STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM
AFDELING ZUIVERE WISKUNDE

ZW 1968-006

Voordracht in de serie
"Elementaire onderwerpen vanuit hoger standpunt belicht"

door

Prof. dr. G. Zoutendijk (R.U. Leiden)

De Wiskunde van het Bergbeklimmen

1. Inleiding

Wij zullen ons in dit rapport bezighouden met het bepalen van een maximum van een functie $f(x)$, waarbij x een vector is met n componenten. We zullen aannemen:

C1: f differentiëerbaar met continue partiële afgeleiden

C2: er is een μ zodanig dat $\{x | f(x) \geq \mu\}$ begrensd en niet leeg. Het infimum van alle μ, waarvoor dit geldt, geven we met μ_0 aan.

Nodig (doch niet voldoende) voor een maximum is:

$$\frac{\partial f}{\partial x_j} = 0, \; j = 1, \ldots, n, \quad \text{i.e.} \quad \nabla f(x) = 0 \qquad (1)$$

Het numeriek oplossen van een stelsel meestal niet-lineaire vergelijkingen is echter geen eenvoudige zaak. Het ligt daarom voor de hand rechtstreekse numerieke methoden te ontwikkelen voor het maximaliseringsprobleem. De te behandelen methoden verlopen alle volgens het volgende schema:

1. Gegeven een startpunt x^0.
2. Stel x^h al berekend voor $h = 0, 1, \dots, k$, dan
   a. bepaal een richting $s^k \neq 0_k$ met $\nabla f(x^k)^T s^k \geqq 0$.
   b. bepaal de staplengte λ_k in de richting s^k zodanig dat $\nabla f(x^k + \lambda_k s^k)^T s^k = 0$.
   c. $x^{k+1} = x^k + \lambda_k s^k$.

We zullen spreken van een methode van toelaatbare richtingen. Uit C2 volgt dat als $f(x^0) > \mu_0$ de rij x^k minstens één verdichtingspunt heeft. Een methode zal bruikbaar zijn wanneer voor elk verdichtingspunt $\bar{x}$ van de rij x^k geldt $\nabla f(\bar{x}) = 0$. Er is dan convergentie naar één of meer (locale) maxima (of in uitzonderingsgevallen naar een zadelpunt).
De staplengte bepaling is een ééndimensionaal maximumprobleem (in λ) dat door herhaalde Hermite (kubieke) interpolatie kan worden opgelost (maximum-bepaling van een derdegraadskromme).
Een voorbeeld is de van Cauchy afkomstige optimale gradientmethode, ook wel methode van de grootste stijging genoemd, waarbij steeds $s^k = \nabla f(x^k)$ gekozen wordt.

2. Convergentie van de methoden

Stelling

Als geldt:

a. C1 en C2, $f(x^0) > \mu_0$.
b. $\nabla f(x^k)^T s^k = \theta_k |\nabla f(x^k)| . |s^k|$, $\theta_k \geqq 0$.
c. $\sum_l \theta_{k_l} \to \infty$ voor elke convergentie deelrij x^{k_l} van de x^k.

Dan zal $\nabla f(\bar{x}) = 0$ zijn voor elk eendichtingspunt $\bar{x}$.

Bewijs: Stel $\exists \bar{x}$. $|\nabla f(\bar{x})| \geqq \varepsilon > 0$. Dan is er een $\delta(\bar{x}, \varepsilon)$ zodanig dat als $N(\bar{x}, \delta) = \{x \mid |x-\bar{x}| \leq \delta\}$, $x^k \in N(\bar{x}, \delta)$ en $x \in N(\bar{x}, \delta)$, willekeurig, er geldt $\nabla f(x)^T s^k \geqq \frac{\theta_k \varepsilon}{2} |s^k|. > 0$.
Laat $x^{k_l} \to \bar{x}$ $(l = 1, 2, \dots)$, dan is er een l' zodanig dat voor $l \geqq l'$ geldt $x^{k_l} \in N(\bar{x}, \frac{\delta}{2})$. Dus $x^{k_l+1} \notin N(\bar{x}, \delta)$ daar anders $\nabla f(x^{k_l+1})^T s^k > 0$ zou gelden volgens het voorgaande, in strijd met de regel 2b, § 1.
Laat de verbindingslijn van x^{k_l} naar x^{k_l+1} de grens van $N(\bar{x}, \delta)$ snijden in y^l.

