boeten" op te leggen, die opgeteld worden bij de te minimaliseren objectfunctie. De tweede categorie betreft die methoden die iteratief nieuwe en betere schattingen van het optimum zoeken langs speciale aan de beperkingen aangepaste zoekrichtingen. De derde categorie omvat die methoden die expliciet gebruik maken van de eigenschappen van de Lagrangefunctie van het niet-lineaire minimaliseringsprobleem.

Voor het oplossen van niet-lineaire minimaliseringsproblemen werd enige jaren geleden door F.A. Lootsma de Algolprocedure MINFUN (LOOTSMA [1972]) ontwikkeld waarin diverse boetefunctiemethoden zijn gerealiseerd. Min of meer in navolging daarvan werd door C.J.B. Dirkx en de schrijver in 1975 een begin gemaakt met de ontwikkeling van de in deze voordracht te bespreken procedure NONLINMIN (DIRKX [1975]), waarbinnen op analoge wijze een aantal verschillende methoden uit de tweede categorie zijn samengebracht. Bij deze ontwikkeling werd in het bijzonder gestreefd naar een opzet waarbij enerzijds de gebruiker de keuzemogelijkheid heeft uit een aantal verschillende algoritmen, terwijl anderzijds deze algoritmen zoveel mogelijk gebruik maken van dezelfde hulpprocedures. Een dergelijke opzet heeft als extra voordeel dat met slechts kleine aanpassingen andere soortgelijke algoritmen kunnen worden ingepast en dat nieuwe ideeën kunnen worden uitgeprobeerd op een goed vergelijkbare manier.

De procedure NONLINMIN is op dit moment nog slechts beperkt operationeel. Gereed zijn alle (hulp-)procedures waaruit de procedure NONLINMIN zelf weer is opgebouwd. Sommige van deze bestaan echter nog slechts in hun meest elementaire vorm. Gegeven de modulaire opbouw van het geheel, is het mogelijk deze (hulp-)procedures nog individueel te verfijnen en daarmee de totale procedure naar wens en behoefte uit te bouwen.

In de hierna te geven beschrijving zal eerst worden ingegaan op de aan de totale procedure ten grondslag liggende algoritme. Vervolgens zal aandacht worden besteed aan de theoretische achtergrond van enkele stappen binnen deze algoritme en tenslotte zal een korte discussie worden gewijd aan de practische uitvoering van diezelfde stappen. Voor programmatische details, listings etc. van de procedure moet worden verwezen naar een in de toekomst nog te schrijven handleiding voor het gebruik van de procedure.

5.2. Probleemstelling en algoritme

5.2.1. Probleem, notatie en definities

De hierna te beschrijven Algolprocedure NONLINMIN heeft betrekking op het oplossen van niet-lineaire optimaliseringsproblemen van de vorm

(5.2.1) $$\min\{f(x) \mid c_j(x) = 0,\ j = 1,\dots,m_1;\ c_j(x) \geq 0,\ j = m_1+1,\dots,m,\ x \in \mathbb{R}^n\}$$

waar $f\colon \mathbb{R}^n \to \mathbb{R}^1$ en $c_j\colon \mathbb{R}^n \to \mathbb{R}^1$, $j = 1,\dots,m$. Verondersteld wordt dat zowel $f \in C^1$ als $c_j \in C^1$, $j = 1,\dots,m$. Een rol bij de beschrijving hierna spelen de volgende probleemgrootheden, genoteerd in een grotendeels uit GILL & MURRAY [1974] overgenomen notatie:

$\hat{x}$: optimale punt, oplossing van het probleem (5.2.1)
$x^{(k)}$: k-de benadering voor $\hat{x}$, startpunt (k+1)-de iteratiestap
$g := \nabla f(x) = (\frac{\partial f}{\partial x_1} \dots \frac{\partial f}{\partial x_n})^T$: gradiënt van de objectfunctie
$g^{(k)}$: gradiënt van de objectfunctie in het punt $x^{(k)}$
$d^{(k)}$: zoekrichting vanuit $x^{(k)}$
$\alpha^{(k)}$: stapgrootte(factor) in de richting $d^{(k)}$
$G(x) := \nabla^2 f(x)$: Hessiaan van de objectfunctie
$a_j^{(k)} := \nabla c_j(x^{(k)})$: gradiënt van de j-de beperking in het punt $x^{(k)}$
$A^{(k)} := [a_1^{(k)} \dots a_m^{(k)}]$: $n \times m$-matrix met als kolommen de gradiënten van de beperkingen in het punt $x^{(k)}$

$N^{(k)}$ $:= [n_1^{(k)} \dots n_q^{(k)}]$: n × q-matrix met als kolommen q gradiënten (of normalen) van actieve beperkingen in $x^{(k)}$ die samen een basis vormen voor de deelruimte opgespannen door de normalen van alle actieve beperkingen

$N^{(k)+} := (N^{(k)T}N^{(k)})^{-1}N^{(k)T}$: pseudo-of gegeneraliseerde inverse van $N^{(k)}$

$P^{(k)}$ $:= N^{(k)}N^{(k)+}$: matrix van projectie op de normalen van de actieve beperkingen

$\bar{P}^{(k)}$ $:= I - N^{(k)}N^{(k)+}$: matrix van projectie op de doorsnede van de raakvlakken aan de actieve beperkingen

$\lambda^{(k)}$: j-de component van een benadering voor de oplossing van het stelsel $N^{(k)}\lambda - \nabla f(x^{(k)}) = 0$, of, equivalent, benadering voor de Lagrange multiplicator corresponderend met de j-de beperking in het optimale punt in de doorsnede van de actieve beperkingen in het punt $x^{(k)}$.

Van belang voor het navolgende zijn ook de volgende begrippen:

- Een punt x heet een *toegelaten* (feasible) *punt* als in x aan alle beperkingen voldaan wordt, d.w.z. indien geldt

$$c_j(x) = 0 \quad (j=1,\dots,m_1) \quad c_j(x) \geq 0 \quad (j=m_1+1,\dots,m).$$

- Een beperking $c_j(x)$ heet in het punt x respectievelijk

actief	als	$c_j(x) = 0$ en $j \in \{1,\dots,m\}$
passief	als	$c_j(x) > 0$ en $j \in \{m_1+1,\dots,m\}$
overschreden	als	$c_j(x) \neq 0$ en $j \in \{1,\dots,m_1\}$ of $c_j(x) < 0$ en $j = \{m_1+1,\dots,m\}$.

- Coördinaatbeperkingen van de vorm

$$a_i \leq x_i \quad \text{of} \quad x_i \leq b_i ,$$

waarvan de gradiënten de corresponderende eenheidsvectoren zijn, worden *triviale beperkingen* genoemd.

5.2.2. Basisalgoritme

Bij de meeste primale methoden bestaat de eerste stap naar de oplossing van het probleem (5.2.1) uit het bepalen van een toegelaten startpunt. In dat punt wordt gekeken welke beperkingen actief zijn, waarna de objectfunctie wordt geminimaliseerd over de doorsnede van de raakvlakken aan deze actieve beperkingen. De passieve beperkingen leveren een begrenzing van het zoekgebied. Er zijn dan 2 mogelijkheden: ofwel het minimum wordt aangenomen op de rand van het zoekgebied, ofwel het minimum wordt gevonden in de doorsnede van de raakvlakken aan de actief veronderstelde beperkingen. In het nieuw gevonden punt worden de beperkingen geëvalueerd en wordt nagegaan of het gevonden punt een toegelaten punt is. Zo nee, dan wordt uitgaande van het gevonden punt een nieuw toegelaten punt gegenereerd (dit proces wordt het *restauratieproces* genoemd). Zo ja, en dat is steeds (bij benadering) het geval bij uitsluitend lineaire beperkingen, dan wordt opnieuw vooraan begonnen. Kan daarna geen verbetering meer worden bewerkstelligd doordat in het betreffende punt het minimum gevonden wordt van de restrictie van de objectfunctie tot de doorsnede van de actieve beperkingen, dan wordt gekeken of het mogelijk is de objectfunctie verder te verlagen door het passief veronderstellen van een van de actieve beperkingen. Als dit niet kan dan is de oplossing van het probleem gevonden; als het wel kan, dan wordt verder gezocht over de doorsnede van de nieuwe verzameling actieve beperkingen.

