Convexe Analyse en Optimalisering

Bernd Heidergott
Vrije Universiteit Amsterdam en Tinbergen Institute
WEB: http://staff.feweb.vu.nl/bheidergott

Overzicht

Literatuur

- *Calculus, a complete course*, Robert A. Adams, Pearson Addison Wesley, 2006;
- *Optimization: Insights and Applications*, Jan Brinhhuis en Vladimir Tikhomirov, gaat verschijnen bij Princeton University Press, najaar 2005;
- *Inleiding in de Analyse*, B. Kaper en H. Norde, Academic Service, Schoonhoven, 1996;
- *Lagrange Multipliers and Optimality*, R. Rockafeller, SIAM Review. vol. 35, 1993, pp. 183-238.

Opbouw: 6 weken college á twee uur en 6 weken practicum á twee uur. Studiepunten 3 (ETCS), aantal uren dat de student wordt geacht voor het vak te studeren is 80 (4 x 6 = 24 contact uren + tentamenvoorbereiding + tentamen + zelfstudie).

College zaal: 4A-04, tijd: 13:30-15:15; data: 7.9, 14.9., 21.9., 28.9., 5.10., 12.10.

Practicum zaal: 4A-04, tijd: 11:00-12:45, data: 9.9., 16.9., 23.9., 30.9., 7.10., 14.10.

1 Basisbegrippen

1.1 Het Limietbegrijp

Definitie 1 *Zij* $a \in \mathbb{R}$ *en* $\varepsilon > 0$. *De verzameling* $U_\varepsilon(a) = \{x \in \mathbb{R} : |x - a| < \varepsilon\} = (a - \varepsilon, a + \varepsilon)$ *wordt de* ε*-omgeving van* a *genoemd.*

Zij $V \subset \mathbb{R}$. En getal $a \in V$ wordt een inwendig punt van V genoemd als er een $\varepsilon > 0$ bestaat zó, dat $U_\varepsilon \subset V$. Een getal $a \in V$ dat geen inwendig punt is wordt randpunt genoemd.

Definitie 2 *(Limiet) Zij* $\{x(n) : n \in \mathbb{N}\}$ *een rij in* $\mathbb{R}$ *and* a *een reëel getal. Het getal* a *wordt de limiet van de rij* $\{x(n) : n : \mathbb{N}\}$ *genoemd als er voor ieder* $\varepsilon > 0$ *een getal* $N_\varepsilon \in \mathbb{N}$ *bestaat zó, dat voor ieder natuurlijk getal* $n > N_\varepsilon$ *geldt*

$$|x(n) - a| < \varepsilon.$$

Wij noteren dit als

$$\lim_{n\to\infty} x(n) = a$$

of $x(n) \to a$ *als* $n \to \infty$.

Een rij, die een limiet heeft, wordt een convergente rij genoemd. Een rij, die geen limiet heeft, wordt een divergente rij genoemd.

De rij $\{x(n) : n \in \mathbb{N}\}$ *divergeert naar (plus) oneindig als voor iedere* $u \in \mathbb{R}$ *een getal* $N_u \in \mathbb{N}$ *bestaat zó, dat voor ieder natuurlijk getal* $n > N_u$ *geldt*

$$x(n) > u.$$

Wij noteren dit als

$$\lim_{n\to\infty} x(n) = \infty$$

of $x(n) \to \infty$ *als* $n \to \infty$.

Schrijfwijzen voor rijen: $\{x(n) : n \in \mathbb{N}\}$, $\{x(n)\}$, $\{x_n : n \in \mathbb{N}\}$, $\{x_n\}$, $(x(n))_{n=1}^\infty, \ldots$

Voorbeeld 1 *Wij tonen aan dat de limiet van de rij* $\{\frac{1+n}{n}\}$ *gelijk aan* 1 *is.*

Er geldt:

$$\left|\frac{1+n}{n} - 1\right| = \left|\frac{1}{n} + 1 - 1\right| = \left|\frac{1}{n}\right| = \frac{1}{n}$$

De fractie $1/n$ *wordt klein voor* n *groot. Zij* $\varepsilon > 0$. *Neem* $N_\varepsilon = \frac{1}{\varepsilon}$ *en vervolgens* $n > N_\varepsilon = \frac{1}{\varepsilon}$ *dan volgt*

$$\left|\frac{1+n}{n} - 1\right| = \frac{1}{n} < \frac{1}{\frac{1}{\varepsilon}} = \varepsilon.$$

Voorbeeld 2 [Optional, identiek met Opgave 1, Practicum week 1] *(Voorbeeld, p 71 K&N)*
Met de definitie van limiet tonen we aan dat $\lim_{n\to\infty} \frac{2n^2+3n+4}{2n^2+3} = 1$. *Zij* $\epsilon > 0$. *Neem* $N_\epsilon = \frac{2}{\epsilon}$. *Zij vervolgens* $n > N_\epsilon$. *Er geldt achtereenvolgens:*

$$\left|\frac{2n^2+3n+4}{2n^2+3} - 1\right| = \frac{3n+1}{2n^2+3} < \frac{3n+n}{2n^2} = \frac{2}{n} < \frac{2}{\frac{2}{\epsilon}} = \epsilon.$$

Hiermee hebben we bewezen dat $\lim_{n\to\infty} \frac{2n^2+3n+4}{2n^2+3} = 1$. **Klad:** *Hoe kom je aan* $N_\epsilon = \frac{2}{\epsilon}$ *? We hebben al gezien dat*

$$\left|\frac{2n^2+3n+4}{2n^2+3} - 1\right| < \frac{2}{n}.$$

Dus als $\frac{2}{n} < \epsilon$, *dan is*

$$\left|\frac{2n^2+3n+4}{2n^2+3} - 1\right| < \epsilon,$$

oftewel als $n > \frac{2}{\epsilon}$, *dan is*

$$\left|\frac{2n^2+3n+4}{2n^2+3} - 1\right| < \epsilon.$$

Het probleem is dus opgelost als we nemen $N_\epsilon = \frac{2}{\epsilon}$.

Rekenregels voor limieten: Als $\{x_n\}$ convergeert naar a en $\{y_n\}$ convergeert naar b, dan

(i) convergeert de somrij $\{x_n + y_n\}$ naar $a + b$

(ii) convergeert de produktrij $\{x_n y_n\}$ naar $a\,b$

(iii) convergeert de quotiëntrij $\{\frac{x_n}{y_n}\}$ naar $\frac{a}{b}$ $(b \neq 0)$

Opmerking: dit geldt ook voor divergente rijen.

Voorbeeld 3 *Wij tonen aan dat de limiet van de rij* $R = \{\frac{2^n+4\cdot3^n}{3^n+1}\}$ *gelijk aan* 4 *is. Wij gebruiken daarbij de eigenschap dat voor iedere* $a \in \mathbb{R}$ *met* $|a| < 1$ *de rij* $\{a^n\}$ *convergeert naar* 0. *Als we de teller en noemer van de* n^{de}*-term delen door* 3^n, *dan is de* n^{de}*-term te herschrijven als*

$$\frac{2^n + 4\cdot 3^n}{3^n+1} = \frac{\left(\frac{2}{3}\right)^n + 4}{1 + \left(\frac{1}{3}\right)^n}.$$

De rij R *laat zich ten behoeve van het convergentie-onderzoek als volgt ontleden:*

(stap 1) $\lim_{n\to\infty} \left(\frac{2}{3}\right)^n = 0$ *en* $\lim_{n\to\infty} \left(\frac{1}{3}\right)^n = 0$

(stap 2) *volgens de rekenregels voor de somrij geldt:* $\lim_{n\to\infty}(\left(\frac{2}{3}\right)^n + 4) = 0 + 4 = 4$ *en* $\lim_{n\to\infty}(1 + \left(\frac{1}{3}\right)^n) = 1 + 0 = 1$

(stap 3)   *volgens de quotiëntenregel convergeert de rij* R *naar* $\frac{4}{1} = 4$

Voorbeeld 4 **[Optional]** *(Voorbeeld, p80 K&N) Wij onderzoeken de convergentie van de rij* $R = \{\sqrt{n^2 + n} - n\}$ *en willen, in het geval van convergentie, weten wat de limiet is van de rij. [Om enigszins een idee te krijgen van het gedrag van de rij rekenen wij eerst een aantal termen uit. Het lijkt er inderdaad op dat er rij een 'plafond' bereikt.] Om de limiet daadwerkelijk te berekenen passen we de zogenaamde worteltruc toe. Als we* $t_n = \sqrt{n^2 + n} - n$ *vermenigvuldigen én tegelijkertijd delen door* $\sqrt{n^2 + n} + n$ *dan komt in de teller een zogenaamd merkwaardig product te staan van de vorm* $(a - b)(a + b)$ *dat gelijk is aan* $a^2 + b^2$*. Laten we eens kijken wat we daarmee opschieten:*

$$t_n = \left(\sqrt{n^2 + n} - n\right) \frac{\sqrt{n^2 + n} + n}{\sqrt{n^2 + n} + n} = \frac{n^2 + n - n^2}{\sqrt{n^2 + n} + n} = \frac{n}{\sqrt{n^2 + n} + n}.$$

Delen we tot slot teller en noemer door n *krijgen we*

$$t_n = \frac{1}{\sqrt{1 + \frac{1}{n}} + 1}.$$

De rij $(t_n)_{n=1}^{\infty}$ *laat zich aan de hand van de laatste uitdrukking voor* t_n *ten behoeve van het convergentieonderzoek als volgt 'ontleden': de rij* $(1 + \frac{1}{n})$ *convergeert met limiet 1, de rij* $\sqrt{1 + \frac{1}{n}}$ *convergeert met limiet* $\sqrt{1} = 1$ *(oefening), de rij* $\sqrt{1 + \frac{1}{n}} + 1$ *convergeert met limiet* 2*, en* t_n *convergeert met limiet* $\frac{1}{2}$*.*

Voorbeeld 5 **[Optional]** *(Voorbeeld, p107 K&N) We tonen aan dat de functie* f *op* $[0, \infty) \setminus \{1\} =: D_f$*, gedefinieerd door* $f(x) = \frac{3x-3}{\sqrt{x}-1}$*, de limiet* 6 *in* 1 *heeft.*

Er geldt

$$\begin{aligned}
\left| \frac{3x - 3}{\sqrt{x} - 1} - 6 \right| &= \left| \frac{3(x-1)}{\sqrt{x} - 1} \frac{\sqrt{x} + 1}{\sqrt{x} + 1} - 6 \right| \\
&= \left| \frac{3(x-1)(\sqrt{x} + 1)}{x - 1} - 6 \right| \\
&= |3(\sqrt{x} + 1) - 6| \\
&= |3(\sqrt{x} - 1)| \\
&= \left| 3 \frac{x - 1}{\sqrt{x} + 1} \right| \\
&\leq 3|x - 1|.
\end{aligned}$$

Als $|x - 1| < \delta$ *dan geldt* $3|x - 1| < 3\delta$*. Dus voor* $\epsilon > 0$ *volgt voor alle* x *met* $|x - 1| < \frac{\epsilon}{3} =: \delta_\epsilon$ *dat*

$$\left| \frac{3x - 3}{\sqrt{x} - 1} - 6 \right| < \epsilon.$$

Hiermee hebben we bewezen dat $\lim_{x \to 1} \frac{3x-3}{\sqrt{x}-1} = 6$*.*

Cauchy rij: Een $\{x(n) : n \in \mathbb{N}\}$ in $\mathbb{R}$ noemen wij een *Cauchy rij* als er voor ieder $\varepsilon > 0$ een getal $N_\varepsilon \in \mathbb{N}$ bestaat zó, dat voor alle natuurlijke getallen $m, n > N_\varepsilon$ geldt

$$|x(n) - x(m)| < \varepsilon.$$

Een rij $\{x(n) : n \in \mathbb{N}\}$ is een Cauchy rij dan en slechts dan als de rij convergent is.

N.B.: Opbouw van de reëele getallen. Een getal $x \in \mathbb{R}$ is de verzameling van alle Cauchy rijen uit de rationele getallen die dezelfde limiet x bezitten.