$$f(x^{k_1+1}) - f(x^{k_1}) \geq f(y^1) - f(x^{k_1}) = \nabla f(z^1)^T (y^1 - x^{k_1}) \geq \frac{\delta}{2} .$$

$$. \nabla f(z^1)^T \frac{s^{k_1}}{|s^{k_1}|} \geq \frac{\theta_{k_1} \delta . \varepsilon}{4} .$$

Dus volgens c. $\sum\{f(x^{k_1+1}) - f(x^{k_1})\} \to \infty$. Dit is echter in tegenspraak met C2.

Opmerking: Als we in plaats van c eisen:

c' $\qquad \sum \theta_k \to \infty$,

dan kunnen we alleen maar bewijzen dat voor minstens één verdichtingspunt $\bar{x}$ geldt $\nabla f(\bar{x}) = 0$. We zullen dit zwakke convergentie noemen. Het is onbekend of zwakke convergentie normale convergentie impliceert. Daar bij de optimale gradientmethode altijd geldt $\theta_k = 1$ is deze convergent.

Er kunnen echter meer verdichtingspunten voorkomen, zoals blijkt uit het volgende voorbeeld: (in cylinder coördinaten)

$$z = -(r - 1)^2 + \tfrac{1}{2}(r - 1)^2 \cos(\frac{1}{r - 1} - \phi),$$

waarin $r = 1^o$ een cirkel van maxima is waarvan elk punt vanuit elk willekeurig startpunt bij de gradientmethode zal worden benaderd. De punten x^k zullen op een spiraal liggen, die zich om de cirkel heenwindt.

De convergentie van de optimale gradientmethode is meestal erg slecht.

3. <u>Geconjugeerde richtingen</u>

Stel $f(x) = p^T x - \frac{1}{2} x^T C x$, kwadratisch, C positief definiet

$$\nabla f(x) = p - Cx, \quad \lambda_k = \frac{\nabla f(x^k)^T s^k}{(s^k)^T C s^k} .$$

We noemen twee richtingen s^1 en s^2 geconjugeerd als $(s^1)^T C s^2 = 0$.

Stel nu dat we bij onze methode van toelaatbare richtingen naast 2a ook zouden eisen 2a': $(s^k)^T Cs^h = 0$, $h = 0, 1, \ldots, k-1$, dan geldt

<u>Stelling</u>: De zo verkregen methode geeft ons na $m \leq n$ stappen een x^m met $\nabla f(x^m) = 0$.

Bewijs:

a. Voor elke k zijn de vectoren $s^0, s^1, \ldots, s^k$ lineair onafhankelijk.

Stel namelijk $\sum_{i=0}^{k} \alpha_i s^i = 0$ dus $\sum_{i=0}^{k} \alpha_i C s^i = 0$.

Laat $0 \leq h \leq k$. $\sum_{i=0}^{k} \alpha_i (s^h)^T Cs^i = \alpha_h (s^h)^T Cs^h = 0$ dus $\alpha_h = 0$ voor alle h.

b. Uit a volgt dat na een eindig aantal (m) stappen $s^m = 0$ moet gelden

$\nabla f(x^m) = \nabla f(x^k) - \sum_{h=k}^{m-1} \lambda_h Cs^h$ geldt voor elke $k < m$

$\nabla f(x^m)^T s^k = \nabla f(x^k)^T s^k - \lambda_k (s^k)^T Cs^k = 0$ voor elke $k < m$.

c. Stel $\nabla f(x^m) \neq 0$, dan is er een s met $\nabla f(x^m)^T s > 0$.