Dit verbaal beschreven proces wordt in meer detail weergegeven door de volgende basisalgoritme voor de primale methoden in de procedure NONLINMIN:

(0) Zet $\bar{x}^{(0)}$:= gegeven startpunt; zet k := 0.

(i) Met $\bar{x}^{(k)}$ als uitgangspunt bepaal een toegelaten punt $x^{(k)}$, bepaal welke beperkingen actief zijn in $x^{(k)}$ en evalueer de corresponderende normalen $a_j^{(k)} := \nabla c_j(x^{(k)})$.

(ii) Bepaal de functiewaarde $f(x^{(k)})$, de gradiënt $g^{(k)}$ en zo nodig de Hessiaan $G^{(k)}$ van de objectfunctie in $x^{(k)}$.

(iii) Bepaal een met de actieve beperkingen corresponderende matrix $N^{(k)}$ van lineair onafhankelijke normalen en de restrictie van de gradiënt tot de daarmee corresponderende doorsnede van raakvlakken, d.i. of de geprojecteerde gradiënt (vgl. (5.3.7))

$$\bar{P}^{(k)}\nabla f(x^{(k)})$$

of de gereduceerde gradiënt (vgl. (5.3.24))

$$\bar{R}^{(k)}\nabla f(x^{(k)}).$$

(iv) Ga na of het punt $x^{(k)}$ het optimale punt is in de huidige actieve set; zo nee, dan ga door naar stap (v) en start een nieuwe iteratie; zo ja dan evalueer het teken van de Lagrangemultiplicatoren die corresponderen met de actieve ongelijkheidsbeperkingen en die voldoen aan (5.3.16)

$$\lambda_j^{(k)} := ((N^{(k)T}N^{(k)})^{-1}N^{(k)T}\nabla f(x^{(k)}))_j$$

of (5.3.30)

$$\lambda_j^{(k)} := (B^{(k)-T}\nabla_{x_B} f(x^{(k)}))_j.$$

Als geen van deze $\lambda_j^{(k)}$ (met $j \in \{m_1+1,\ldots,m\} \cap I_A(x^{(k)})$ waar $I_A(x^{(k)})$ de indexverzameling is van de actieve beperkingen in $x^{(k)}$) negatief is dan *klaar*; zo niet, dan veronderstel een van de met de negatieve $\lambda_j^{(k)}$'s corresponderende beperkingen passief en ga terug naar stap (iii).

(v) Bepaal een nieuwe zoekrichting $d^{(k)}$.

(vi) Bepaal een maximale stapgroottefactor $\bar{\alpha}^{(k)}$.

(vii) Bepaal een stapgroottefactor $\alpha^{(k)}$ met $0 \leq \alpha^{(k)} \leq \bar{\alpha}^{(k)}$ zodat

(5.2.2) $$f(x^{(k)}+\alpha^{(k)}d^{(k)}) < f(x^{(k)}).$$

(viii) Bepaal een nieuw punt

(5.2.3) $$\bar{x}^{(k+1)} := x^{(k)} + \alpha^{(k)}d^{(k)},$$

zet k := k+1 en ga terug naar stap (i).

Een flowdiagram van de op deze algoritme gebaseerde procedure NONLINMIN is weergegeven in Figuur 5.4.1 op blz.113. In deze figuur is bij alle subprocedures aangegeven met welke stap uit de hier gegeven algoritme deze corresponderen. In de volgende paragraaf wordt nader ingegaan op de details van de stappen (i), (v) en (vi).

Van belang om op te merken is dat de in stap (iv) tot uiting komende strategie voor de behandeling van de beperkingen in de literatuur bekend is als de *actieve-set-strategie* (vgl. GILL & MURRAY [1974], p. 50). Deze strategie bestaat daaruit dat men veronderstelt dat de beperkingen die actief zijn in $x^{(k)}$ ook daarna actief blijven en waarbij eerst dan een beperking uit de verzameling van actieve beperkingen wordt verwijderd wanneer het minimum over de doorsnede van de raakvlakken aan de actieve beperkingen is bereikt. Men voorkomt hiermee het zg. "zigzagging"- of "jamming"-verschijnsel, waarbij steeds tussen twee of meer dezelfde beperkingen op en neer wordt gesprongen en dat zelfs convergentie naar een niet optimaal punt tot gevolg kan hebben.

5.3. Theoretische details

5.3.1. Bepaling van een toegelaten startpunt

Voor de bepaling van een toegelaten startpunt zijn een aantal methoden bekend in de literatuur: Indien alle beperkingen lineair zijn is het mogelijk gebruik te maken van de z.g. Fase 1-methode voor lineaire programmering (vgl. LUENBERGER [1973]). Indien ook niet-lineaire beperkingen aanwezig zijn, is het mogelijk op analoge wijze te werk te gaan door de minimalisering van in te voeren artificiële variabelen dan wel door de minimalisering van een kwadratische of andersoortige boetefunctie. (Voor deze onbeperkte minimaliseringsproblemen kan eventueel het bestaande programma zelf worden gebruikt). Voor het bepalen van een toegelaten startpunt in de procedure NONLINMIN werd een speciale algoritme ontwikkeld, die tevens kon worden gebruikt voor het bepalen van toegelaten punten in latere fasen van het minimaliseringsproces. Deze algoritme bestaat daaruit dat in een aantal opvolgende stappen telkens de minimum-norm kleinste-kwadraten oplossing wordt bepaald van het stelsel lineaire vergelijkingen gegenereerd door linearisatie van de lokale overschreden (eng.: violated) en actieve beperkingen in het laatste gevonden punt. Wordt dit stelsel

$$(5.3.1)\qquad c_j(\bar{x}^{(k)}) + \nabla c_j^T(\bar{x}^{(k)})(x-\bar{x}^{(k)}) = 0, \qquad j \in (I_A(\bar{x}^{(k)}) \cup I_V(\bar{x}^{(k)})),$$

herschreven in de vorm

$$(5.3.2)\qquad \bar{A}^T\Delta x + \bar{c} = 0$$

dan kan de minimum-norm kleinste-kwadraten oplossing met behulp van de pseudo-inverse $(\bar{A}^T)^+$ van de matrix $\bar{A}^T$ worden weergegeven als