1.2 Continuïteit

Een functie f van een verzameling A naar de verzameling B is een voorschrift dat aan ieder element x van A precies één element $f(x)$ van B toevoegt. De verzameling A wordt het definitiegebied of domein van f genoemd en wordt ook aangegeven met D_f. We noteren de functie f van A naar B als

$$f : A \to B \quad \text{of} \quad f : D_f \to B.$$

We noemen $f(x)$ het beeld van x onder f. De verzameling van alle beelden wordt aan waardenverzameling of het bereik van f op A genoemd en aangeduid met R_f (van het Engelse 'Range') of $f(A)$.

Definitie 3 *Zij f een functie op een interval I en $c \in I$. De functie f wordt continu in c genoemd als er voor ieder $\varepsilon > 0$ een $\delta > 0$ bestaat zó, dat* **Def. 1.5, p. 18**

$$\forall x \in I : \quad |x - c| < \delta \Rightarrow |f(x) - f(c)| < \varepsilon.$$

De functie wordt continu genoemd als f continu is in ieder punt van I.

Een functie f op een interval I is dan en slechts dan continu in c als voor ieder rij $\{x_n\}$ in I, waarvoor $\lim_{n\to\infty} x_n = c$, geldt $\lim_{n\to\infty} f(x_n) = f(c)$.

Voorbeeld 6 *We tonen aan dat de functie f op $\mathbb{R}$, gedefinieerd door*

$$f(x) = \begin{cases} x^2 + 1 & \text{als } x \leq 2 \\ x + 4 & \text{als } x > 2 \end{cases}$$

niet continu is in 2. Beschouw de rij $\{x_n = 2 + \frac{1}{n}\}$. Dan geldt $x_n > 2$ voor iedere $n \in \mathbb{N}$ en $\lim_{n\to\infty} x_n = 2$. Voor de beeldrij geldt $f(x_n) = 6 + \frac{1}{n}$ voor iedere $n \in \mathbb{N}$ dus $\lim_{n\to\infty} f(x_n) = 6$. Als f continu was in 2 dan zou volgens het rijencriterium de beeldrij $\{f(x_n)\}$ limiet $f(2) = 5$ moeten hebben. Dit is niet het geval dus f is niet continu in 2.

Voorbeeld 7 **[Optional]** *(Voorbeeld p118 in K&N)*

- *De functie f op $\mathbb{R}$, gedefinieerd door $f(x) = x^3$, is continu in 2 omdat $f(2) = 2^3 = \lim_{x\to 2} x^3$.*

- *De functie f op $[0, \infty)$ gedefinieerd door*

$$f(x) \begin{cases} \frac{3x-3}{\sqrt{x}-1} & als\ x \neq 1 \\ 6 & als\ x = 1 \end{cases}$$

  is continu in 1. Immers, $f(1) = 6 = \lim_{x \to 1} \frac{3x-3}{\sqrt{x}-1}$, zie voorbeeld 5.

Een verzameling $V \subset \mathbb{R}$ wordt een open verzameling genoemd, als ieder punt van V een inwendig punt van V is. Een verzameling $V \subset \mathbb{R}$ wordt een gesloten verzameling genoemd, als het complement van V open is. Een verzameling $V \subset \mathbb{R}$ wordt begrensd genoemd, als er een $m > 0$ bestaat zó, dat voor iedere $x \in V$ geldt $|x| \leq m$. Een deelverzameling V van $\mathbb{R}$ wordt compact genoemd, als V begrensd én gesloten is. Voorbeelden: intervallen $[a, b]$, $a < b$ en $a, b \in \mathbb{R}$, zijn compact. $(a, b]$ en $\mathbb{R}$ zijn niet compact.

Theorem 1 *(stelling van Weierstrass) Iedere continue functie f op een compact interval I heeft een minimum en een maximum.*

Th. 1.6, p. 18 en 544+545 uit B & V volgt Derivation 2, p545 uit B & V

Bewijs: Zij f een continue functie op een compact interval $I = [a, b]$. Dan is f begrensd. Wij kunnen dus zinvol spreken over

$$\inf f(I) = \inf\{f(x) : x \in I\}$$

en

$$\sup f(I) = \sup\{f(x) : x \in I\}.$$

We moeten laten zien dat er $c \in I$ bestaat zó, dat $f(c) = \inf f(I)$ en een $d \in I$ zó, dat $f(d) = \sup f(I)$. We tonen alleen het bestaan van het maximum aan. Uit de bewering $u = \sup f(I)$ volgt dat voor ieder $\varepsilon > 0$ een $x \in I$ bestaat zó dat $u - \varepsilon < f(x)$. Bij achtereenvolgens $\varepsilon = 1, 1/2, 1/3, \ldots$ bestaan dus punten $x_1, x_2, x_3, \ldots$ met de eigenschap $a \leq x_n \leq b$ en $u - \frac{1}{n} < f(x_n) \leq u$ voor ieder $n \in \mathbb{N}$. De beeldrij $\{f(x_n)\}$ convergeert dus naar u. Omdat $a \leq x_n \leq b$ voor iedere $n \in \mathbb{N}$ heeft de rij $\{x_n\}$ een convergente deelrij[1] met limiet $d \in [a, b]$. Omdat f continu is in d convergeert de beeldrij naar $f(d)$. $\square$

N.B. Voor een rij $\{x_n\}$ is de *limiet superior* gedefinieerd als

$$\lim_{N \to \infty} \sup\{x_n : n \geq N\}$$

en de *limiet inferior* gedefinieerd als

$$\lim_{N \to \infty} \inf\{x_n : n \geq N\}.$$

De $\lim \inf$ en $\lim \sup$ bestaan altijd. De rij is convergent dan en slechts dan als $\lim \inf x_n = \lim \sup x_n$.

1.3 De Afgeleide

Definitie 4 *Zij f een functie op een interval $I \subset \mathbb{R}$ en $c \in I$. We noemen de functie f differentieerbaar in c als $\lim_{x \to c} \frac{f(x)-f(c)}{x-c}$ bestaat. In dat geval wordt de waarde van de limiet de afgeleide van f in c genoemd en geschreven als $f'(c)$.* **Def. 1.9, p. 29 uit B & V**

De functie f wordt differentieerbaar genoemd als f differentieerbaar is in ieder punt van I. In dat geval wordt de functie, die aan ieder punt $x \in I$ de afgeleide van f toevoegt, de afgeleide functie van f genoemd en genoteerd met f'.

Andere notaties voor de afgeleide van f in c zijn $\frac{df}{dx}(c)$, $Df(c)$, $\frac{d}{dx}\big|_{x=c} f, \ldots$

Voorbeeld 8 *We tonen aan dat de functie f op $(0, \infty)$, gedefinieerd door $f(x) = \frac{1}{\sqrt{x}}$, differentieerbaar is en dat de afgeleide f' op $(0, \infty)$ gegeven wordt door $f'(x) = \frac{-1}{2x\sqrt{x}}$. Zij $c > 0$. Er geldt voor $\Delta \neq 0$ (met $c + \Delta x > 0$) dat*

$$\begin{aligned}
\frac{\frac{1}{\sqrt{c+\Delta x}} - \frac{1}{\sqrt{c}}}{\Delta x} &= \frac{1}{\Delta x}\left(\frac{\sqrt{c} - \sqrt{c+\Delta x}}{\sqrt{c}\sqrt{c+\Delta}}\right) \\
&= \frac{1}{\Delta x}\left(\frac{c-(c+\Delta x)}{\sqrt{c}\sqrt{c+\Delta}(\sqrt{c}+\sqrt{c+\Delta})}\right) \\
&= \frac{-1}{\sqrt{c}\sqrt{c+\Delta}(\sqrt{c}+\sqrt{c+\Delta})}
\end{aligned}$$

zodat $f'(c) = \frac{-1}{2c\sqrt{c}}$. Aangezien dit geldt voor iedere $c > 0$ is de functie f differentieerbaar en wordt de afgeleide functie f' op $(0, \infty)$ gegeven door $f'(x) = \frac{-1}{2x\sqrt{x}}$.

Voorbeeld 9 *(Voorbeeld p138 K&N en Voorbeeld 1.2.2 p. 8 in B&V)*
De (absolute-waarde) functie f op $\mathbb{R}$, gedefinieerd door $f(x) = |x|$, is continu in 0. Laat zien dat

$$f'(x) = \frac{x}{|x|}, \quad x \neq 0.$$

Voor $x = 0$ moten we de limiet

$$\lim_{\Delta \to 0} \frac{|\Delta|}{\Delta}$$

onderzoeken. Er geldt

$$\lim_{\Delta \downarrow 0} \frac{|\Delta|}{\Delta} = 1 \neq -1 = \lim_{\Delta \uparrow 0} \frac{|\Delta|}{\Delta}.$$

Dus de limiet bestaat niet.

Rekenregels voor differentieerbare functies: Voor functies f en g op een interval geldt, als f en g differentieerbaar in $c \in I$ zijn: **Sect. B.2 p. 523-525 uit B & V**

(i) (somregel) $(f+g)'(c) = f'(c) + g'(c)$

(ii) (productregel) $(fg)'(c) = f'(c)g(c) + f(c)g'(c)$

[1]Bolzano-Weierstrass, ieder begrensde rij in $\mathbb{R}$ heeft tenminste één convergente deelrij.

(iii) (quotiëntenregel)
$$\left(\frac{f}{g}\right)(c) = \frac{f'(c)g(c) - f(c)g'(c)}{(g(c))^2}$$

(iv) (kettingregel) Zij g nu een functie op een interval J zó, dat $g(J) \subset I$ en $c \in J$. Als g differentieerbaar in c is en f in $g(c)$, dan is de samengestelde functie $f \circ g$ differentieerbaar in c en
$$(f \circ g)'(c) = f'\big(g(c)\big) \cdot g'(c).$$

Voorbeeld 10 *(Voorbeeld 1.2.1 p. 8 in B&V) (Alternatief practicum) We tonen aan dat voor iedere $n \geq 1$ de functie $f(x) = x^n$ op $\mathbb{R}$ differentieerbaar is en dat $f'(x) = nx^{n-1}$. We maken gebruik van het principe van volledige inductie. De bewering 'de functie f op $\mathbb{R}$, gedefinieerd door $f(x) = x^n$, is differentieerbaar en $f'(x) = nx^{n-1}$ voor ieder $x \in \mathbb{R}$' noemen wij $P(n)$.*

stap 1 *($P(1)$ is waar) De functie $f(x) = x$ is differentieerbaar op $\mathbb{R}$ en $f'(x) = 1$ ($\frac{f(x+h)-f(x)}{h} = 1$ voor ieder h).*

stap 2 *(voor ieder $n \in \mathbb{N}$ geldt $P(n) \Rightarrow P(n+1)$) Zij $n \in \mathbb{N}$. Neem aan dat $f'(x) = nx^{n-1}$ voor ieder $x \in \mathbb{R}$. De functie g op $\mathbb{R}$ gedefinieerd door $g(x) = x^{n+1}$ is te schrijven als het product van de functie h op $\mathbb{R}$, gedefinieerd door $h(x) = x$ en f. Volgens de productregel voor het differentiëren is g, als product van twee differentieerbare functies h en f, differentieerbaar en geldt voor iedere $x \in \mathbb{R}$*
$$g'(x) = h'(x)f(x) + h(x)f'(x) = x^n + xnx^{n-1} = (n+1)x^n.$$
Daarmee is aangetoond dat de bewering $P(n+1)$ waar is.

Drie visies op de afgeleide:

- de fysieke interpretatie,
- de geometrische interpretatie, en
- de analytische (approximatieve) interpretatie.

Alleen de analytische interpretatie kan uitgebreid worden naar functies van meerdere variabelen.

Zij r een functie op $\mathbb{R}$. Wij schrijven $r(h) = o(h)$ als geldt
$$\lim_{h \to 0} \frac{|r(h)|}{|h|} = 0.$$

Voorbeeld $r(h) = h^2$. Interpretatie: door h klein te kiezen wordt $r(h)$ willekeurig klein.

Definitie 5 *Zij f een functie op een interval $I \subset \mathbb{R}$ en $c \in I$. We noemen de functie f differentieerbaar in c als*

Def. 1.1 p. 7 uit B & V

$$f(c+h) - f(c) = ah + r(h)$$

voor $r(h) = o(h)$ en $a \in \mathbb{R}$. In dat geval wordt a de afgeleide van f in c genoemd en geschreven als $f'(c)$.