Laat $t = s - \sum_{h=0}^{m-1} \mu_h s^h$ met $\mu_h = \dfrac{(s^h)^T Cs}{(s^h)^T Cs^h}$

$k \leq m - 1$: $(s^k)^T Ct = (s^k)^T Cs - \mu_k (s^k)^T Cs^k = 0$

$\nabla f(x^m)^T t = \nabla f(x^m)^T s > 0$. contradictie.

Bij de methoden van geconjugeerde richtingen bestaat nog steeds vrijheid in het kiezen van de s^k. Er moet alleen aan de lineaire betrekkingen 2a' voldaan worden. Een bijzonder geval is de geconjugeerde gradient methode waarbij

$$s^k = s^{k-1} + \frac{g^k}{(g^k)^T g^k} \qquad (g^k = \nabla f(x^k)), \quad s^{-1} = 0 \tag{2}$$

$$s^k = \sum_{h=0}^{k} \frac{g^h}{(g^h)^T g^h}. \tag{3}$$

Direct duidelijk: $(g^k)^T s^k = 1$, $\lambda_k = \dfrac{1}{(s^k)^T Cs^k}$.

<u>Stelling:</u> De geconjugeerde gradient methode is een methode van geconjugeerde richtingen.

Er geldt: a. $(s^h)^T Cs^l = 0, \quad h \neq l$

b. $(g^h)^T g^l = 0, \quad h \neq l$

c. $(g^h)^T s^l = 1, \quad h \leq l$

d. s^k is voor elke k de oplossing van het probleem

$$\text{Min } \{\tfrac{1}{2} s^T s \,|\, (g^h)^T s = 1,\ h = 0, 1, \ldots, k\}$$

e. $|s^k|^2 = |s^{k-1}|^2 + |\dfrac{1}{g^k}|^2$.

<u>Bewijs</u>:

a en b: We zullen volledige inductie toepassen. Er geldt $(g^1)^T g^0 = 0$ $((g^1)^T s^0 = 0$ en $s^0 = \dfrac{g^0}{|g^0|^2}$), alsmede $(s^0)^T Cs^1 =$

$$= \frac{(g^1-g^0)^T}{-\lambda_0} \left(s^0 + \frac{g^1}{|g^1|^2}\right) = \frac{1}{\lambda_0} - \frac{1}{\lambda_0} = 0.$$

Stel a en b gelden voor h, $l \leq k-1$, $h \neq l$

$h \leq k-2$: $(g^h)^T g^k = (g^h)^T(g^{k-1} - \lambda_{k-1}\, Cs^{k-1}) = -\lambda_{k-1}\, (g^h)^T\, Cs^{k-1} = 0$ daar

$$\frac{g^h}{|g^h|^2} = s^h - s^{k-1},\ (s^h)^T\, Cs^{k-1} = 0 \text{ en } (s^{h-1})^T\, Cs^{k-1} = 0$$

$$(g^{k-1})^T g^k = (g^{k-1})^T g^{k-1} - \lambda_{k-1}\, (g^{k-1})^T\, Cs^{k-1} = (g^{k-1})^T g^{k-1} -$$

$$- \frac{1}{(s^{k-1})^T Cs^{k-1}}\, (g^{k-1})^T g^{k-1}\, \{s^{k-1} - s^{k-2}\}^T\, Cs^{k-1} = 0$$

$$(s^h)^T\, Cs^k = \frac{\{g^{h+1} - g^h\}^T s^k}{-\lambda_h} = -\frac{1}{\lambda_h}\, \{g^{h+1} - g^h\}^T \sum_{l=0}^{k} \frac{g^l}{|g^l|^2} =$$

$$= -\frac{1}{\lambda_h}\, (1-1) = 0.$$

c: $$(g^h)^T s^l = (g^h)^T \sum_{i=0}^{l} \frac{g^i}{|g^i|^2} = 1 \quad (h \leq l)$$

d: Volgens de methode van de Lagrange multiplicatoren moet gelden voor de oplossing

$$\tilde{s}^k : \tilde{s}^k = \sum_{h=0}^{k} \alpha_h^k\, g^h.$$

Voor $k = 0$ volgt uit $(g^0)^T\ \tilde{s}^0 = 1$ dat $\alpha_0^0 = \frac{1}{|g^0|^2}$ dus dat $\tilde{s}^0 = s^0$.