$$(5.3.3)\qquad \Delta\hat{x} = -\ (\bar{A}^T)^+\bar{c}.$$

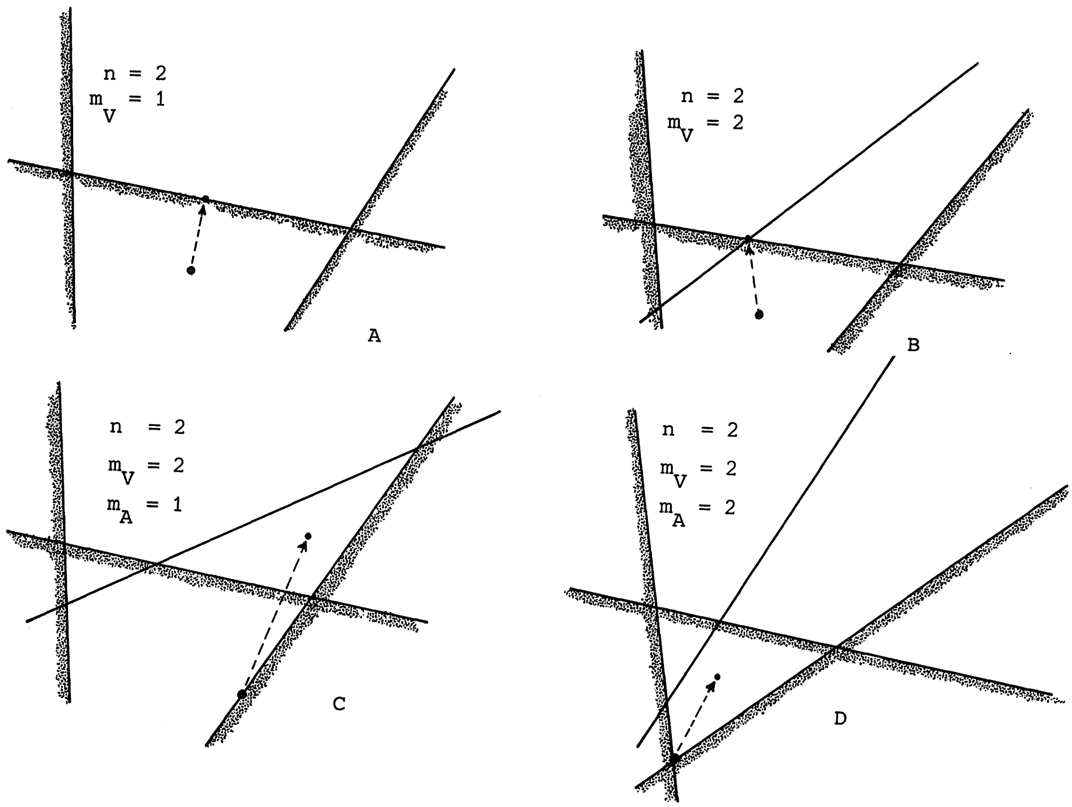
Als $\bar{A}^T$ een (p×n)-matrix is en verondersteld wordt dat $\bar{A}^T$ maximale rang bezit dan volgt bij uitwerking van de pseudo-inverse in het geval dat $p < n$ (onderbepaald stelsel) dat

$$(5.3.3) \qquad \Delta\hat{x} = - \bar{A}(\bar{A}^T\bar{A})^{-1}c$$

en in het geval dat $p \geq n$ (overbepaald stelsel) dat

$$(5.3.4) \qquad \hat{x} = - (\bar{A}\bar{A}^T)^{-1}\bar{A}c.$$

In het eerste geval is de correctie een lineaire combinatie van de gradiënten van de beperkingen, in het tweede geval is de correctie zodanig dat slechts in kleinste-kwadraten zin aan de vergelijkingen wordt voldaan. Ten grondslag aan dit proces 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 algoritme is geïllustreerd in Figuur 5.3.1 waarin voor het geval van $\mathbb{R}^2$ enkele verschillende mogelijkheden met betrekking tot de overschreden (m_V) en actieve (m_A) beperkingen zijn weergegeven.



Figuur 5.3.1. Voorbeelden van enige mogelijkheden voor niet toegelaten punten.

5.3.2. Eerste-orde zoekrichtingen

Uitgangspunt voor de afleiding van zowel eerste- als hoger-orde zoekrichtingen voor primale methoden is de lokale linearisatie van de actieve beperkingen in het toegelaten punt $x^{(k)}$ waarmee het originele probleem (5.2.1) overgaat in het probleem

$$(5.3.5) \qquad \min\{f(x) \mid N^{(k)T}(x-x^{(k)}) = 0,\ x \in \mathbb{R}^n\}.$$

Voor dit basisprobleem bestaan drie equivalente formuleringen die aanleiding vormen voor evenzovele formuleringen voor eerste- en hoger-orde zoekrichtingen. De eerst-orde zoekrichtingen volgen bij minimalisering van de gelineariseerde objectfunctie over een hyperbol. De meest bekende eerste-orde zoekrichting volgt als oplossing van het probleem

$$(5.3.6) \qquad \min\{f(x^{(k)}+\nabla^T f(x^{(k)})(x-x^{(k)}) \mid N^{(k)T}(x-x^{(k)}) = 0,\ \|x-x^{(k)}\| \le r\}$$

en staat in de literatuur bekend als de *geprojecteerde-gradient* (van ROSEN [1960])

$$(5.3.7) \qquad d^{(k)} := -(I-N^{(k)}(N^{(k)T}N^{(k)})^{-1}N^{(k)T})\nabla f(x^{(k)}) =: -\bar{P}^{(k)}\nabla f(x^{(k)}).$$

Een andere formulering van het basisprobleem (5.3.5) resulteert, wanneer gebruik wordt gemaakt van een $n \times (n-q)$-matrix $Z^{(k)}$ waarvan de kolommen een orthonormale basis vormen van het orthogonale complement van de deelruimte opgespannen door de kolommen van $N^{(k)}$, d.i. de normalen van de actieve beperkingen in $x^{(k)}$. Voor deze matrix $Z^{(k)}$ geldt

$$(5.3.8) \qquad N^{(k)T}Z^{(k)} = 0_{q\times(n-q)}$$

en

$$(5.3.9) \qquad Z^{(k)T}Z^{(k)} = I_{(n-q)\times(n-q)}.$$

Het basisprobleem (5.3.5) kan daarmee worden geschreven als een onbeperkt minimaliseringsprobleem

$$(5.3.10) \qquad \min\{f(x^{(k)}+Z^{(k)}w) \mid w \in \mathbb{R}^{n-q}\}$$

en het aan (5.3.6) analoge probleem

$$(5.3.11) \qquad \min\{f(x^{(k)}) + \nabla^T f(x^{(k)}) Z^{(k)} w \mid \|w\| < r\}$$

levert dan in termen van w de zoekrichting

$$(5.3.12) \qquad d_w^{(k)} := - Z^{(k)T} \nabla f(x^{(k)})$$

en in termen van het originele probleem de zoekrichting

$$(5.3.13) \qquad d^{(k)} := - Z^{(k)} Z^{(k)T} \nabla f(x^{(k)}).$$

Aangezien zowel

$$(I-N^{(k)}(N^{(k)T}N^{(k)})^{-1}N^{(k)T})[N^{(k)}|Z^{(k)}] = [0|Z^{(k)}]$$

als

$$Z^{(k)}Z^{(k)T}[N^{(k)}|Z^{(k)}] = [0|Z^{(k)}]$$

volgt onmiddellijk dat

$$(5.3.14) \qquad Z^{(k)}Z^{(k)T} = (I-N^{(k)}(N^{(k)T}N^{(k)})^{-1}N^{(k)}) =: \bar{P}^{(k)} .$$

Dit laatste impliceert dat beide uitdrukkingen (5.3.7) en (5.3.13) exact dezelfde zoekrichting representeren. De formulering (5.3.10) van het basisprobleem (5.3.5) werd vooral door GILL & MURRAY [1974] benut voor de afleiding van een aantal verschillende hoger-orde zoekrichtingen. In verband daarmee wordt de formulering (5.3.13) van de geprojecteerde gradiënt met behulp van de projectiematrix $Z^{(k)}Z^{(k)T}$ hierna ook wel aangeduid als de geprojecteerde gradiënt van GILL en MURRAY of wel de *Gill-Murray-gradiënt*.

Het zoekproces met de geprojecteerde gradiënt (van Rosen of van Gill en Murray) stopt indien geen nieuwe zoekrichting meer kan worden gegenereerd d.i. als

$$(5.3.15) \qquad d^{(k)} := - \bar{P}^{(k)} \nabla f(x^{(k)}) = - \nabla f(x^{(k)}) + N^{(k)}(N^{(k)T}N^{(k)})^{-1}N^{(k)T}\nabla f(x^{(k)}) = 0.$$

Daaruit volgt dan onmiddellijk dat in dat geval

(5.3.16) $\lambda^{(k)} = (N^{(k)T}N^{(k)})^{-1}N^{(k)T}\nabla f(x^{(k)}) = N^{(k)+}\nabla f(x^{(k)})$.