Alternatief kunnen wij schrijven

$$f(c+h) = f(c) + ah + r(h),$$

$$f(c+h) = f(c) + ah + o(h)$$

of

$$|f(c+h) - f(c) - ah| \to 0 \quad \textit{als} \quad h \to 0.$$

Theorem 2 *(de stelling van Taylor) Zij f een tweemaal differentieerbare functie op een open interval $I \subset \mathbb{R}$ en $c \in I$. Er geldt: voor iedere $x \in I$, $x \neq c$, bestaat er een punt ξ_x tussen c en x zó, dat*

zie ook Voorbeeld 1.6.39 p. 83 uit B & V

$$f(x) = f(c) + f'(c)(x-c) + \frac{1}{2} f"(\xi_x)(x-c)^2.$$

De functie $f(c) + f'(c)(x-c)$ wordt het eerste order Taylor polynoom van f in c genoemd.

Bewijs: De rest r op I, gedefinieerd door

$$r(x) = f(x) - \big(f(c) - f'(c)(x-c)\big),$$

is tweemaal differentieerbaar. Bovendien is $r(c) = r'(c) = 0$ en $r''(x) = f''(x)$ voor iedere $x \in I$. We zullen nu aantonen dat er voor iedere $x \in I \setminus \{c\}$ een punt ξ_x bestaat tussen c en x zó, dat

$$r(x) = \frac{r''(\xi_x)}{2}(x-c)^2.$$

Definieer de helpfunctie h op $\mathbb{R}$ door $h(x) = (x-c)^2$. Zij $x \in I \setminus \{c\}$. Neem aan dat $x > c$. (Het bewijs als $x < c$ gaat analoog.) Volgens de Cauchy middelwaardestelling[2], toegepast op de rest r en de functie h, beperkt tot het interval $[c, x]$, bestaat er een $\xi_1 \in (c, x)$ zodat

$$\frac{r(x)}{h(x)} = \frac{r(x) - r(c)}{h(x) - h(c)} = \frac{r'(\xi_1)}{h'(\xi_1)}. \tag{1}$$

Passen we nogmaals de Cauchy middelwaardestelling toe, maar nu op de functies r' en h' beperkt tot het interval $[c, \xi_1]$, dan vinden we een $\xi_2 \in (c, \xi_1)$ met

$$\frac{r'(\xi_1)}{h'(\xi_1)} = \frac{r'(x_1) - r'(c)}{h'(x_1) - h'(c)} = \frac{r''(\xi_2)}{h''(\xi_2)}. \tag{2}$$

[2]stelling van Rolle $\to$ Cauchy middelwaardestelling

Omdat $h''(\xi_2) = 2$, is

$$\frac{r(x)}{h(x)} \overset{(1)}{=} \frac{r'(\xi_1)}{h'(\xi_1)} \overset{(2)}{=} \frac{r''(\xi_2)}{h''(\xi_2)},$$

dus

$$r(x) = \frac{r''(x_2)}{2} h(x) = \frac{r''(\xi_2)}{2} (x-c)^2.$$

Voor $\xi_x = \xi_2$ hebben we de stelling aangetoond. $\square$

2 Practicum Week 1

Opgave 1 (Voorbeeld, p 71 K&N) Toon aan dat

$$\lim_{n\to\infty} \frac{2n^2+3n+4}{2n^2+3} = 1.$$

Opgave 2 (Voorbeeld, p80 K&N) Onderzoek de convergentie van de rij

$$\left\{\sqrt{n^2+n}-n\right\}.$$

(Veronderstel dat $\sqrt{x}$ continu is.)

Opgave 3 (Voorbeeld, p107 K&N) Toon aan dat de functie f op $[0,\infty)\setminus\{1\}$, gedefinieerd door $f(x)=\frac{3x-3}{\sqrt{x}-1}$, de limiet 6 in 1 heeft.

Opgave 4 (Voorbeeld, p118 K&N) Toon aan dat de functie f op $\mathbb{R}$, gedefinieerd door $f(x)=x^3$, continu is. (Hint: toepassing van de rekenregels voor limieten.)

Opgave 6 Laat zien dat de limiet $\lim_{x\to 0}\sin\left(\frac{1}{x}\right)$ niet bestaat en de limiet $\lim_{x\to 0} x\sin\left(\frac{1}{x}\right)$ wél.

Opgave 7 (Repititorium [B], p191) Laat zien dat

$$f(x)=\begin{cases} x\sin\left(\frac{1}{x}\right) & \text{als } x\neq 0\\ 0 & \text{als } x=0\end{cases}$$

niet differentieerbaar in 0.

Opgave 8 (Stelling van Rolle) Zij f een continue functie op het interval $[a,b]$, die differentieerbaar is op (a,b). Toon aan dat geldt: als $f(a)=f(b)$, dan bestaat er tenminste één (tussen)-punt $\xi\in(a,b)$ zó, dat $f'(\xi)=0$.

Opgave 9 (Middelwaardstelling) Zij f een continue functie op het interval $[a,b]$, die differentieerbaar is op (a,b). Toon aan dat geldt: er bestaat tenminste één (tussen)-punt $\xi\in(a,b)$ zó, dat

$$\frac{f(b)-f(a)}{b-a}=f'(\xi) \quad \text{of} \quad f(b)-f(a)=f'(\xi)(b-a).$$

Opgave 10 (Cauchy middelwaardstelling) Zijn f en g continue functies op het interval $[a,b]$, die differentieerbaar zijn op (a,b) en waarvoor $g(x)\neq 0$ voor iedere $x\in(a,b)$. Toon aan dat geldt: er bestaat tenminste één (tussen)-punt $\xi\in(a,b)$ zó, dat

$$\frac{f(b)-f(a)}{g(b)-(a)}=\frac{f'(\xi)}{g'(\xi)}.$$

Opgave 11 (Supremum/Infimum) Geef een definitie van het begrip supremum (infimum). **zie ook p. 520 in Sect. B.1 uit B & V**

3 Optimalisering over $\mathbb{R}$

3.1 Fermat's theorema voor een variable

Probleem $(P_{1.1})$ Zij f een functie op $\mathbb{R}$: **Def. 1.2 p. 14 uit B & V**

$$f(x) \to \text{ extr.}$$

noemen wij een optimaliseringsprobleem zonder restricties.

Definitie 6 *Zij f een functie op een interval $I \subset \mathbb{R}$ en $c \in I$. Wij noemen de functiewaarde $f(c)$ een lokaal minimum van f als een voor $\varepsilon > 0$ geldt $f(x) \geq f(c)$ voor ieder $x \in (c-\varepsilon, c+\varepsilon) \cap I$. Het punt c wordt een minimumlokatie van f genoemd en $f(c)$ de waarde van c.* **Def. 1.3 p. 14 uit B & V**

Wij noemen de functiewaarde $f(c)$ een lokaal maximum van f als een voor $\varepsilon > 0$ geldt $f(x) \leq f(c)$ voor ieder $x \in (c-\varepsilon, c+\varepsilon) \cap I$. Het punt c wordt een maximumlokatie van f genoemd en $f(c)$ de waarde van c.

Wij noemen een minimum (maximum) globaal als voor ieder $x \in I$ geldt $f(x) \geq f(c)$ ($f(x) \leq f(c)$).

Zij f een differentieerbare functie op een interval $I \subset \mathbb{R}$. Een punt $c \in I$ heet stationair punt als $f'(c) = 0$.

Theorem 3 *(Theorem 1.4 p. 15 uit B&V) Zij probleem $(P_{(1.1)})$ gegeven en zij f differentieerbaar in c. Als c een lokaal extremum is, dan geldt $f'(c) = 0$. [Ieder extremum is een stationair punt.]*

Bewijs: Neem aan dat c een inwendig punt is van D_f en dat f differentieerbaar is in c. Neem bovendien aan dat $f(c)$ een lokaal maximum is van f. Dan bestaat een $\varepsilon > 0$ zó, dat voor iedere $x \in U_\varepsilon(c)$ geldt $f(x) \leq f(c)$, oftewel $f(x) - f(c) \leq 0$. Omdat f differentieerbaar is in c, is $\lim_{x\to\infty} \frac{f(x)-f(c)}{x-c} = f'(c)$. Er geldt dus voor iedere rij $\{x_n\}$ in D_f, waarvoor $x_n \neq c$ voor $n \in \mathbb{N}$ en $\lim_{n\to\infty} x_n = c$, dat

$$\lim_{n\to\infty} \frac{f(x_n) - f(c)}{x_n - c} = f'(c).$$

Neem een rij waarvoor $c - \varepsilon < x_n < c$ voor iedere $n \in \mathbb{N}$ en $\lim_{n\to\infty} x_n = c$. Dan is

$$\frac{f(x_n) - f(c)}{x_n - c} \geq 0$$

voor iedere $n \in \mathbb{N}$, en dus $f'(c) \geq 0$. Neem vervolgens een rij waarvoor $c < x_n < \varepsilon + c$ voor iedere $n \in \mathbb{N}$ en $\lim_{n\to\infty} x_n = c$. Dan is

$$\frac{f(x_n) - f(c)}{x_n - c} \leq 0$$

voor iedere $n \in \mathbb{N}$, en dus $f'(c) \leq 0$. Samenvattend geldt dus $f'(c) = 0$. Als $f(c)$ een minimum is verloopt het bewijs analoog. $\square$

Voorbeeld 1.3.1 p. 17 uit B&V. ($f(x) = x + 5x^{-1} \to \min$).

Voorbeeld 1.3.2 p. 17 uit B&V. $f'(c) = 0$ is niet toereikend voor global extrema (alleen noodzakelijk).

Theorem 4 *(tweede orde voorwaarde voor extremum) Zij f een tweemaal differentieerbare functie op een interval I en c een inwendig punt van I zó, dat $f'(c) = 0$ en f'' continu is in c. Er geldt:*

a) als $f''(c) > 0$, dan is $f(c)$ een lokaal minimum van f

b) als $f''(c) < 0$, dan is $f(c)$ een lokaal maximum van f.

Bewijs: Deel a). Neem aan dat $f''(c) > 0$. Omdat f'' continu is in c, bestaat er een ε-omgeving van c, $U_\varepsilon(c)$, zó, dat $f''(c) > 0$ voor iedere $x \in U_\varepsilon(c)$. Zij $x \in U_\varepsilon(c) \setminus \{c\}$. Volgens de stelling van Taylor bestaat er bij x een punt ξ_x tussen c en x zó, dat

$$f(x) = f(c) + f'(c)(x-c) + \frac{f''(\xi_x)}{2}(x-c)^2 = f(c) + \frac{f''(\xi_x)}{2}(x-c)^2$$

omdat $f'(c) = 0$. Omdat $x \in U_\varepsilon(c)$, behoort het tussenpunt ξ_x ook tot $U_\varepsilon(c)$ en is dus $f''(\xi_x) > 0$. Omdat $(x-c)^2 > 0$, is dus

$$f(x) = f(c) + \frac{f''(\xi_x)}{2}(x-c)^2 \geq f(c) + 0 = f(c)$$

en $f(c)$ is een lokaal minimum.

Deel b). Neen aan dat $f'(c) = 0$ en $f''(c) < 0$. Dan is $(-f)'(c) = 0$ en $(-f)''(c) > 0$. Volgens onderdeel (a) is $(-f)(c)$ dan een lokaal minimum van $-f$, zodat $f(c)$ een lokaal maximum is van f.

$\square$

Wat betekend het als $f''(x) = 0$? In dit geval kan het punt x een maximumlokatie, een minimumlokatie of een buigpunt zijn. Een buigpunt $\hat{x}$ is een lokaal extremum van $f'(x)$, dat wil zeggen, $\hat{x}$ is een buigpunt wanneer $f''(\hat{x}) = 0$ en $f'''(\hat{x}) \neq 0$ is. In het geval $f'''(\hat{x}) = 0$, moet de functie verder onderzoekt worden. Het is niet mogelijk om een gesloten algoritme aan te geven voor het bepalen van de aard van een stationair punt. Bijvoorbeeld heeft de functie

$$f(x) = \begin{cases} e^{-\frac{1}{x^2}} & x \neq 0 \\ 0 & x = 0, \end{cases}$$

$\hat{x} = 0$ als stationair punt en $\hat{x}$. Alle afgeleiden zijn echter bij $\hat{x}$ null, $f^{(n)}(0) = 0$, $n \geq 1$.