Stel dat voor $h \leq k-1$ geldt $\tilde{s}^h = s^h$, d.w.z. dat tot en met stap k-1 F.R. en de door d. bepaalde methode equivalent zijn. Volgens het reeds bewezene volgt daar dus uit dat $(g^l)^T\ g^h = 0$ voor $l \leq k$, $h \leq k-1$, $l \neq h$.

Dus $1 = (g^l)^T\ \tilde{s}^k = \sum_{h=0}^{k} \alpha_h^k\, (g^l)^T\ g^h = \alpha_l^k |g^l|^2$. Hieruitvolgt $\alpha_l^k = \frac{1}{|g^l|^2}$ of $\tilde{s}^k = s^k$, q.e.d.

e. Volgt direct uit de definitie van s^k, alsmede uit $(g^k)^T\ s^{k-1} = 0$.

Uit d volgt dat s^k tevens de oplossing is van het probleem

$$\text{Min } \{\tfrac{1}{2}(g^k-s)^T\ (g^k-s) \mid (g^k)^T s = 1,\ (s^k)^T\ Cs = 0,\ h = 0, 1, \ldots, k-1\},$$

d.w.z. dat $\frac{s}{|s|^2}$ de projectie is van g^k op de lineaire deelruimte, bepaald door de betrekkingen $(s^h)^T\ Cs = 0$, $h \leq k-1$.

4. Methoden voor onbeperkte maximalisering

De in deze paragraaf te behandelen methoden berusten alle op de veronderstelling dat een generalisatie van een methode van geconjugeerde richtingen tot meer algemene functies zal leiden tot snellere convergentie. Deze veronderstelling is in vele gevallen in de praktijk juist gebleken.

4.1. Methode van Fletcher en Reeves

s^k wordt bepaald volgens (2). Van de 5 eigenschappen van de geconjugeerde gradient methode blijft alleen e gelden (alsmede $(g^k)^T s^k = 1$).

Stelling. De methode van Fletcher en Reeves is zwak convergent.

Bewijs: Voor de in de stelling van §2 gebruikte θ_k geldt $\theta_k = \dfrac{1}{|g^k|\ |s^k|} =$

$= \dfrac{1}{|g^k|\sqrt{\sum\limits_{h=0}^{k} \dfrac{1}{|g^h|^2}}}$. Als de methode niet zwak convergent is, dan zal er

een $\varepsilon > 0$ zijn, alsmede een k' zodanig dat voor alle $k \geqq k'$ geldt $|g^k| \geqq \varepsilon$.

In dat geval geldt voor $k \geqq k'$:

$$\theta_k \geqq \frac{1}{M\sqrt{\alpha+(k-k')\frac{1}{\varepsilon^2}}} = \frac{1}{\sqrt{\beta\ k+j}}\ , \text{ zodat } \sum \theta_k \to \infty.$$

Uit de opmerking na de convergentiestelling van §2 volgt dan echter dat de methode zwak convergent moet zijn. Dit laatste moet dus het geval zijn.

In de praktijk is de convergentieversnelling nogal teleurstellend, zelfs bij "nette" functies. Dit is vermoedelijk het gevolg van het star vasthouden aan de formule (2) inplaats van aan het aan de formule ten grondslag liggende principe.

4.2. "Variabele metriek" methode van Davidson

Dit is een voorbeeld van een matrix methode, zo geheten omdat per stap een matrix moet worden herberekend.

$$s^k = H_k g^k,\quad H_0 = I \tag{4}$$

$$H_{k+1} = H_k + \lambda_k \frac{s^k (s^k)^T}{(g^k)^T s^k} - \frac{H_k\ \Delta g^k . (\Delta g^k)^T H_k}{(\Delta g^k)^T H_k\ \Delta g^k} \tag{5}$$

$(\Delta g^k = g^{k+1} - g^k)$.

Er geldt:

a. H_k positief definiet (dus $(g^k)^T s^k > 0$ voor alle k).