Een derde formulering van het basisprobleem (5.3.5) resulteert, indien gebruik wordt gemaakt van de mogelijkheid om de q lineair onafhankelijke gelijkheidsvoorwaarden

(5.3.17) $N^{(k)T}x = N^{(k)T}x^{(k)} =: b^{(k)}$

te gebruiken om q variabelen op te lossen als functie van de n-q overige variabelen. De vector x wordt in dat geval gesplitst in een q-vector x_B die de afhankelijke of *basisvariabelen* bevat en een (n-q)-vector x_D die de onafhankelijke of *niet-basisvariabelen* bevat. De matrix $N^{(k)T}$ wordt analoog gesplitst in een reguliere $q \times q$-matrix $B^{(k)}$ en een $q \times (n-q)$-matrix $D^{(k)}$. Het stelsel (5.3.17) is dan te schrijven als

(5.3.18) $B^{(k)}x_B + D^{(k)}x_D = b^{(k)}$

waaruit direct volgt dat

(5.3.19) $x_B = B^{(k)-1}b^{(k)} - B^{(k)-1}D^{(k)}x_D$.

De objectfunctie van het probleem (5.3.5) is daarmee te schrijven als een functie van alleen de vector x_D

(5.3.20) $f(x) = f(x_B,x_D) = f(B^{-1}b-B^{-1}Dx_D,x_D) =: \tilde{f}(x_D)$

(in welke uitdrukking, evenals in de hierna volgende, de indices (k) achterwege werden gelaten). Het basisprobleem (5.3.5) kan opnieuw worden beschouwd als een onbeperkt minimaliseringsprobleem, ditmaal in de variabele x_D (de nevenvoorwaarde (5.3.18) fungeert in dit geval uitsluitend als een voorschrift voor de bepaling van de afhankelijke variabele x_B)

(5.3.21) $\min\{\tilde{f}(x_D) \mid x_D \in \mathbb{R}^{n-q}\}$.

Minimalisering van de linearisering van de objectfunctie $\tilde{f}(x_D)$ rond $x_D^{(k)}$ over een hyperbol in $\mathbb{R}^{n-q}$ geeft als zoekrichting de negatieve gradiënt van de functie $\tilde{f}(x_D)$ naar x_D. Deze gradient, die in de literatuur bekend staat als de *gereduceerde gradient* (vgl. WOLFE [1963]) wordt gegeven door

$$(5.3.22)\qquad \nabla_{x_D} f = -D^T B^{-T} \nabla_{x_B} f + \nabla_{x_D} f.$$

Als zoekrichting in de originele ruimte volgt daaruit

$$(5.3.23)\qquad \begin{aligned} d_B &:= -B^{-1}D(-\nabla_{x_D}\tilde{f}) = -B^{-1}DD^TB^{-T}\nabla_{x_B} f + B^{-1}D\nabla_{x_D} f \\ d_D &:= -\nabla_{x_D}\tilde{f} = D^TB^{-T}\nabla_{x_B} f - \nabla_{x_D} f \end{aligned}$$

of in matrix-vector notatie

$$(5.3.24)\qquad d := -\begin{bmatrix} B^{-1}DD^TB^{-T} & -B^{-1}D \\ -D^TB^{-T} & I \end{bmatrix} \begin{bmatrix} \nabla_{x_B} f \\ \nabla_{x_D} f \end{bmatrix} =: -\bar{R}^{(k)}\nabla f(x^{(k)})$$

of, nog anders, met de matrix

$$(5.3.25)\qquad M := \begin{bmatrix} -B^{-1}D \\ I \end{bmatrix}$$

uitgeschreven

$$(5.3.26)\qquad d := -MM^T\nabla f.$$

Deze laatste uitdrukking vertoont grote overeenkomst met de uitdrukking voor de Gill-Murray-gradiënt (5.3.13), een overeenkomst die nog wordt versterkt door de observatie dat

$$(5.3.27)\qquad N^{(k)T}M = [B|D] \begin{bmatrix} -B^{-1}D \\ I \end{bmatrix} = O_{q\times(n-q)}.$$

Het zoekproces met de gereduceerde gradiënt stopt indien

$$(5.3.28)\qquad \nabla_{x_D}\tilde{f} = -D^TB^{-T}\nabla_{x_B} f + \nabla_{x_D} f = 0.$$

In dat geval geldt

$$(5.3.29) \qquad -\begin{bmatrix} B^T \\ \\ D^T \end{bmatrix} B^{-T}\nabla_{x_B} f + \begin{bmatrix} \nabla_{x_B} f \\ \\ \nabla_{x_D} f \end{bmatrix} = 0$$

en daaruit volgt weer

$$(5.3.30) \qquad \lambda = B^{-T}\nabla_{x_B} f \; .$$

De gereduceerde gradiënt werd voor het eerst als zoekrichting gesuggereerd door WOLFE [1963]. De (gegeneraliseerde) gereduceerde-gradiënt methode die er op gebaseerd is werd vooral verder ontwikkeld door ABADIE & CARPENTIER [1969]. De hierboven weergegeven afleiding wijkt af van de gebruikelijke doordat alleen de actieve beperkingen in de matrix $N^{(k)T} = [B^{(k)} | D^{(k)}]$ worden opgenomen. In de gebruikelijke situatie worden, in navolging van de simplexmethode voor lineaire programmering, alle passieve beperkingen met behulp van "slack"-variabelen actief gemaakt, waarna als uitgangsprobleem fungeert een probleem van de vorm

$$(5.3.31) \qquad \min\{f(x) \mid A^T x - b = 0, \;\; x \geq 0\}.$$

In deze laatste situatie worden alleen die componenten van de gereduceerde gradiënt gebruikt als zoekrichting die geen overschrijding van de coördinaatbeperkingen tot gevolg hebben. D.w.z. de niet-basiscomponenten van de zoekrichting worden gedefinieerd als

$$(5.3.32) \qquad \begin{aligned} \bar{d}_{D,i} &:= -(\nabla_{x_D}\tilde{f})_i \quad \text{als} \quad x_i \neq 0 \vee (\nabla_{x_D}\tilde{f})_i < 0 \\ &:= 0 \qquad\qquad\qquad\quad\;\; x_i = 0 \wedge (\nabla_{x_D}\tilde{f})_i > 0. \end{aligned}$$

Waarna als basiscomponenten dan volgen

$$(5.3.33) \qquad \bar{d}_{B,j} := -(B^{-1}D\bar{d}_D)_j .$$

Wanneer alle beperkingen lineair zijn dan zullen de verzamelingen van actieve beperkingen in opvolgende iteraties slechts in één (of geen) element verschillen en wel doordat of een eerdere passieve beperking actief wordt of een eerdere actieve beperking passief verondersteld wordt. Het is in dit geval niet nodig om de totale berekeningen die samenhangen met de matrices $\bar{P}^{(k)}$ (5.3.7), $Z^{(k)}Z^{(k)T}$ (5.3.14) en $\bar{R}^{(k)}$ (5.3.24) geheel opnieuw van voor af aan uit te voeren. Door diverse auteurs, waaronder o.a. ROSEN [1960] en GILL & MURRAY [1974] werden aanpassingsformules opgesteld die er voor zorgen dat een efficient gebruik gemaakt wordt van de aanwezige informatie over de actieve beperkingen die actief blijven.

5.3.3. Hoger-orde zoekrichtingen

Tweede-orde of Newton-zoekrichtingen kunnen worden afgeleid door gebruik te maken van tweede-orde benaderingen van de objectfunctie van het basisprobleem (5.3.5). Quasi-Newton-zoekinrichtingen kunnen daar weer van worden afgeleid door gebruik te maken van benaderingen in de uitdrukkingen van de Newton-zoekrichtingen. In deze paragraaf zullen alleen voor het meest simpele geval (positief definiete Hessiaan van de objectfunctie) twee tweede-orde of Newton-zoekrichtingen worden besproken om een indruk te geven van de mogelijkheden in dit verband. Voor meer gedetailleerde discussies over modificaties van deze zoekrichtingen en van daarvan afgeleide quasi-Newton-zoekrichtingen wordt de lezer verwezen naar elders (bv. GILL & MURRAY [1974], DE JONG [1976], DIRKX [1975]).