Test via de Nde afgeleide: Zij f N keer differentieerbaar in een punt c zó dat $f^{(n)}(c) = 0$ voor $1 \leq n < N$ en $f^{(N)}(c) \neq 0$. Voor N even (dus $N \in \{2n : n \in \mathbb{N}\}$)

geldt: c is een lokaal minimum als $f^{(N)}(c) > 0$, en c is een lokaal maximum als $f^{(N)}(c) < 0$. Voor N oneven ($N \in \{2n+1 : n \in \mathbb{N}\}$) is c een buigpunt.

Voorbeeld 1.3.1 nog een keer via tweede orde voorwaarden.

Vinden van het globaal minimum: via tweede orde voorwaarden alle (lokale) minima berekenen en door vergelijking van de waarden van de minima het globale minimum identificeren.

3.2 De vier stappen methode

Eenvoudiger (en numeriek efficiënter) is het eerst aan te tonen dat er een globaal minimum bestaat en dan de waarden van de stationaire punten te checken.

Een functie op $\mathbb{R}$ heet *coercive* ('dwingend') (voor minimalisering) als $\lim_{|x|\to\infty} f(x) = +\infty$.

Corollary 1 *(Corollary 1.7 p. 19 uit B&V) Zij f een 'coercive' en continue functie op* $\mathbb{R}$*, dan heeft f een globaal minimum.*

De vier stappen methode:

1. wiskundig model voor het probleem
2. bepaal de noodzakelijke voorwaarden voor een extrema
3. los de vergelijkingen op
4. conclusie

Voorbeeld 1.3.4 p. 20 (voorbeeld 1.3.1 voortgezet) uit B&V p.29. ($f(x) = x+5x^{-1} \to \min$).

Voorbeeld 1.3.7 p. 24 uit B&V ($f(x) = ax^2 + 2bx + c \to \min$).

Corollary 1.8 + voorbeeld 1.3.6 p. 23 uit B&V (($x = \cos x$) beschrijft globaal minimum.)

Remark p. 31 uit B&V [**Optional**]

4 Practicum Week 2

Opgave 1 (Remark p. 31 uit B&V) Laat zien dat voor $a > 0$ de oplossing voor

$$\frac{1}{2}x_1x_2 \to \max,\ x_1 + x_2 = a,\ x_1, x_2 > 0$$

gegeven is door $\hat{x}_1 = \hat{x}_2 = a/2$.

Opgave 2 (Exercise 1.6.2 uit B&V)

$$f(x) = \sqrt{|x^2 - 1|} - x \to \text{extr}$$

Opgave 3 (Exercise 1.6.3 uit B&V)

$$f(x) = \sqrt{x^2 - x^4} \to \text{extr},\ -1 \leq x \leq 1$$

Opgave 4 (Exercise 1.6.4 uit B&V)

$$f(x) = \sqrt{x^2 + 1} - \frac{1}{2}x \to \text{min}$$

Opgave 5 (Exercise 1.6.5 uit B&V)

$$f(x) = e^{x_1 x_2} \to \text{extr}, x_1 + x_2 = 1$$

Opgave 6 (Exercise 1.6.15 uit B&V)

$$f(n) = \sqrt[n]{n} \to \text{extr},\ n \in \mathbb{N}$$

5 Fermat's theorema voor meerdere variabele)

5.1 De n dimensionale ruimte

Definitie 7 *Zij $x \in \mathbb{R}^n$. De (euclidische) norm van x, notatie $|x|$, is gedefinieerd door*

$$|x| = \sqrt{x_1^2 + \cdots + x_n^2}.$$

Voor het product van een getransponeerde vector uit $\mathbb{R}^n$ met een vector uit $\mathbb{R}^n$ is geldt:

$$x^\top y = \sum_{i=1}^n x_i y_i.$$

(Dit product heet "dot" product, $x \bullet y$, in R. Adams.) Verder is de lengte van de vector x gegeven door $|x| = \sqrt{x^\top x}$.

Geometrische interpretatie: gegeven $x, y \in \mathbb{R}^n$ en zij de hoek tussen x en y gelijk aan ϑ, dan geldt

$$x^\top y = |x|\,|y|\,\cos(\vartheta)$$

Th. 1 in Sec. 10.2, p. 549, Adams

Wij noemen vectoren $x, y \in \mathbb{R}$ orthogonaal als $x^\top y = 0$.

Er geldt: $|x+y|^2 = |x|^2 + 2x^\top y + |y|^2$. Dit ziet men zó:

$$\begin{aligned} |x+y|^2 &= \sum (x_i + y_i)^2 \\ &= \sum (x_i^2 + 2x_i + y_i + y_i^2) \\ &= \sum x_i^2 + 2\sum x_i y_i + \sum y_i^2 \\ &= |x|^2 + 2x^\top y + |y|^2\,. \end{aligned}$$

Theorem 2.1 p. 89 uit het B. & T. (Cauchy-Bunyakovskii-Schwartz ongelijkheid en driehoeksongelijkheid)
Uitwerking:

$$(1) \quad x^\top y \le |x^\top y| \le |x|\,|y|$$

De linker ongelijkheid is triviaal omdat $x^\top y \in \mathbb{R}$ is. Zij $x \neq 0$ ($x = 0$ triviaal). Beschouw het probleem

$$f(\alpha) = |\alpha x + y|^2 \to \min\,.$$

Er geldt

$$f(\alpha) = |\alpha x + y|^2 = \alpha^2 |x|^2 + \alpha 2x^\top y + |y|^2$$

en f is coercief voor minimaliseering: $\lim f_{|\alpha| \to \infty} f(\alpha) = \infty$.

$$f'(\alpha) = 2\alpha |x|^2 + 2x^\top y$$

$$f'(\alpha) = 0 \Leftrightarrow \alpha = -\frac{x^\top y}{|x|^2}$$

De waarde van het minimum is

$$\begin{aligned} f\left(-\frac{x^\top y}{|x|^2}\right) &= \frac{(x^\top y)^2}{|x|^4}|x|^2 - 2\frac{x^\top y}{|x|^2}x^\top y + |y|^2 \\ &= \frac{(x^\top y)^2}{|x|^2} - 2\frac{(x^\top y)^2}{|x|^2} + |y|^2 \\ &= \frac{|x|^2|y|^2 - (x^\top y)^2}{|x|^2} . \end{aligned}$$

Omdat $f(\alpha) \geq 0$ volgt

$$\frac{|x|^2|y|^2 - (x^\top y)^2}{|x|^2} \geq 0$$

oftewel

$$|x|^2|y|^2 \geq (x^\top y)^2$$

worteltrekken

$$|x|\,|y| \geq |x^\top y|$$

Let op $x^\top y$ kan negatieve waarden aannemen.

Bewijs van onderdeel (2): Met onderdeel (1) volgt

$$\begin{aligned} |x+y|^2 &= |x|^2 + 2x^\top y + |y|^2 \\ &\leq |x|^2 + 2|x|\,|y| + |y|^2 \qquad \text{(gebruik hier (1))} \\ &= (|x|+|y|)^2. \end{aligned}$$

Definitie 8 *Zij $a \in \mathbb{R}^n$ en $\varepsilon > 0$. De verzameling $U_n(a,\varepsilon)$,*

$$U_n(a,\varepsilon) = \{x \in \mathbb{R}^n : |x-a| < \varepsilon\}$$

wordt de ε-omgeving van a in $\mathbb{R}^n$ genoemd of de open bal in $\mathbb{R}^n$ om a met radius ε.

Zij $r : \mathbb{R}^n \to \mathbb{R}$. Wij schrijven $r(h) = o(h)$ als geldt

$$\lim_{|h|\to 0} \frac{|r(h)|}{|h|} = 0.$$

Een verzameling $V \subset \mathbb{R}^n$ wordt een open verzameling genoemd, als ieder punt van V een inwendig punt van V is. Een verzameling $V \subset \mathbb{R}^n$ wordt een gesloten verzameling genoemd, als het complement van V open is. Een verzameling $V \subset \mathbb{R}^n$ wordt begrensd genoemd, als er een $m > 0$ bestaat zó, dat voor iedere $x \in V$ geldt $|x| \leq m$. Een deelverzameling V van $\mathbb{R}^n$ wordt compact genoemd, als V begrensd én gesloten is.

Definitie 9 *Een symmetrische $n \times n$ matrix A heet*

positief definiet [3] *als voor ieder $x \in \mathbb{R}^n$, $x \neq 0$, geldt: $x^\top Ax > 0$*

[3]Een positief definiete matrix A is richtingsbehoudend: $x^T Ax > 0 \Leftrightarrow |x||Ax|\cos(\eta) > 0$, met η de hoek tussen x en Ax; omdat $\cos(\eta) > 0$ volgt dat $\eta \in (-\pi/2, \pi/2)$.

positief semi-definiet *als voor ieder* $x \in \mathbb{R}^n$*,* $x \neq 0$*, geldt:* $x^\top Ax \geq 0$

negatief definiet *als voor ieder* $x \in \mathbb{R}^n$*,* $x \neq 0$*, geldt:* $x^\top Ax < 0$

negatief semi-definiet *als voor ieder* $x \in \mathbb{R}^n$*,* $x \neq 0$*, geldt:* $x^\top Ax \leq 0$

indefiniet *als er een* $x_1 \in \mathbb{R}^n$*, bestaat waarvoor* $x_1^\top Ax_1 > 0$ *en een* $x_2 \in \mathbb{R}^n$ *waarvoor* $x_2^\top Ax_2 < 0$

Voorbeeld 11 *De symmetrische matrix*

$$A = \begin{pmatrix} 3 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}$$

is positief definiet. Immers,

$$x^\top Ax = (x_1, \ldots, x_n) \begin{pmatrix} \sum_j a_{1j}x_j \\ \sum_j a_{2j}x_j \\ \vdots \\ \sum_j a_{nj}x_j \end{pmatrix} = \sum_j x_1 a_{1j} x_j + \sum_j x_2 a_{2j} x_j + \cdots + \sum_j x_n a_{nj} x_j$$

en de kwadratische functie

$$x^\top Ax = 3x_1^2 + 2x_2^2 + 2x_3^2 + 2x_1x_2 + 2x_2x_3 + 2x_1x_3$$

is door afsplitsen van de kwadraten te schrijven als

$$x_1^2 + (x_1 + x_2)^2 + (x_1 + x_3)^2 + (x_2 + x_3)^2.$$

De som van kwadraten is > 0 *tenzij* $x_1 = x_1 + x_2 = x_1 + x_3 = 0$ *oftewel* $x = 0$*.*

5.2 Afgeleiden

Definitie 10 *Zij* f *een functie op een open bal* $U_n(c, \varepsilon)$*. We noemen de functie* f *differentieerbaar in* c *als er een rijvector* $f'(c)$ *bestaat zó, dat*

$$f(c + h) = f(c) + f'(c) \cdot h + r(h)$$

voor $r(h) = o(h)$ *en* $c + h \in U_n(c, \varepsilon)$*. In dat geval wordt* a *de afgeleide van* f *in* c *genoemd en geschreven als* $f'(c)$ *of* $Df(c)$*.*

Voorbeeld 2.2.1 p. 93 uit B&V ($f(x) = x_1^2 + x_2^2$).

Definitie 11 *(eerste afgeleide) Zij* $f : U_n(c, \varepsilon) \to \mathbb{R}$*. Als de snedefunctie* g *van* f *in* c *met betrekking tot de* k*-de variable* x_k*, gedefinieerd door*

$$g(x_k) = f(c_1, \ldots, c_{k-1}, x_k, c_{k+1}, \ldots, c_n)$$

differentieerbaar is in c_k dan wordt de functie f partieel differentieerbaar naar de k-de variable in c genoemd. De afgeleide $g'(c_k)$ wordt de partieële afgeleide van f naar de k-de variable in c genoemd en genoteerd als $\frac{\partial f(c)}{\partial x_k}$. Wij noemen f partieel differentieerbaar in c als $\frac{\partial f(c)}{\partial x_k}$, $1 \leq k \leq n$, bestaan:

$$f'(c) = \left(\frac{\partial f(c)}{\partial x_1}, \ldots, \frac{\partial f(c)}{\partial x_n} \right).$$

De rijvector $f'(c)$ wordt gradiënt genoemd[4]*. Wij noemen f partieel differentieerbaar als f partieel differentieerbaar in ieder punt van $U_n(c, \varepsilon)$.*

Een functie, die differentieerbaar is in een punt, is continu in dat punt. (Dit volgt uit $f(c+h) - f(c) = f'(c) \cdot h + r(h)$.)