Immers H_0 is positief definiet. Stel dit geldt reeds voor H_i, $i \leqq k$.

$$x^T H_{k+1} x = x^T H_k x - \frac{x^T H_k \Delta g^k (\Delta g^k)^T H_k x}{(\Delta g^k)^T H_k \Delta g^k} + \lambda_k \frac{x^T s^k (s^k)^T x}{(g^k)^T s^k} \geqq 0$$

daar volgens de ongelijkheid van Schwartz het verschil van de eerste twee termen van het rechterlid niet negatief is, hetgeen ook voor de derde term geldt. Volgens de ongelijkheid van Schwartz zal het gelijkteken alleen gelden wanneer $x = \mu \Delta g^k$. In dat geval is $(x^T s^k)^2 > 0$ zodat altijd geldt $x^T H_{k+1} x > 0$ als $x \neq 0$, q.e.d.

b. Beschouw het kwadratische geval: $f(x) = p^T x - \frac{1}{2} x^T Cx$, C positief definiet.

Nu geldt dat voor elke $k \leqq n-1$, $s^0, s^1, \ldots, s^k$ lineair onafhankelijke eigenvectoren zijn van de matrix $H_{k+1}C$ bij de eigenwaarde 1. De s^i zijn bovendien geconjugeerd.

b1. We bewijzen eerst dat $H_{k+1} C s^k = s^k$. Gebruik hiertoe (5):

$$H_k Cs^k = -\frac{1}{\lambda_k} H_k \Delta g^k$$

$$\lambda_k \frac{s^k (s^k)^T Cs^k}{(g^k)^T s^k} = \lambda_k \frac{s^k}{\lambda_k} = s^k$$

$$-\frac{H_k \Delta g^k (\Delta g^k)^T H_k Cs^k}{(\Delta g^k)^T H_k \Delta g^k} = + \frac{H_k \Delta g^k (\Delta g^k)^T H_k \Delta g^k}{\lambda_k (\Delta g^k)^T H_k \Delta g^k} = \frac{1}{\lambda_k} H_k \Delta g^k$$

Optelling geeft het gewenste resultaat.

b2. Vervolgens bewijzen we voor alle l $(s^i)^T Cs^j = 0$, $0 \leqq i < j \leqq l$

$H_k Cs^i = s^i$, $0 \leqq i \leqq l-1$.

De betrekkingen gelden voor l=1; ($H_1 Cs^0 = s^0$ volgens b1, $(s^1)^T Cs^0 = (g^1)^T H_1 Cs^0 = (g^1)^T s^0 = 0$).

Stel dat de betrekkingen gelden voor $l \leqq k$

$$i \leqq k-1: H_{k+1} Cs^i = H_k Cs^i + \lambda_k \frac{s^k (s^k)^T Cs^i}{(g^k)^T s^k} - \frac{H_k \Delta g^k (\Delta g^k)^T H_k Cs^i}{(\Delta g^k)^T H_k \Delta g^k} =$$

$$= s^i + 0 - \frac{H_k \Delta g^k (\Delta g^k)^T s^i}{(\Delta g^k)^T H_k \Delta g^k} = s^i$$

omdat $(\Delta g^k)^T s^i = -\lambda_k (s^k)^T C s^i = 0$. Het geval i = k is reeds bewezen.

$$(s^i)^T C s^{k+1} = (s^i)^T C H_{k+1} g^{k+1} = (s^i)^T g^{k+1} = (s^i)^T \{g^{i+1} -$$

$$- \sum_{j=i+1}^{k} \lambda_j C s^j\} = 0 \quad \text{q.e.d.}$$

b3. Na n stappen moet gelden:

$$H_n C = I \text{ dus } H_n = C^{-1}$$

$$g^n = 0 \quad \text{want } (g^n)^T s^j = 0 \text{ voor } j = 0, 1, \ldots, n-1,$$