Analoog aan de situatie bij de eerste orde zoekrichtingen resulteren de eerder besproken verschillende formuleringen van het basisprobleem in evenzo vele verschillende formuleringen van Newton-zoekrichtingen. In de eerste, gebruikelijke formulering wordt de tweede-orde benadering van de objectfunctie rond het punt $x^{(k)}$ gegeven door de uitdrukking

$$(5.3.34) \qquad f(x^{(k)})+\nabla^T f(x^{(k)})(x-x^{(k)})+\tfrac{1}{2}(x-x^{(k)})^T G(x^{(k)})(x-x^{(k)}).$$

Het minimum $x^{(k+1)}$ van deze kwadratische functie onder de nevenvoorwaarde

$$(5.3.35) \qquad N^{(k)T}(x-x^{(k)}) = 0$$

kan worden gevonden met behulp van de noodzakelijke voorwaarde dat in het optimale punt $x^{(k+1)}$ moet gelden

$$(5.3.36) \qquad \nabla f(x^{(k+1)}) - N^{(k)}\lambda^{(k+1)} = 0$$

waar

(5.3.37) $\quad \nabla f(x^{(k+1)}) = \nabla f(x^{(k)}) + G(x^{(k)})(x^{(k+1)}-x^{(k)})$.

In het geval dat $G(x^{(k)})$ niet singulier is volgt daaruit dat

(5.3.38ä) $\quad x^{(k+1)}-x^{(k)} = G^{-1}(x^{(k)})(N^{(k)}\lambda^{(k+1)}-\nabla f(x^{(k)}))$

of met een (ook hierna toe te passen) vereenvoudigde notatie

(5.3.38b) $\quad x^{*} - x = G^{-1}(N\lambda^{*}-\nabla f)$.

Substitutie van dit resultaat in de nevenvoorwaarde (5.3.35) geeft de mogelijkheid de onbekende parameter λ^{*} te bepalen uit

(5.3.39) $\quad N^TG^{-1}N\lambda^{*} - N^TG^{-1}\nabla f = 0$

met als resultaat de *Newton-zoekrichting I*

(5.3.40) $\quad d^{(k)} := -(I-G^{(k)-1}N^{(k)}(N^{(k)T}G^{(k)-1}N^{(k)})N^{(k)T})G^{(k)-1}\nabla f(x^{(k)})$.

Een tweede-orde benadering van de alternatieve formulering (5.3.10) van het basisprobleem met behulp van de matrix $Z^{(k)}$ (5.3.8) resulteert in het onbeperkte kwadratische minimaliseringsprobleem

(5.3.41) $\quad \min\{f(x^{(k)})+\nabla^Tf(x^{(k)})Z^{(k)}w+\frac{1}{2}w^TZ^{(k)T}G(x^{(k)})Z^{(k)}w \mid w \in \mathbb{R}^{n-q}\}$

waarvan de oplossing, indien $Z^{(k)T}G(x^{(k)})Z^{(k)}$ positief definiet is, gevonden wordt voor

(5.3.42) $\quad w^{(k+1)} := -(Z^{(k)T}G(x^{(k)})Z^{(k)})^{-1}Z^{(k)T}\nabla f(x^{(k)})$.

In de originele coördinaten levert dit direct de *Newton-zoekrichting II*

(5.3.43) $\quad d^{(k)} := -Z^{(k)}(Z^{(k)T}G(x^{(k)})Z^{(k)})^{-1}Z^{(k)T}\nabla f(x^{(k)})$.

Juist als in het geval bij de eerste orde zoekrichtingen kan worden aangetoond dat, als $G(x^{(k)})$ niet singulier is, de beide uitdrukkingen (5.3.40) en (5.3.43) dezelfde zoekrichting representeren. Het verschil tussen beide zit voornamelijk in de grotere toepasbaarheid van de laatste formulering die een gevolg is van het feit dat alleen de eigenschappen van de matrix $Z^{(k)T}G^{(k)}Z^{(k)}$ meespelen en niet die van de gehele originele Hessiaan $G(x^{(k)})$. De tweede formulering leent zich ook beter voor modificaties in het geval de matrix $Z^{(k)T}G^{(k)}Z^{(k)}$ niet positief definiet is, als ook voor de ontwikkeling van quasi-Newtonmethoden (vgl. GILL & MURRAY [1974]).

5.3.4. Bepaling van de maximale stapgroottefactor

Voor de bepaling van de maximale stapgrootte wordt uitgegaan van de lokale linearisering van alle beperkingen, zowel actieve als passieve, in het toegelaten punt $x^{(k)}$. De in de voorgaande punten besproken zoekrichtingen werden zo geconstrueerd dat alle gelineariseerde actieve beperkingen actief blijven ongeacht de grootte van de stap. Bepalend voor de maximale stapgrootte zijn daarom uitsluitend de gelineariseerde passieve beperkingen

$$c_j(x^{(k)})+a_j^{(k)T}(x-x^{(k)}) \geq 0, \quad j \in I_p(x^{(k)}). \tag{5.3.44}$$

Als voorwaarde volgt daaruit voor de maximale stap $\bar{\alpha}$ dat

$$\bar{\alpha}a_j^{(k)T}d^{(k)} \geq -c_j(x^{(k)}) \tag{5.3.45}$$

hetgeen onmiddellijk leidt tot de uitdrukking

$$\bar{\alpha} := \min_j\left\{ \frac{-c_j(x^{(k)})}{a_j^{(k)T}d^{(k)}} \;\middle|\; j \in I_p(x^{(k)}) \wedge a_j^{(k)T}d^{(k)} < 0\right\}. \tag{5.3.46}$$

5.3.5. Bepaling van toegelaten punten tijdens het iteratieproces

In het geval dat alle beperkingen lineair zijn en gebruik wordt gemaakt van de in het voorgaande besproken zoekrichtingen en stapgroottebegrenzingen 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

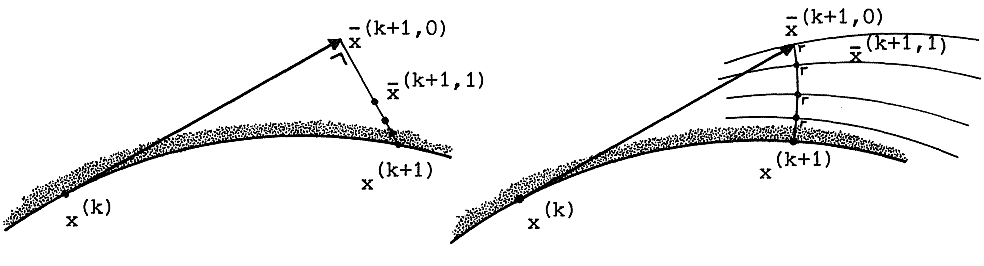
$$\bar{x}^{(k+1,0)} = x^{(k)} + \alpha^{(k)} d^{(k)} \tag{5.3.47}$$

zal in het algemeen dan ook een niet-toegelaten punt zijn en speciale procedures zijn nodig om uitgaande van het niet-toegelaten punt (5.3.47) een toegelaten punt te bepalen. Zoals gezegd in par. 5.2.2 worden deze procedures aangeduid als *restauratieprocedures*. 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 of door alleen in het eerste niet-toegelaten punt in de betreffende iteratieslag alle informatie opnieuw te bepalen en deze in eventuele volgende restauratiestappen te gebruiken. De mogelijkheden genoemd onder a en b worden geïllustreerd in Figuur 5.3.2.



Figuur 5.3.2. Twee mogelijkheden om vanuit een niet toegelaten punt $\bar{x}^{(k+1,0)}$ een toegelaten punt $x^{(k+1)}$ te vinden.