Voorbeeld 12 *De functie f op $\mathbb{R}^2$, gedefinieerd door*

$$f(x,y) = \begin{cases} \frac{xy}{x^2+y^2} & \textit{als } (x,y) \neq (0,0) \\ 0 & \textit{als } (x,y) = (0,0) \end{cases}$$

is in $(0,0)$ niet continu en dus ook niet differentieerbaar. De functie f is wel partieel differentieerbaar in $(0,0)$. De snedefunctie van f in $(0,0)$ wordt gegeven door

$$g(x) = f(x,0) = \begin{cases} 0 & \textit{als } x \neq 0 \\ 0 & \textit{als } x = 0. \end{cases}$$

Dus $\frac{\partial f}{\partial x}(0,0) = 0$. Op analoge wijze vind je dat $\frac{\partial f}{\partial y}(0,0) = 0$. Dus $f'(0,0) = (0,0)$.

Theorem 5 *Zij $f : D_f \subset \mathbb{R}^n \to \mathbb{R}$ en c een inwendig punt van D_f. Er geldt: als f partieel differentieerbaar is op een omgeving U van c en de partiële afgeleiden zijn continu zijn in c, dan is f differentieerbaar in c en is $f'(c)$ is gelijk aan de gradiënt van f in c.*

(Zonder bewijs)

Definitie 12 *(tweede afgeleide) Zij $f : U_n(c,\varepsilon) \to \mathbb{R}$ een reëelwaardige functie, die tweemaal continu differentieerbaar is op een omgeving van $x \in U_n(c,\varepsilon)$. De $n \times n$ matrix $Hf(x)$,*

$$Hf(x) = \begin{pmatrix} \frac{\partial}{\partial x_1}\left(\frac{\partial}{\partial x_1} f(x)\right) & \frac{\partial}{\partial x_2}\left(\frac{\partial}{\partial x_1} f(x)\right) & \cdots & \frac{\partial}{\partial x_n}\left(\frac{\partial}{\partial x_1} f(x)\right) \\ \frac{\partial}{\partial x_1}\left(\frac{\partial}{\partial x_2} (x)\right) & \frac{\partial}{\partial x_2}\left(\frac{\partial}{\partial x_2} f(x)\right) & \cdots & \frac{\partial}{\partial x_n}\left(\frac{\partial}{\partial x_2} f(x)\right) \\ \vdots & \vdots & & \vdots \\ \frac{\partial}{\partial x_1}\left(\frac{\partial}{\partial x_n} f(x)\right) & \frac{\partial}{\partial x_2}\left(\frac{\partial}{\partial x_n} f(x)\right) & \cdots & \frac{\partial}{\partial x_n}\left(\frac{\partial}{\partial x_n} f(x)\right) \end{pmatrix}$$

[4] Als $f : U_n(c,\varepsilon) \to \mathbb{R}^m$ dan wordt de matrix J met $J_{ij} = \partial f_i / \partial x_j$ de matrix van Jacobi of de Jacobiaan van f genoemd.

of te wel

$$(Hf(x))_{ij} = \frac{\partial}{\partial x_j}\left(\frac{\partial}{\partial x_i} f(x)\right)$$

wordt de matrix van Hesse of de Hessiaan van f in x genoemd.

Onder de voorwaarden in opgave 7 geldt: $\frac{\partial^2 f}{\partial x_i \partial x_j}(x) = \frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_j}(x)\right)$.
N.B:

$$Hf(x) = \left(\left(\frac{\partial}{\partial x_1} f'(x)\right)^\top \left(\frac{\partial}{\partial x_2} f'(x)\right)^\top \ldots \left(\frac{\partial}{\partial x_n} f'(x)\right)^\top\right).$$

Voorbeeld 13 *Voor een symmetrische $n \times n$ matrix A definiëren we de kwadratische functie f op $\mathbb{R}^n$ door $f(x) = x^\top Ax$. We tonen aan dat $H(x) = 2A$ voor iedere $x \in \mathbb{R}^n$. Er geldt*

$$\begin{aligned}
\frac{\partial}{\partial x_k} f(x) = \frac{\partial}{\partial x_k}\left(\sum_i x_i \sum_j a_{ij} x_j\right) &= \sum_j a_{kj} x_j + \sum_i x_i a_{ik} \\
&= \sum_j x_j a_{jk} + \sum_i x_i a_{ik} \qquad \textit{Asym.} \\
&= 2\sum_i x_i a_{ik} \\
&= 2x^\top A.
\end{aligned}$$

Dus,

$$f'(x) = Df(x) = 2x^\top A = 2\left(\sum_{i=1}^n x_i a_{i1}, \sum_{i=1}^n x_i a_{i2}, \ldots, \sum_{i=1}^n x_i a_{in}\right),$$

zodat

$$Hf(x) = 2\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} = 2A\ (= 2A^\top).$$

voor iedere $x \in \mathbb{R}^n$.

5.3 Fermat's theorema voor meerdere variabelen

Probleem $(P_{2.1})$ Zij f een functie op $\mathbb{R}^n$: **Sect. 2.3 in B & V**

$$f(x) \to \text{ extr.}$$

noemen wij een n-dimensionaal operationaliseringsprobleem zonder restricties.

Definitie 13 *Zij f een functie, die partieel differentieerbaar is in een inwendig punt c. Wij noemen c een stationair punt van f als*

$$Df(c) = 0^\top.$$

Theorem 6 *(Theorem 2.5 p. 97 uit B &V) Zij f een reëelwaardige functie en $c \in D_f$ zó dat $f(c)$ een lokaal extremum is van f. Er geldt: als c een inwendig punt is van D_f en f differentieerbaar in c, dan is c een stationair punt van f.*

Bewijs: Neem aan dat c een inwendig punt van D_f is en dat f differentieerbaar is in $c \in D_f$. Neen een vector $u \in \mathbb{R}$ met $|u| = 1$ (een richtingsvector dus) en beschouw de functie F op een open interval $I = (-\delta, \delta)$, gedefinieerd door

$$F(t) = f(c + tu).$$

Omdat $f(c)$ een lokaal extremum is van f, is $F(0)(= f(c))$ een lokaal extremum van F. Omdat F differentieerbaar is in 0 met $F'(0) = Df(c)u$ en 0 een inwendig punt is van I, geldt volgens $F'(0) = 0$ oftewel $Df(c)u = 0$ (dit is Fermat's theorema voor een variabele). De richtingsvector u is willekeurig genomen, zodat $Df(c)u = 0$ voor iedere u. Neem achtereenvolgens $u = e_i$, $i = 1, \ldots, n$, dan geldt

$$\frac{\partial}{\partial x_i} f(c) = 0\,,\ 1 \le i \le n,$$

oftewel $Df(c) = 0^\top$. $\square$

Voorbeeld 2.3.1 uit B&V p. 99.

Voorbeeld 2.3.2 uit B&V p. 99. [**Optional**]

Theorem 7 *(de stelling van Taylor) Zij f een tweemaal continu differentieerbare, reëlwaardige functie met een open convex definitiegebied $D_f \subset \mathbb{R}^n$ en $c \in D_f$. Er geldt: voor ieder punt $x \in D_f$, $x \neq c$, bestaat er een punt ξ_x tussen c en x zó, dat*

$$f(x) = f(c) + Df(c)(x - c) + \frac{1}{2}(x - c)^\top Hf(\xi_x)(x - c).$$

(Zonder bewijs.)

Theorem 8 *(tweede orde voorwaarde voor extremum) Zij f een reëelwaardige functie die tweemaal continu differentieerbaar is op een omgeving van een stationair punt c van f (c is inwendig punt). Er geldt:*

a) als $Hf(c)$ positief definiet is, dan is $f(c)$ een lokaal minimum van f

b) als $Hf(c)$ negatief definiet is, dan is $f(c)$ een lokaal maximum van f

c) als $Hf(c)$ indefiniet is, dan is c een zadelpunt van f.

Bewijs: We volstaan met het geven van een schets van het bewijs van deel a). Neem aan dat $Hf(c)$ positief definiet is. Omdat de tweede orde partiële afgeleiden van f continu zijn, bestaat er een omgeving U van c zó, dat $Hf(c)$ positief definiet is voor iedere $x \in U$. Zij $x \in U$, $x \neq c$. Volgens de stelling van Taylor bestaat er een punt ξ_x tussen c en x zó, dat

$$f(x) = f(c) + \frac{1}{2}(x-c)^\top Hf(\xi_x)(x-c),$$

(merk op dat $Df(c) = 0$). Omdat ξ_x gelegen is tussen c in x geldt $\xi_x \in U$. De Hessian $Hf(\xi_x)$ is dus positief definiet zodat

$$(\xi_x - c)^\top Hf(\xi_x)(\xi_x - c) \geq 0.$$

Bijgevolg is dus $f(x) \geq f(c)$ voor iedere $x \in U$. De waarde $f(c)$ is dus een lokaal minimum. □

Voortzetting van voorbeeld 2.3.1 p. 99 uit B&V. Er geldt

$$Hf(x) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

is pos. definiet.

5.4 De vier stappen methode

Het theorema van Weierstrass geldt ook in $\mathbb{R}^n$, Theorem 2.6 uit B&V p. 98. Een functie op $\mathbb{R}^n$ heet *coercive* ('dwingend') (voor mimmalisering) als $\lim_{|x|\to\infty} f(x) = +\infty$. Een coercieve en continue functie f op $\mathbb{R}^n$, dan heeft f een globaal minimum.

5.5 Convexe en concave functies

Sect. B.3 uit B & V

Definitie 14 *Zij f een reëelwaardige functie met een convex definitiegebied $D_f \subset \mathbb{R}^n$. De functie wordt concaaf genoemd als voor ieder tweetal punten $a, b \in D_f$ en voor iedere $t \in [0,1]$ geldt*

$$f((1-t)a + tb) \geq (1-t)f(a) + tf(b).$$

De functie wordt convex genoemd als voor ieder tweetal punten $a, b \in D_f$ en voor iedere $t \in [0,1]$ geldt

$$f((1-t)a + tb) \leq (1-t)f(a) + tf(b).$$

Theorem 9 *Zij f een reëelwaardige functie die tweemaal continu differentieerbaar is op een convex verzameling S. Er geldt:*

a) als Hf positief definiet is op S, dan is f strikt convex op S

b) als Hf negatief definiet is, dan is f strikt concaaf op S

c) f is convex op S dan en slechts dan als Hf positief semidefiniet op S

d) f is concaaf op S dan en slechts dan als Hf negatief semidefiniet op S

(Zonder bewijs.)

Theorem 10 *Zij f een differentieerbare, reëelwaardige functie met een convex definitiegebied $D_f \subset \mathbb{R}^n$ en c een stationair punt van f. Er geldt:*

a) als f convex is, dan is $f(c)$ het (globale) minimum van f

b) als f concaaf is, dan is $f(c)$ het (globale) maximum van f.

(Zonder bewijs.)

5.6 Geometrische interpretatie van de gradiënt

Analytische geometrie

Voor $a, b \in \mathbb{R}^3$ is het kruijsproduct van a en b gedefinieerd als volgt:

$$a \times b = (a_2\, b_3 - a_3\, b_2, a_3\, b_1 - a_1\, b_3, a_1\, b_2 - a_2\, b_1).$$

Geometrische interpretatie: de vector $a \times b$ staat loodrecht op de door a en b opgespannen hypervlakte. Er geldt $|a \times b| = |a|\,|b| \sin(\eta)$. Zie Sectie 10.3 in R. Adams, p. 553.