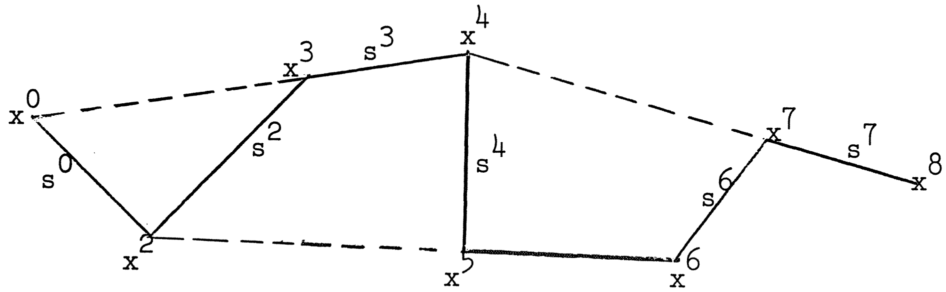
s^j lineair onafhankelijk dus $g^n = 0$.

Bij een kwadratische functie wordt het maximum dus na n stappen bereikt en is er sprake van een methode van geconjugeerde richtingen.

c. Bij een willekeurige functie f(x) kunnen dezelfde formules (4) en (5) worden toegepast. Een convergentiebewijs ontbreekt. In de praktijk blijkt de methode echter goed te voldoen, beter dan de methode van Fletcher en Reeves.

4.3. Partan, ontwikkeld door Shah, Buehler en Kempthorne

De volgende tekening geeft een geometrisch beeld van deze methode:



Voor s^{2k} moet gelden $(g^{2i})^T s^{2k} = 0$ ($k > 0$, $i = 0, 1, \ldots, k-1$).
De s^{2k+1} wordt gekozen in de richting $x^{2k+1} - x^{2k-2}$ ($k = 1, 2, \ldots$).
Het kan bewezen worden (zie het artikel in JSIAM 12 (1964) 74-92) dat de s^{2k} onderling geconjugeerd zijn bij f(x) kwadratisch, zodat we wederom te maken hebben met een methode van geconjugeerde richtingen. Uitbreiding tot niet-kwadratische functies ligt voor de hand.
Een eenvoudige keuze voor s^{2k} is: $s^{2k} = g^{2k}$ (voldoet aan de eisen).

4.4. Methoden van Zoutendijk

Kies:

$$s^0 \text{ zodanig dat } (g^0)^T s^0 = 1$$
$$s^k \text{ zodanig dat } (g^h)^T s^k = 1,\ h \leq k.$$

Het is direct duidelijk dat bij een kwadratische functie weer een methode van geconjugeerde richtingen verkregen is, zodat x^n het maximum zal zijn.

Bij een niet-kwadratische functie zijn twee varianten mogelijk:

a. Als $k \geq n$ laat h lopen van $k-n+1$ tot k (steeds de "oudste" weglaten)

b. Begin op nieuw na n stappen met $x^0 = x^n$.

In geval b is de algemene convergentiestelling toepasbaar.

Bij de keuze van de s^k hebben we nog vrijheid, zodat additionele eisen mogelijk zijn, bijvoorbeeld:

a. Minimalisering van $||s||$ of $||g-s||$

b. Laat G_k de $k+1$ bij n matrix zijn van de gradienten g^h.
   $G_k = (G_k^{\ 1}, G_k^{\ 2})$ met $G_k^{\ 1}$ vierkant en niet-singulier.
   $G_k^{\ 1} s_1 + G_k^{\ 2} s_2 = 1$. Neem $s_2^{\ k} = 0$, $s_1^{\ k} = (G_k^{\ 1})^{-1}.e$ (waarbij e een vector, geheel bestaande uit enen, is). Op de numerieke uitwerking wordt niet ingegaan.

Bij een kwadratische functie zal minimalisering van $||s||_2$ leiden tot dezelfde methode als Fletcher en Reeves.

Bij methode bb zijn per stap gemiddeld n^2 vermenigvuldigingen nodig (excl. de staplengtebepaling) tegen $3n^2$ bij Davidon (eveneens excl. de staplengtebepaling). Over het verschil in convergentiesnelheid is niets bekend.