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 geprojekteer-de-gradiëntmethode. Wanneer $d^{(k)}$ bepaald is met de gereduceerde-gradiënt-methode dan kan men de niet-basiscomponenten van $\bar{x}^{(k+1)} = x^{(k)} + \alpha^{(k)} d^{(k)}$ vasthouden en alleen de basiscomponenten aanpassen aan de niet-lineaire beperkingen. Dat wil zeggen dat de restauratie alleen via de basisvariabelen wordt uitgevoerd. Als geldt dat tijdens het hele 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

$$(5.3.48) \qquad \begin{aligned} B^{(k)} s_B^{(k+1,r)} &= - c(\bar{x}^{(k+1,r)}) \\ s_D^{(k+1,r)} &= 0 \end{aligned}$$

waar $s^{(k+1,r)} = \bar{x}^{(k+1,r+1)} - \bar{x}^{(k+1,r)}$ 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. (5.3.2))

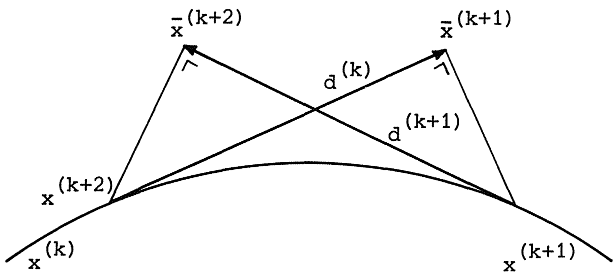
$$(5.3.49) \qquad \bar{N}^{(k)T} s^{(k+1,r)} = \bar{N}^{(k)T} \bar{N}^{(k)} \kappa^{(k+1,r)} = - c(\bar{x}^{(k+1,r)})$$

waaruit volgt

$$(5.3.50) \qquad s^{(k+1,r)} = - \bar{N}^{(k)} (\bar{N}^{(k)T} \bar{N}^{(k)})^{-1} c(\bar{x}^{(k+1,r)}).$$

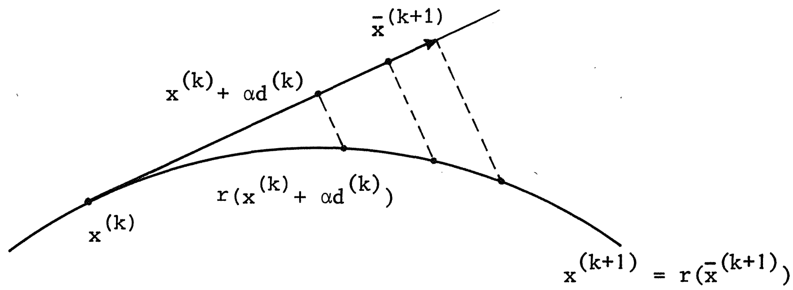
De matrix $\bar{N}^{(k)}$ die vergelijkbaar is met de in par. 5.3.1 besproken matrix $\bar{A}$ bevat in eerste instantie de normalen van de actieve beperkingen in $x^{(k)}$. Als tijdens de hele restauratieprocedure geen andere beperkingen overschreden worden dan die, die in $x^{(k)}$ actief waren, kan 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 par. 5.3.1 beschreven voor het geval van een toegelaten startpunt.

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 5.3.3). Voorkomen kan men dat door in de gebruikte lijnminimaliseringsprocedure een andere strategie



Figuur 5.3.3. Een mogelijk geval van cykelen

toe te passen zoals geïllustreerd in Figuur 5.3.4. Zij $r(x)$ het toegelaten punt, dat gevonden wordt bij restauratie uitgaande van het punt x. Dan moeten we in de gebruikte lijnminimaliseringsprocedure het minimum bepalen van $f(r(x^{(k)}+\alpha d^{(k)}))$ in plaats van van $f(x^{(k)}+\alpha d^{(k)})$.



Figuur 5.3.4. Lijnminimaliseringsmethode om cykelen te voorkomen.

5.4. Practische details: Beschrijving van de belangrijkste procedures

In deze paragraaf zullen de belangrijkste procedures beschreven worden waaruit de Algol-procedure NONLINMIN is opgebouwd. De plaats die deze procedures innemen in het totaal volgt direct uit het flowdiagram van de procedure NONLINMIN dat weergegeven is in Figuur 5.4.1. De procedure RESTORE bepaalt zowel voor de eerste als voor alle andere iteraties een toegelaten punt $x^{(k)}$ als startpunt voor de komende iteratie. Tevens bepaalt de procedure welke beperkingen actief zijn in het betreffende toegelaten punt. De procedure bestaat zelf uit een viertal subprocedures a) VIOLACT, b) NEW NORMALS, c) MNSOLN en d) LINESEARCH, welke aan elkaar gekoppeld zijn volgens het in Figuur 5.4.2 geschetste flowdiagram, waarin QV het in de procedure VIOLACT gevonden aantal overschreden beperkingen voorstelt. De functies van deze subprocedures zijn als volgt:
De procedure VIOLACT bepaalt de status van alle beperkingen (actief, passief dan wel overschreden) in de actuele benadering x van de oplossing. Hierbij wordt onderscheid gemaakt tussen de triviale en de niet-triviale beperkingen. Eerst worden alle triviale beperkingen bekeken en bij overschrijding van een ervan, wordt het punt x aangepast door de betreffende component op de dichtsbijliggende grenswaarde te zetten. Daarmee wordt deze triviale beperking actief. Daarna worden in het, eventueel aangepaste, punt x alle niet-triviale beperkingen geëvalueerd en genoteerd of deze actief, passief dan wel overschreden zijn. Ook worden de aantallen overschreden, actieve triviale en actieve niet-triviale beperkingen bijgehouden.

(o)

(i)

(ii)

(iii)

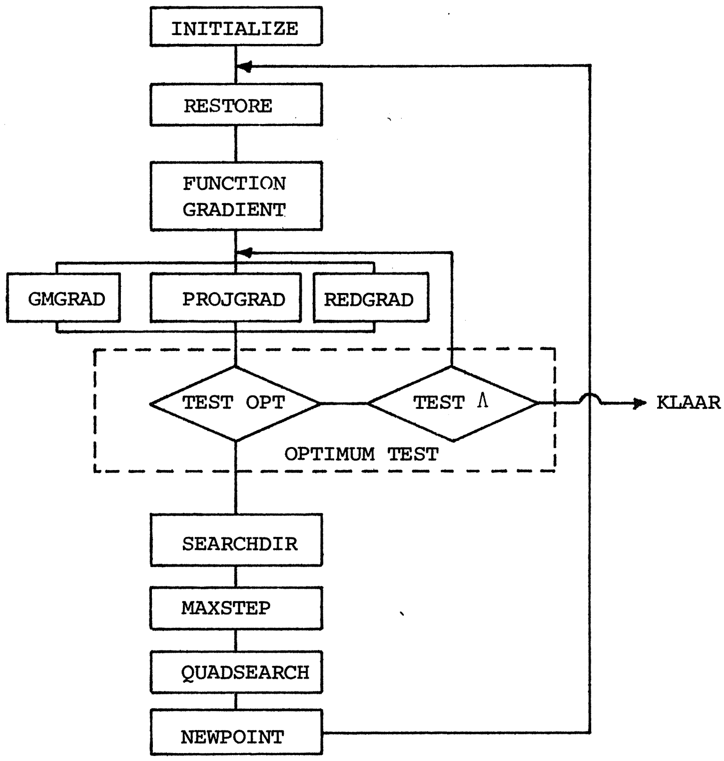
(iv)

(v)

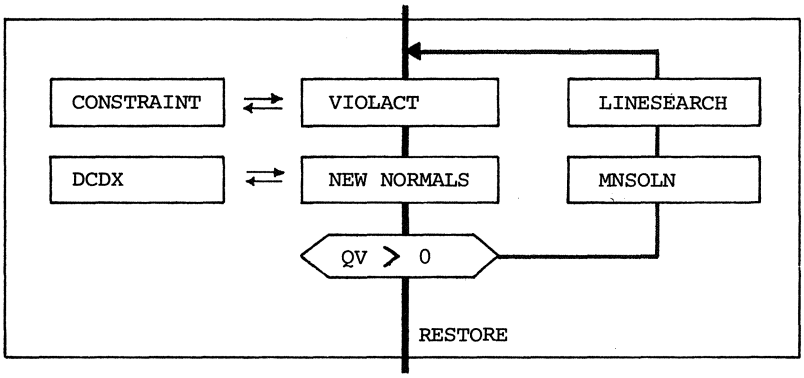
(vi)

(vii)

(viii)



Figuur 5.4.1. Flowdiagram van de procedure NONLINMIN (De nummers tussen ( ) corresponderen met de stappen in de basisalgoritme (pt. 5.2.2)).