Voorbeeld: Gegeven zijn de vectoren $a = (1, 2, 3)^\top$ en $b = (4, 5, 6)^\top$. Dan geldt

$$(a \times b)^\top = (1, 2, 3) \times (4, 5, 6) = (2\times6-3\times5, 3\times4-1\times6, 1\times5-2\times4) = (-3, 6, -3).$$

Zijn P_1, P_2 en P_3 drie punten op een oppervlakte Q. Dan is Q de verzameling van alle punten

$$P_1 + s(P_2 - P_1) + t(P_3 - P_1), \qquad (3)$$

voor $s, t \in \mathbb{R}$. Alternatief kan Q geschreven worden als de verzameling van alle punten x voor wie geldt dat

$$Nx = NP_1 \quad (\Leftrightarrow N(x - P_1) = 0), \qquad (4)$$

waarbij

$$N = (P_2 - P_1) \times (P_3 - P_1).$$

Met $N = (N_x, N_y, N_z)$ en $P_1 = (P_{1x}, P_{1y}, P_{1z})$ geld dat $(x, y, z)^\top$ in Q ligt als

$$N_x x + N_y y + N_z z = N_x P_{1x} + N_y P_{1y} + N_z P_{1z}.$$

De representatie in (3) heet de *parameter*representatie en de representatie in (4) heet de *normalen*representatie.

Voorbeeld: Zijn $P_1 = (0,0,1)^\top$, $P_2 = (2,0,0)^\top$ en $P_3 = (0,3,0)^\top$. Dan geldt $P_2 - P_1 = (2,0,-1)^\top$ en $P_3 - P_1 = (0,3,-1)^\top$, verder is

$$N = (P_2 - P_1) \times (P_3 - P_1) = (3,2,6)^\top, |N| = 7.$$

Dus de door $(P_3 - P_1)$ en $(P_3 - P_1)$ opgespannen oppervlakte door het punt P_1 (de unieke oppervlakte welke door P_1, P_2 en P_3 beschrijven wordt) is de verzameling van $(x,y,z)^\top \in \mathbb{R}^3$ zodat

$$3x + 2y + 6z = 6.$$

Tangenten hypervlakte

Beschouw de functie $F(x,y) : \mathbb{R}^2 \to \mathbb{R}$ in het punt (x_0, y_0). De punten $x, y, z)$ in de tangentenhypervlakte worden gegeven door de afbeelding $FL : \mathbb{R}^2 \to \mathbb{R}$, met $FL(x,y) = z$:

$$FL_{(x_0,y_0)}(x,y) = F(x_0,y_0) + \nabla F(x_0,y_0)(x - x_0, y - y_0)^\top. \tag{5}$$

De tangentenhypervalkte is de verzameling van alle punten $(x, y, FL(x,))^\top$ voor $(x,y)^\top \in \mathbb{R}^2$.

Lemma 1 *De normale N van de tangentenhypervlakte in $(x_0,y_0)^\top$ is geven door*

$$N = (F_x(x_0,y_0), F_y(x_0,y_0), -1) = (\nabla F(x_0,y_0), -1).$$

Bewijs: Gegeven x en y, luidt de vergelijking voor de z-coördinate van een punt op FL:

$$z = F(x_0,y_0) + \nabla F(x_0,y_0)(x - x_0, y - y_0)^\top$$

oftewel

$$\begin{aligned}
& z = F(x_0,y_0) + F_x(x_0,y_0)(x - x_0) + F_y(x_0,y_0)(y - y_0) \\
\Leftrightarrow\quad & z - F_x(x_0,y_0)x - F_y(x_0,y_0)y = F(x_0,y_0) - F_x(x_0,y_0)x_0 - F_y(x_0,y_0)y_0 \\
\Leftrightarrow\quad & F_x(x_0,y_0)x + F_y(x_0,y_0)y - z = -F(x_0,y_0) + F_x(x_0,y_0)x_0 + F_y(x_0,y_0)y_0 \\
\Leftrightarrow\quad & F_x(x_0,y_0)x + F_y(x_0,y_0)y - z = F_x(x_0,y_0)x_0 + F_y(x_0,y_0)y_0 - F(x_0,y_0)
\end{aligned}$$

We herschrijven de vergelijking als volgt

$$N(x,y,z)^\top = N(x_0, y_0, F(x_0,y_0))^\top.$$

Het punt $(x_0, y_0, F(x_0,y_0))^\top$ ligt op de tangentiaaloppervalkte en bovenstaande vergelijking is de normalenrepresentatie van de tangentenhypervlakte. □

Het bovenstaande lemma kunnen we als volgt in worden vatten: de gradiënt is de projectie van de normale van de tangentenhypervlakte in de coördinaten hypervlakte. (De gradiënt staat loodrecht (**orthogonaal**) op de projectie van de tangentenvlakte in de coördinatenstelsel.)

Voorbeeld 14 *(Geometrische interpretatie van de gradiënt) Zij* $F : \mathbb{R}^2 \to \mathbb{R}$ *met* $F(x,y) = xy$*. Dan is de gradiënt van* F *gegeven door*

$$\nabla F(x,y) = (y,x).$$

De tangentenhypervalkte in het punt (x_0,y_0) *is dan de verzameling van allen punten* $(x,y,z)^\top \in \mathbb{R}^3$ *zodat*

$$(y_0,x_0,-1)^\top(x-x_0,y-y_0,z-x_0y_0)^\top = 0 \qquad \textit{zie (4)}$$

oftewel

$$z = xy_0 + yx_0 - x_0y_0.$$

Het tangentenvlakte in het punt $(\frac{1}{2},2)$ *is dus*

$$z = 2x + \frac{y}{2} - 1 \qquad \textit{zie (5).}$$

Alternatief kunnen we in de coördinatenvlakte de niveaukromme berekenen. De tangent van de niveaukromme $xy = 1$ *in het punt* $(\frac{1}{2},2)$ *is gegeven door*

$$\nabla F(\frac{1}{2},2)\left(x-\frac{1}{2},y-2\right)^\top = \left(2,\frac{1}{2}\right)\left(x-\frac{1}{2},y-2\right)^\top$$

oftewel

$$2x - 1 + \frac{1}{2}y - 1 \Leftrightarrow y = -4x + 4.$$

Zie ook:

http://www-math.mit.edu/18.013A/HTML/chapter06/section06.html

Niveaukrommen

Zij F gedefinieerd op $\mathbb{R}^2$. De niveaukromme $F(x,y) = c$ zij differentieerbaar op $\mathbb{R}^2$. De helling van de niveaukromme in het punt (x,y) is gegeven door

$$F(x,y) = c \Rightarrow y' = -\frac{\frac{\partial}{\partial x}F(x,y)}{\frac{\partial}{\partial y}F(x,y)}.$$

De helling van de tangent in het punt (x_0,y_0) is dan

$$-\frac{\frac{\partial}{\partial x}F(x_0,y_0)}{\frac{\partial}{\partial y}F(x_0,y_0)}$$

en de vergelijking voor de tangent is

$$y - y_0 = -\frac{\frac{\partial}{\partial x}F(x_0,y_0)}{\frac{\partial}{\partial y}F(x_0,y_0)}(x-x_0)$$

oftewel

$$y = -\frac{\frac{\partial}{\partial x}F(x_0, y_0)}{\frac{\partial}{\partial y}F(x_0, y_0)}(x - x_0) + y_0$$

(dus (x, y) is een punt op de tangent als (x, y) aan de bovenstaande vergelijking voldoet). Dit kan worden herschrijven als

$$\frac{\partial}{\partial x}F(x_0, y_0)(x - x_0) + \frac{\partial}{\partial y}F(x_0, y_0)(y - y_0) = 0$$

oftewel

$$\left(\frac{\partial}{\partial x}F(x_0, y_0), \frac{\partial}{\partial y}F(x_0, y_0)\right)(x - x_0, y - y_0)^\top = 0 \Leftrightarrow \nabla F(x_0, y_0)(x - x_0, y - y_0)^\top = 0.$$

De gradiënt staat dus loodrecht (**orthogonaal**) op de tangent in de coördinatenvlakte.

Zij $x^0 = (x_1^0, \ldots, x_n^0)$ een punt op het niveau-oppervlak $F(x) = C$ voor $F : \mathbb{R}^n \to \mathbb{R}$. Men noemt de verzameling van alle $x = (x_1, \ldots, x_n)$ zó dat

$$\frac{\partial}{\partial x_1}F(x^0)(x_1 - x_1^0) + \ldots + \frac{\partial}{\partial x_1}F(x^0)(x_n - x_n^0) = 0 \Leftrightarrow F'(x^0)(x - x^0)^\top = 0$$

de tangenten hyperruimte van het niveau-oppervlak $F(x) = C$ in het punt x^0.

6 Practicum Week 3

Opgave 1 De kwadratische functie f op $\mathbb{R}^3$ is gedefinieerd door $f(x_1, x_2, x_3) = 4x_1^2 + 3x_1x_3 - 2x_2^2 - 5x_2x_3 + 4x_2 - 2x_3 + 28$.

a) Bepaal de matrix van Hesse van f.

b) Bepaal de symmetrische 3×3 matrix A, de vector $b \in \mathbb{R}^3$ en het getal c zó, dat $f(x) = x^\top Ax + b^\top x + c$, waarbij $x = (x_1, x_2, x_3)$.

c) Laat door berekening zien dat $Hf(x) = 2A$.

Opgave 2 (Exercise 2.6.1 uit B&V p. 128)

$$x_1x_2 + 50x_1^{-1} + 20x_2^{-1} \to \text{min}, \; x_1, x_2 > 0$$

Opgave 3 (Problem 2.2.1 uit B&V p.93) Bepaal de afgeleide van $f(x) = x^\top Ax + 2b \cdot x + c$.

Opgave 4 (Exercise 2.6.11 uit B&V p.132)

$$\int_{-1}^{1} (t^2 - x_2t - x_1)^2 dt \to \text{min}$$

Opgave 5 Zij f een functie gedefinieerd op $\mathbb{R}^n$ met waarden in $\mathbb{R}$ en c een inwendig punt van het definitiegebied van f zó dat f tweemaal continu differentieerbaar is op een omgeving U van c. Laat zien dat

$$\left.\frac{\partial^2 f}{\partial x_i \partial x_j}\right|_{x=c} (x) = \left.\frac{\partial^2 f}{\partial x_j \partial x_i}\right|_{x=c} (x).$$

Opgave 6 Zijn f en g gedefinieerd op een convex gebied $S \subset \mathbb{R}^n$. Laat zien dat

(a) als f en g concaaf zijn op S en $a \geq 0$, $b \geq 0$, dan is $af + bg$ concaaf op S

(b) als f en g convex zijn op S en $a \geq 0$, $b \geq 0$, dan is $af + bg$ convex op S

Opgave 7 Laat zien dat de Cobb-Douglas functie $f(x, y) = x^a y^b$ voor $0 \leq a, b$, $a + b \leq 1$ gedefinieerd op $S = \{(x, y) : x > 0, y > 0\}$ concaaf op S is.

Opgave 8 Zijn f gedefinieerd op een convex gebied $S \subset \mathbb{R}^n$ en zij F gedefinieerd op een interval in $\mathbb{R}$ dat de verzameling $\{f(x) : x \in S$ omvat. Laat zien dat

(a) als f concaaf op S is en F concaaf en monotoon stijgend op S is, dan is $F(f(x))$ concaaf op S

(b) als f convex op S is en F convex en monotoon stijgend op S is, dan is $F(f(x))$ convex op S

(c) als f concaaf op S is en F convex en monotoon dalend op S is, dan is $F(f(x))$ convex op S

(d) als f convex op S is en F concaaf en monotoon dalend op S is, dan is $F(f(x))$ concaaf op S

Opgave 9 Bepaal de tangentvlakte van $f(x, y) = \sin(xy)$ in het punt $x_0 = \sqrt{\pi} = y_0$.

7 De Lagrangemethode

Probleem $(P_{3.1})$ Zij $f : D \to \mathbb{R}$, met $D \subset \mathbb{R}^n$: **Def. 3.1, p. 140**

$$f_0(x) \to \text{ extr}, \quad f_i(x) = 0,\ 1 \le i \le m,$$

noemen wij een n-dimensionaal optimaliseringsprobleem met nevenvoorwaarden in de vorm van een stelsel bestaande uit meer dan één vergelijking.

Voorbeeld 15 *Voorbeeld 3.2.1 p. 140 uit B&V.*

$$f_0(x) = x_1 x_2 \to \max, \quad a > 0$$
$$f_1(x) = x_1^2 + x_2^2 - a^2 = 0, \quad x_i > 0 \textit{ voor } i = 1, 2.$$

Let op dat wij $x_i = 0$ kunnen toelaten zonder de oplossing te beïnvloeden. $x_i = 0$ levert immers $f_0(x) = 0$ op en dit is zeker geen maximumlocatie.

De verzameling $V = \{x_1, x_2 \ge 0 : x_1^2 + x_2^2 - a^2 = 0\}$ is een kwart cirkel met radius a en is dus compact. Met de stelling van Weierstrass volgt dan dat er een maximum bestaat en de randpunten en de stationaire punten kandidaat maxima zijn.

Uitwerking via Fermat:

$$x_1 = \sqrt{a^2 - x_2^2}$$

en wij kunnen f_0 herschrijven als

$$f_0(x_2) = x_2\sqrt{a^2 - x_2^2}.$$

Wij vereenvoudigen de notatie door $x_2 = x$ te schrijven. Differentiëren geeft

$$\begin{aligned} f_0'(x) &= \sqrt{a^2 - x^2} + \frac{1}{2}\frac{(-2x)x}{\sqrt{a^2 - x^2}} \\ &= \sqrt{a^2 - x^2} - \frac{x^2}{\sqrt{a^2 - x^2}}. \end{aligned}$$

Bepalen van de stationaire punten:

$$\begin{aligned} f_0'(x) = 0 \Rightarrow \sqrt{a^2 - x^2} &= \frac{x^2}{\sqrt{a^2 - x^2}} \\ a^2 - x^2 &= x^2 \\ a^2 &= 2x^2 \\ x = \frac{a}{\sqrt{2}}. \end{aligned}$$

Dus $x_2 = \frac{a}{\sqrt{2}}$ en invullen in f_1 geeft $x_1 = \frac{a}{\sqrt{2}}$ en

$$f_0\left(\frac{a}{\sqrt{2}}, \frac{a}{\sqrt{2}}\right) = \frac{a^2}{2}.$$

De randpunten $(a,0)$ en $(0,a)$ geven $f_0 = 0$. Dus

$$\left(\frac{a}{\sqrt{2}}, \frac{a}{\sqrt{2}}\right)$$

is maximumlocatie.

7.1 De Lagrangemethode

Wij voeren op de productruimte $D \times \mathbb{R}^{m+1}$ de Lagrangefunctie $\mathcal{L}$ in, gedefinieerd door

$$\mathcal{L}(x,\lambda) = \sum_{i=0}^{m} \lambda_i f_i(x).$$

De rijvector λ wordt de Lagrange multiplicator genoemd. Wij schrijven $\mathcal{L}_x(y,\lambda)$ voor de rijvector van partiële afgeleiden naar x in y:

$$\mathcal{L}_x(y,\lambda) = \left(\frac{\partial \mathcal{L}}{\partial x_1}(y,\lambda), \ldots, \frac{\partial \mathcal{L}}{\partial x_n}(y,\lambda)\right).$$

Theorem 11 *(Theorem 3.3 p. 143 uit B&V.) Zij f_0 differentieerbaar in $y \in D$ en f_i, $1 \leq i \leq m$, continu differentieerbaar in y. Zij y een inwendig punt van D zó, dat $f(y)$ een lokaal extremum is van f op de verzameling $T = \{x \in D : f_i(x) = 0, 1 \leq i \leq m\}$. Er geldt: er bestaat een Lagrange multiplicator $\lambda \neq 0_{m+1}$ zó, dat y een stationair punt is van $\mathcal{L}(x,\lambda)$, oftewel*

$$\mathcal{L}_x(y,\lambda) = 0_n^\top$$
$$\Leftrightarrow \sum_{i=0}^{m} \lambda_i f_i'(y) = 0_n^\top = \sum_{i=0}^{m} \lambda_i \left(\frac{\partial f_i}{\partial x_1}(y), \ldots, \frac{\partial f_i}{\partial x_n}(y)\right)$$
$$\Leftrightarrow \sum_{i=0}^{m} \lambda_i \frac{\partial f_i}{\partial x_j}(y) = 0\,, 1 \leq j \leq n.$$

N.B.: de vectoren $f_i'(y)$, $0 \leq i \leq m$, zijn lineair afhankelijk.

In woorden: een lokaal extremum van het optimaliseeringsprobleem is ook altijd een stationair punt van de Lagrangefunctie. Het is echter niet altijd zo dat een lokaal extremum van het optimaliseeringsprobleem ook een lokaal extremum van de Lagrangefunctie is.

Voorbeeld 16 *Voorbeeld 3.2.3 p. 144 uit B&V (voortzetting van voorbeeld 15)*
Weierstrass is al gedaan (minimum lokaties zijn de randpunten, we zoeken maximum).

$$\mathcal{L}(x,\lambda) = \lambda_0 f_0(x) + \lambda_1 f_1(x) = \lambda_0 x_1 x_2 + \lambda_1 (x_1^2 + x_2^2 - a^2).$$

$$\begin{aligned} \mathcal{L}_x(x,\lambda) = 0_2^\top \Leftrightarrow \frac{\partial}{\partial x_1}\mathcal{L}(x,\lambda) &= \lambda_0 x_2 + 2\lambda_1 x_1 = 0 \\ \frac{\partial}{\partial x_2}\mathcal{L}(x,\lambda) &= \lambda_0 x_1 + 2\lambda_1 x_2 = 0 \end{aligned}$$

2 (a) Zij $\lambda_0 = 0$, dan $2\lambda_1 x_1 = 2\lambda_1 x_2$. Omdat $\lambda_1 \neq 0$ geeft dit $x_1 = 0 = x_2$ en dat is tegenstrijdig met f_1.

2 (b) Zij $\lambda_0 = 1$, dan

$$\left.\begin{array}{ll} x_2 + 2\lambda_1 x_1 & = 0 \\ x_1 + 2\lambda_1 x_2 & = 0 \end{array}\right\} \Rightarrow \begin{array}{ll} 2\lambda_1 & = -\frac{x_2}{x_1} \\ 2\lambda_1 & = -\frac{x_1}{x_2} \end{array}$$

Dus

$$\frac{x_2}{x_1} = \frac{x_1}{x_2} \Leftrightarrow x_1^2 = x_2^2.$$

Met f_1 volgt dan $x_1 = x_2 = \frac{a}{\sqrt{2}}$.

Voordeel van Lagrange in vergelijk met Fermat: (i) algebra is makkelijker (no 'roots'), (ii) $n + m$ vergelijkingen in precies zo veel onbekenden waarvan $\lambda_1, \ldots, \lambda_m$ niet hoeven berekend te worden.

Voorbeeld 3.2.6 p. 147 uit B&V. ($\lambda_0 = 0$)

$$f_0(x) \to \min, \quad x_1^2 - x_2^3 = 0$$

en $\hat{x}_0 = 0_2$ voor $f_0 : \mathbb{R}^2 \to \mathbb{R}$ differentieerbaar in het punt 0_2. Wij veronderstellen verder dat $\frac{\partial}{\partial x_1} f_0 = \frac{\partial}{\partial x_2} f_0$.

Uitwerking: $f_0(x) \to \min$ onder $f_1(x) = x_1^2 - x_2^3 = 0$.

1) (Zonder Weierstrass)

2)

$$\begin{aligned} \mathcal{L}(x) &= \lambda_0 f_0(x) + \lambda_1(x_1^2 - x_2^3) \\ \frac{\partial}{\partial x_1}\mathcal{L}(x) &= \lambda_0 \frac{\partial}{\partial x_1} f_0(x) + 2\lambda_1 x_1 \\ \frac{\partial}{\partial x_2}\mathcal{L}(x) &= \lambda_0 \frac{\partial}{\partial x_2} f_0(x) + 3\lambda_1 x_2^2 \end{aligned}$$

2 (a) Zij $\lambda_0 = 0$, dan $\lambda_1 \neq 0$ en

$$2\lambda_1 x_1 = 3\lambda_1 x_2^2 \Leftrightarrow x_1 = \frac{3}{2} x_2^2$$

invullen in f_1 geeft

$$\frac{9}{4} x_2^4 - x_2^3 = \left(\frac{9}{4} x_2 - 1\right) x_2^3 = 0 \Leftrightarrow x_1 = x_2 = 0 \text{ of } x_2 = \frac{4}{9}, x_1 = \frac{8}{27}.$$

2 (b) Zij $\lambda_0 = 1$, dan

$$\frac{\partial}{\partial x_1} f_0(x) + 2\lambda_1 x_1 = 0 \text{ en } \frac{\partial}{\partial x_2} f_0(x) + 3\lambda_1 x_2 = 0$$

Omdat $\frac{\partial}{\partial x_1} f_0 = \frac{\partial}{\partial x_2} f_0$ volgt hieruit

$$2x_1 = 3x_2^2.$$

Dit leidt tot de zelfde oplossing als 2(a).

7.2 Werken met de Lagrangemethode

Voorbeeld 3.2.8 p. 149 uit B&V. Laat zien dat $C, D \geq 0$ bestaan zó dat voor ieder x_1, x_2

$$C\sqrt[4]{x_1^4 + x_2^4} \leq \sqrt{x_1^2 + x_2^2} \leq C\sqrt[4]{x_1^4 + x_2^4}.$$

Uitwerking: Herschrijf

$$C\,\sqrt[4]{x_1^4 + x_2^4} \leq \sqrt{x_1^2 + x_2^2} \leq D\,\sqrt[4]{x_1^4 + x_2^4}$$

als

$$C \leq \frac{\sqrt{x_1^2 + x_2^2}}{\sqrt[4]{x_1^4 + x_2^4}} \leq D.$$

Zij $y = \alpha x$, dan is

$$\frac{\sqrt{y_1^2 + y_2^2}}{\sqrt[4]{y_1^4 + y_2^4}} = \frac{\sqrt{x_1^2 + x_2^2}}{\sqrt[4]{x_1^4 + x_2^4}}$$

Wij kunnen dus volstaan met $x_1^4 + x_2^4 = 1$ (de ‘lengte’ van de x-vector is niet belangrijk). Wij lossen op

$$\begin{aligned} \tilde{f}_0(x) &= \sqrt{x_1^2 + x_2^2} \rightarrow \text{extr.} \\ & x_1^4 + x_2^4 - 1 = 0 \end{aligned}$$

of

$$\begin{aligned} f_0(x) &= x_1^2 + x_2^2 \rightarrow \text{extr.} \\ & x_1^4 + x_2^4 - 1 = 0 \end{aligned}$$

1) (Weierstrass) f_0 is continu en $V = \{x : x_1^4 + x_2^4 - 1 = 0\}$ is compact:

- afgesloten: $V = g^{-1}(\{0\})$ met $g(x) = x_1^4 + x_2^4 - 1$ continu
- niet leeg: $(1, 0) \in V$
- begrensd: voor ieder $x \in V$ geldt $-1 \leq x_1, x_2 \leq 1$

2)

$$\begin{aligned} \mathcal{L}(x, \lambda) &= \lambda_0(x_1^2 + x_2^2) + \lambda_1(x_1^4 + x_2^4 - 1) \\ \frac{\partial}{\partial x_1}\mathcal{L}(x) &= 2\lambda_0 x_1 + 4\lambda_1 x_1^3 \\ \frac{\partial}{\partial x_2}\mathcal{L}(x) &= 2\lambda_0 x_2 + 4\lambda_1 x_2^3 \end{aligned}$$

2 (a) Zij $\lambda_0 = 0$, dan $\lambda_1 \neq 0$ en $4\lambda_1 x_1^3 = 0 = 4\lambda_1 x_2^3$. Dus $x_1 = 0 = x_2$ en dat is in tegenspraak met $x_1^4 + x_2^4 = 1$.