Figuur 5.4.2. De procedure RESTORE

De procedure NEW NORMALS vult de in par. 5.3.1 en par. 5.3.5 besproken matrices $\bar{A}$ of $\bar{N}^{(k)}$. Afhankelijk van de gevolgde strategie (vgl. par. 5.3.5) worden bij de methoden, die gebruik maken van een geprojecteerde gradiënt, de normalen van de niet-lineaire beperkingen gehandhaafd, dan wel opnieuw geëvalueerd voor iedere nieuwe benadering $x^{(k,r)}$ van het toegelaten punt $x^{(k)}$. Bij de methoden, die gebruik maken van de gereduceerde gradiënt wordt in de procedure NEW NORMALS of (1) de basismatrix gehandhaafd, of (2) er vindt een basiswisseling plaats, of (3) van het onderscheid tussen basis- en niet-basisvariabelen wordt geheel afgezien. Deze drie mogelijkheden doen zich respectievelijk voor (1) wanneer de strategie het handhaven van de oude normalen voorschrijft en geen, eerder passieve triviale of passieve niet-triviale beperking actief wordt; (2) wanneer de strategie het handhaven van de oude normalen voorschrijft en er een triviale beperking actief wordt en (3) in alle andere gevallen. In het laatste geval gaat de restauratieprocedure verder als bij de methoden die gebruik maken van de geprojecteerde gradiënt.

De procedure MNSOLN bepaalt, uitgaande van een niet-toegelaten punt en met behulp van de matrix van normalen van actieve en overschreden beperkingen, de stap naar een toegelaten punt, dan wel de stap naar een punt met minder overschreden beperkingen. Afhankelijk van de omstandigheid of de som van het aantal actieve en het aantal overschreden beperkingen groter of gelijk, dan wel kleiner is dan het aantal variabelen (vgl. par. 5.3.1) bepaalt de procedure de kleinste-kwadraten dan wel de minimumnorm oplossing van de vergelijking (5.3.2)

$$\bar{A}^T \Delta x = - \bar{c}$$

of, equivalent, van de vergelijking (5.3.49)

$$\bar{N}^{(k)T} \underline{s}^{(k+1,r)} = - c(\bar{x}^{(k+1,r)}).$$

De gevonden stap wordt als zoekrichting gebruikt in de procedure LINESEARCH. De procedure RESTORE wordt verlaten indien een toegelaten punt is gevonden, dan wel indien meer dan een vooraf aangegeven aantal restauratiestappen zijn gedaan zonder resultaat. In dit laatste geval wordt het iteratieproces gestopt.

In de procedure NONLINMIN zijn drie procedures, te weten: PROJGRAD, GMGRAD en REDGRAD, opgenomen, die evenzoveel transformaties van de gradiënt van de objectfunctie genereren. Deze getransformeerde gradiënten worden zowel gebruikt voor het genereren van een zoekrichting in de procedure SEARCHDIR als, samen met de in dezelfde procedures bepaalde Lagrangemultiplicatoren, in de test of een lokaal optimaal punt is bereikt. Dit gebeurt in de procedure OPTIMUM TEST. De drie procedures voor het genereren van een transformatie van de gradiënt van de objectfunctie fungeren in meer detail als volgt: De procedure GMGRAD berekent de geprojecteerde gradiënt volgens de methode van Gill en Murray. Gebruik wordt gemaakt van een matrix Z, waarvan de kolommen een orthonormale basis vormen van het orthogonale complement van de deelruimte opgespannen door de normalen van actieve beperkingen. Zij A de $(n \times m_A)$-matrix met in de kolommen de normalen van de actieve beperkingen, dan geeft QR-decompositie, als k de rang is van de matrix A, het resultaat:

$$\text{(5.4.1)} \qquad A = Q \left[\begin{array}{c|c} R & D \\ \hline 0 & 0 \end{array}\right] P^T$$

waar R een $k \times k$ bovendriehoeksmatrix is, Q een $n \times n$ matrix waarvoor $Q^TQ = I$, P een $m_A \times m_A$ permutatiematrix en D een $k \times (m_A-k)$ matrix. Splitsen we Q in een $n \times k$ matrix $Q^{(1)}$ en een $n \times (n-k)$ matrix $Q^{(2)}$, dan volgt

$$\text{(5.4.2)} \qquad \begin{aligned} Q^{(1)^T} A &= [R|D]P^T \\ Q^{(2)^T} A &= 0. \end{aligned}$$

Hieruit en uit

$$(5.4.3) \qquad \begin{bmatrix} Q^{(1)^T} \\ \hdashline Q^{(2)^T} \end{bmatrix} (Q^{(1)} \vdots Q^{(2)}) = \begin{bmatrix} I_k & 0 \\ 0 & I_{n-k} \end{bmatrix}$$

blijkt dat $Q^{(2)}$ een goede keuze is voor de in par. 5.3.2 gepostuleerde matrix Z. De Gill-Murray-gradiënt (5.3.13) wordt met de matrix $Q^{(2)}$ gemakkelijk gevonden als

$$(5.4.4) \qquad g_{GM}^{(k)} = Q^{(2)} Q^{(2)^T} \nabla f(x^{(k)}).$$

De vector $\lambda^{(k)}$ van Lagrangemultiplicatoren wordt gevonden als de kleinste-kwadraten-oplossing van

$$(5.4.5) \qquad A\lambda^{(k)} - \nabla f(x^{(k)}) = 0.$$

Met behulp van de resultaten van de QR-decompositie van A wordt hiervoor gevonden

$$(5.4.6) \qquad \lambda^{(k)} = P \begin{bmatrix} R^{-1} Q^{(1)^T} \\ \hdashline 0 \end{bmatrix} \nabla f(x^{(k)})$$

als we de Lagrangemultiplicatoren voor de afhankelijke actieve beperkingen a priori gelijk aan nul kiezen.

De procedure PROJGRAD bepaalt de in par. 5.3.2 besproken geprojecteerde gradiënt eveneens met behulp van de QR-decompositie (5.4.1) van de matrix A. Na splitsing van Q op dezelfde manier als in GMGRAD blijkt dat de projectiematrix (5.3.14)

$$\bar{P}^{(k)} = (I - N^{(k)} (N^{(k)^T} N^{(k)})^{-1} N^{(k)^T})$$

gegeven wordt door de uitdrukking

$$(5.4.7) \qquad \bar{P}^{(k)} = I - Q^{(1)} Q^{(1)^T}.$$

Daarmee is de geprojecteerde gradiënt te schrijven als

(5.4.8) $$g_{PROJ}^{(k)} = \nabla f(x^{(k)}) - Q^{(1)}Q^{(1)^T}\nabla f(x^{(k)}).$$

De vector van Lagrangemultiplicatoren wordt weer gevonden uit de hierboven bij de procedure GMGRAD besproken relatie (5.4.6)

$$\lambda^{(k)} = P \begin{bmatrix} R^{-1}\ Q^{(1)^T} \\ \hline 0 \end{bmatrix} \nabla f(x^{(k)}) \ .$$

De procedure REDGRAD bepaalt een gereduceerde gradiënt op de manier, zoals besproken in par. 5.3.2. Begonnen wordt met het maken van een keuze voor een basis. Het criterium daarbij is, dat geen van de basisvariabelen op zijn onder- of bovengrens ligt. Vervolgens wordt nagegaan of de matrix B niet-singulier is. Is dit wel het geval, dan wordt een basiswisseling uitgevoerd. Zoniet, dan wordt een vector λ van Lagrangemultiplicatoren berekend (vgl. (5.3.30))

(5.4.9) $$\lambda = B^{-T}\nabla_{x_B} f$$

en daarmee de gereduceerde gradiënt

(5.4.10) $$\nabla_{x_D}\tilde{f} = \nabla_{x_D} f - D^T\lambda.$$

Daarna wordt afgeweken van de in par. 5.3.2 beschreven formulering en wordt gelet op de mogelijkheid om de componenten van $-\nabla_{x_D}\tilde{f}$ voor de niet-basisvariabelen te gebruiken als componenten van de zoekrichting. Voor de niet-basisvariabelen waarvoor deze componenten van teken zo zijn, dat actieve triviale beperkingen overschreden dreigen te worden, wordt de overeenkomstige component van de aangepaste gereduceerde gradiënt $[\nabla_{x_D}\tilde{f}]$ op nul gesteld. De andere componenten worden ongewijzigd gehandhaafd. Tenslotte worden de basiscomponenten van de aangepaste gereduceerde gradiënt bepaald uit (5.3.33)

$$g_{R,B} = -\ B^{-1}D[\nabla_{x_D}\tilde{f}]$$

waarmee als "gereduceerde gradiënt" resulteert

$$(5.4.11) \qquad g_R = \begin{bmatrix} g_{R,B} \\ ---- \\ g_{R,D} \end{bmatrix} = \begin{bmatrix} -B^{-1}D[\nabla_{x_D}\tilde{f}] \\ ---------- \\ [\nabla x_D \tilde{f}] \end{bmatrix} .$$

In de procedure OPTIMUM TEST wordt gecontroleerd of het laatst bepaalde toegelaten punt het optimale punt is. Hierbij wordt gebruik gemaakt van drie verschillende criteria

a) (1): $\|g^{(k)}_{GM/PROJ/R}\| < \varepsilon_1 \|\nabla f(x^{(0)})\| + \varepsilon_2$ en (2): $\lambda \geq 0$

b) $\|x^{(k)} - x^{(k-1)}\| \leq \varepsilon_3$

c) $|f(x^{(k)}) - f(x^{(k-1)})| \leq \varepsilon_4 |f(x^{(k-1)})| + \varepsilon_5$.

Het eerst wordt nagegaan of aan criterium a) (1) voldaan is. Is dat het geval, dan wordt gecontroleerd of alle Lagrangemultiplicatoren behorend bij de actieve ongelijkheidsbeperkingen groter dan of gelijk aan nul zijn. Wanneer dit niet het geval is, wordt de beperking die overeenkomt met de meest negatieve Lagrangemultiplicator passief gemaakt en wordt zonder de andere convergentiecriteria te beschouwen teruggegaan naar een der procedures GMRAD, PROJGRAD of REDGRAD waar een nieuwe getransformeerde gradiënt wordt bepaald. In het andere geval worden beide andere convergentiecriteria gecontroleerd. Meegegeven moet worden of men het proces convergent noemt zodra behalve aan criterium a) ook aan nul, één of twee van de onder b) en c) genoemde criteria voldaan is.

In de procedure SEARCHDIR wordt de zoekrichting $d^{(k)}$ bepaald. Op dit moment wordt in deze procedure nog niets anders gedaan, dan de zoekrichting gelijkstellen aan de negatieve getransformeerde gradiënt

$$(5.4.12) \qquad d^{(k)} = - g^{(k)}_{GM/PROJ/R}.$$

De procedure MAXSTEP bepaalt de maximale toegelaten stap vanuit het punt $x^{(k)}$ in de zoekrichting $d^{(k)}$, binnen het door de passieve gelineariseerde beperkingen bepaald gebied. Dit wordt gedaan volgens de methode beschreven in par. 5.3.4, d.w.z. door evaluatie van de uitdrukking (5.3.46)

$$\alpha_{max} = \min_j \left\{ - \frac{c_j(x^{(k)})}{a_j^{(k)T} d^{(k)}} \;\middle|\; j \in I_p(x^{(k)}),\ a_j^{(k)T} d^{(k)} < 0 \right\} .$$

In het geval geen enkele beperking passief is, wordt α_{max} gelijkgesteld aan een meegegeven constante.

De procedure QUADSEARCH tenslotte bepaalt een benadering voor het minimum van de objectfunctie langs de zoekrichting $d^{(k)}$. Daartoe worden, binnen het door α_{max} beperkte gebied, drie punten op deze lijn gezocht, zodanig dat het mogelijk is door de drie bijbehorende functiewaarden een dalparabool te construeren. Het minimum van deze dalparabool wordt gebruikt als benadering voor het gezochte minimum van de objectfunctie. Een meer gedetailleerde beschrijving van deze procedure wordt gegeven door EILERS [1975].

5.5. Slotopmerking

In het voorgaande is een beknopte beschrijving gegeven van de opbouw van een algemene Algolprocedure voor het minimaliseren van niet-lineaire functies onder niet-lineaire nevenvoorwaarden. De ontwikkeling daarvan is nog in volle gang. Numerieke ervaring werd slechts in zeer beperkte mate opgedaan. De gekozen modulaire opbouw van de procedure is tot nu toe geschikt gebleken voor de beoogde ontwikkeling en diverse deelprocedures werden inmiddels reeds verder verfijnd en uitgebouwd. De belangrijkste ervaring tot op dit moment is de enorm grote hoeveelheid tijd en energie die nodig is voor het ontwikkelen en operationeel maken van dit soort universele programma's. Een woord van dank aan R. Kool van de Onderafdeling der Wiskunde, die samen met C.J.B. Dirkx en de schrijver hier reeds vele uren aan besteedde, lijkt in dit verband zeker op zijn plaats.

LITERATUUR

ABADIE, J. & J. CARPENTIER, [1969]: *Generalization of the Wolfe reduced gradient to the case of nonlinear constraints*, in: Fletcher, J. (ed.): *Optimization*, Academic Press, New York.

DIRKX, C.J.B. [1975]: *NONLINMIN, een Algol-procedure voor het minimaliseren van niet-lineaire functies onder niet-lineaire nevenvoorwaarden*, Afstudeerverslag, Onderafdeling der Wiskunde, Technische Hogeschool Eindhoven.

EILERS, G.A.M. [1975]: *Het minimaliseren van sommen van kwadraten van niet-lineaire functies*, Memorandum COSOR 75.08, Onderafdeling der Wiskunde, Technische Hogeschool Eindhoven.

GILL, P.E. & W. MURRAY, [1974]: *Numerical methods for constrained optimization*, Academic Press, London.

DE JONG, J.L. [1976]: *Numerieke algoritmen voor niet-lineaire optimaliseringsproblemen*, Syllabus by het college Capita Optimaliseringsmethoden, voorjaar 1976, Onderafdeling der Wiskunde, Technische Hogeschool Eindhoven.

LOOTSMA, F.A. [1972]: *The Algol-procedure MINIFUN for solving non-linear optimization problems*, Report 4761, Philips Research Laboratories, Eindhoven.

LUENBERGER, D.G. [1973]: *Introduction to linear and nonlinear programming*, Addison Wesley Publ. Cy., Reading, Mass.

ROSEN, J.B. [1960]: *The gradient projection method for nonlinear programming, Part I:Linear constraints*, SIAM J. Appl. Math. 8, pp. 181-217.

WOLFE, PH. [1963]: *Methods of nonlinear programming*, in: GRAVES, R.L. & PH. WOLFE, (eds.) *Recent advances in mathematical programming*, McGraw-Hill, New York.