2 (b) $\lambda_0 = 1$, dan

$$\begin{aligned} 2x_1 + 4\lambda_1 x_1^3 = 0 &\quad \Leftrightarrow 4\lambda_1 = -2\tfrac{x_1}{x_1^3} \\ 2x_2 + 4\lambda_1 x_2^3 = 0 &\quad \Leftrightarrow 4\lambda_1 = -2\tfrac{x_2}{x_2^3} \\ \Rightarrow &\quad 2x_1 x_2^3 = 2x_2 x_1^3 \end{aligned}$$

Mogelijke oplossingen:

- $x_1 = 0$ en $x_2 = \pm 1$
- $x_1 = \pm 1$ en $x_2 = 0$
- zij $x_1 \neq 0$ en $x_2 \neq 0$ (en wij mogen door x_1, x_2 delen) dan

  $$2x_1 x_2^3 = 2x_2 x_1^3 \Rightarrow x_1^2 = x_2^2 \Rightarrow |x_1| = |x_2| =: C$$

  Uit de nevenvoorwaarde volgt

  $$C^4 + C^4 = 1 \Leftrightarrow C = \sqrt[4]{\frac{1}{2}}$$

Uit de bovenstaande analyse volgt voor de extrema $\hat{x}$

$$\hat{x} \in \{(0,1), (1,0), (-1,0), (0,-1), (C,C), (-C,C), (C,-C), (-C,-C)\}$$

De waarden van f_0 voor $\hat{x} \in \{(0,1), (1,0), (-1,0), (0,-1)\}$ zijn gelijk aan 1 en de andere waarden zijn gelijk aan $\sqrt{2} > 1$. Dus de punten $\{(C,C), (-C,C), (C,-C), (-C,-C)\}$ zijn maximumlocaties en de punten $\{(0,1), (1,0), (-1,0), (0,-1)\}$ zijn minimumlocaties. Wij hebben dus aangetoond:

$$1 \leq x_1^2 + x_2^2 \leq \sqrt{2}, \qquad \text{voor } x_1^4 + x_2^4 = 1$$

oftewel

$$1 \leq \sqrt{x_1^2 + x_2^2} \leq \sqrt[4]{2}, \qquad \text{voor } x_1^4 + x_2^4 = 1$$

en we hebben

$$1 \leq \frac{\sqrt{x_1^2 + x_2^2}}{\sqrt{x_1^4 + x_2^4}} \leq \sqrt[4]{2}, \qquad \text{voor ieder } x$$

(wij hebben boven al aangetoond dat we met $x_1^4 + x_2^4 = 1$ kunnen volstaan, de 'lengte' van de x-vector is niet belangrijk). Er geldt dus $C = 1$ en $D = \sqrt[4]{2}$.

7.3 Bewijs van de Lagrangemethode

Theorem 12 *(Tangent Space Theorem, zie ook Theorem 2.13 p. 127 uit B & V) Zij $\hat{x} \in \mathbb{R}^n$, $U \subset \mathbb{R}^n$ een open omgeving van $\hat{x}$, $f_i : U \to \mathbb{R}$, $1 \leq i \leq m$, en*

$$F(x_1, \ldots, x_n) = \begin{pmatrix} f_1(x_1, \ldots, x_n) \\ f_2(x_1, \ldots, x_n) \\ \vdots \\ f_m(x_1, \ldots, x_n) \end{pmatrix}$$

een vector afbeelding van U naar $\mathbb{R}^m$ zó dat F continu differentieerbaar is in $\hat{x}$, rank$F'(\hat{x}) = m$ en $F(\hat{x}) = 0_m$.

Zij $F'(\hat{x})v = 0$, dan bestaat een afbeelding $r : \mathbb{R}^n \to \mathbb{R}^n$ zó dat voor (een wat kleinere) omgeving U geldt dat

$$F\big(tv + r(tv)\big) = 0_m$$

voor v orthogonaal op de tangenten hyperruimte in $\hat{x}$ [5]*, $tv \in U$ en $r(tv) = o(t)$.*

Het Tangent Space Theorem laat zien dat de verzameling $T = \{x \in D : f_i(x) = 0, 1 \le i \le m\}$ lokaal in een omgeving van 0_n door een lineaire functie benadert worden kan.

Bewijs van Lagrange: De nevenvoorwaarden geven een functie

$$F(x_1, \ldots, x_n) = \begin{pmatrix} f_1(x_1, \ldots, x_n) \\ f_2(x_1, \ldots, x_n) \\ \vdots \\ f_m(x_1, \ldots, x_n) \end{pmatrix}.$$

Wij veronderstellen dat $\hat{x} = 0_n$ de lokatie van het globaal minimum is en dat $F(0_n) = 0_m$. Merk op

$$F'(x_1, \ldots, x_n) = \begin{pmatrix} f_1'(x_1, \ldots, x_n) \\ f_2'(x_1, \ldots, x_n) \\ \vdots \\ f_m'(x_1, \ldots, x_n) \end{pmatrix} = \begin{pmatrix} \frac{\partial}{\partial x_1} f_1(x_1, \ldots, x_n) & \ldots & \frac{\partial}{\partial x_n} f_1(x_1, \ldots, x_n) \\ \frac{\partial}{\partial x_1} f_2(x_1, \ldots, x_n) & \ldots & \frac{\partial}{\partial x_n} f_2(x_1, \ldots, x_n) \\ \vdots & \vdots & \vdots \\ \frac{\partial}{\partial x_1} f_m(x_1, \ldots, x_n) & \ldots & \frac{\partial}{\partial x_n} f_m(x_1, \ldots, x_n) \end{pmatrix}.$$

Het niet regulair geval: De rijen van $F'(0_n^\top)$ zijn lineair afhankelijk. Dan bestaat een combinatie λ_i, $1 \le i \le m$, zó dat

$$\sum_{i=1}^m \lambda_i f_i'(0_n^\top) = 0_n^\top.$$

Als wij $\lambda_0 = 0$ kiezen, geldt dan $\sum_{i=1}^m \lambda_i f_i'(0_n^\top) = 0_n^\top$. (Wij vinden dus een oplossing zonder de voorwaarde $f_0'(0_n^\top) = 0_n^\top$).

Het regulair geval: De rijen van $F'(0_n)$ zijn lineair onafhankelijk. Zij $v \in \mathbb{R}^n$ zó dat $F'(0_n)v = 0_m$. Met het Tangent-Spade-Theorem volgt dan dat

$$tv + r(tv)$$

een oplossing is van $F(x) = 0_m$ voor $|t|$ voldoende klein, en $r(tv) = o(t)$. Wij hebben verondersteld dat 0_n een minimumlokatie is dus

$$0 \le f_0(tv + r(tv)) - f_0(0_n)$$

[5]andere verwoordingen: $v \in \text{kern}F'(\hat{x})$, of $v \perp y$ voor ieder y in de door $F'(\hat{x})$ opgespannen ruimte

en uit de differentieerbaarheid van f_0 volgt

$$0 \leq f_0(tv + r(tv)) - f_0(0_n) = tf_0'(0_n)v + o(t).$$

Omdat t positieve en negatieve waarden kan aannemen volgt dat $f_0'(0_n)v = 0$. Omdat $F'(0_n)v = 0_m$ volgt uit de theorie der lineaire vergelijkingen dat f_0' een lineair combinatie van de rijen van $F'(0_n)$ is. Omdat de rijen van F' de vectoren $f_i'(0_n)$, $1 \leq i \leq m$, zijn volgt het bewijs van de stelling van Lagrange.

8 Practicum Week 4

Opgave 1 (Probleem 3.3.7 uit B&V p.163) Ieder symmetrische matrix $A \in \mathbb{R}^{n\times n}$ heeft een eigenwaarde in $\mathbb{R}$.

Opgave 2 (Exercise 3.6.1 uit B&V p.190)

$$e^{x_1x_2} \to \max,\ x_1^3 + x_2^3 = 1$$

Opgave 3 (Exercise 3.6.2 uit B&V p.190)

$$x_1^2 + 12x_1x_2 + 2x_2^2 \to \text{extr},\ 4x_1^2 + x_2^2 = 25$$

Opgave 4 (Exercise 3.6.3 uit B&V p.190)

$$x_1x_2^2x_3^3 \to \text{extr},\ x_1^2 + x_2^2 + x_3^2 = 1$$

Opgave 5 (Exercise 3.6.13 uit B&V p.195) (!) Laat zien dat de Lagrangemethode de verkeerde oplossing geeft (namelijk $(x_1, x_2) = (9, 4)$) voor

$$2x_1 + 3x_2 \to \max,\ \sqrt{x_1} + \sqrt{x_2} = 5$$

9 Kuhn-Tucker

Probleem $(P_{4.1})$ Zij $f_0 : D \to \mathbb{R}$, met $D \subset \mathbb{R}^n$ een convexe verzameling: **Def. 4.1 in p.206**

$$f_0(x) \to \text{ extr}, \quad f_i(x) \le 0,\ 1 \le i \le m,$$

noemen wij een n-dimensionaal optimaliseringsprobleem met nevenvoorwaarden in vorm van een stelsel bestaande uit meer dan één ongelijkheid.

Voorbeeld 4.2.2 uit B&V p.206:

$$f_0(x) = (x_1 - 2)^2 + (x_2 - 3)^2 \to \min$$
$$f_1(x) + x_1 + x_2 - 2 \le 0, f_2(x) = x_1^2 - 4 \le 0$$

Uitwerking volgt later.

9.1 Het Karush-Kuhn-Tucker Theorem

Wij voeren op de productruimte $D \times \mathbb{R}^{m+1}$ de Lagrangefunctie $\mathcal{L}$ in, gedefinieerd door

$$\mathcal{L}(x, \lambda) = \sum_{i=0}^{m} \lambda_i f_i(x).$$

De rijvector λ wordt de Lagrange multiplicator genoemd. Wij veronderstellen dat de functies f_i, $0 \le i \le m$, differentieerbaar zijn in $\hat{x}$. Wij schrijven $\mathcal{L}_x(y, \lambda)$ voor de rijvector van partiële afgeleiden naar x in y:

$$\mathcal{L}_x(y, \lambda) = \left(\frac{\partial \mathcal{L}}{\partial x_1}(y, \lambda), \ldots, \frac{\partial \mathcal{L}}{\partial x_n}(y, \lambda) \right).$$

De Karush-Kuhn-Tucker voorwaarden (KKT-voorwaarde) in punt $\hat{x}$ zijn:

(α) $\mathcal{L}_x(\hat{x}, \lambda) = 0_n^\top$ stationariteit;

(β) $\lambda_i \ge 0$, $1 \le i \le m$, niet-negativiteit;

(γ) $\lambda_i f_i(\hat{x}) = 0$, $1 \le i \le m$, complementair slackness

Theorem 13 *(Theorem 4.4 p. 210 uit B&V.) Zijn $f_i \le 0$, $0 \le i \le m$, differentieerbaar in $\hat{x} \in \mathbb{R}^n$ en convex. Er geldt: als $\hat{x}$ een minimumlokatie is dan bestaat er een Lagrange multiplicator $\lambda \ne 0_{m+1}$ zó, dat $\hat{x}$ aan de KKT voorwaarden voldoet.*

N.B. Convexiteit levert dat wij geen onderscheid moeten maken tussen globaal en lokaal minimum.

Definitie 15 *Een punt $x \in \mathbb{R}^n$ wordt Slater punt genoemd als $f_i(x) < 0$, $1 \le i \le m$.*

Theorem 4.5 p.211 uit B&V: *Zijn $f_i \le 0$, $0 \le i \le m$, differentieerbaar in $\hat{x} \in \mathbb{R}^n$ en convex.*

(i) Als $\hat{x}$ aan de KKT voorwaarden voldoet voor λ met $\lambda_0 = 1$, dan is $\hat{x}$ een minimumlokatie van f_0 (het globaal minimum van de Lagrange functie is ook het globaal minimum van de waardefunctie).

(ii) Veronderstel dat er een Slater punt $\bar{x}$ bestaat. Er geldt: iedere vector λ die aan de KKT-voorwaarden voldoet heeft $\lambda_0 \neq 0$.

Bewijs: Volgens (α) is $\hat{x}$ een stationair punt van de Lagrangefunctie. Omdat $\mathcal{L}(x, \lambda)$ convex is in x voor iedere λ, volgt verder uit (α) dat $\hat{x}$ een globaal minimum is van $\mathcal{L}(\hat{x}, \lambda)$:

$$\mathcal{L}(\hat{x}, \lambda) \leq \mathcal{L}(x, \lambda)\ . \qquad (\delta)$$

Bewijs (i): Zij x een toegestaan punt. Als wij $\lambda_0 = 1$ kiezen geeft dit

$$f_0(x) \overset{(\beta)+\mathrm{f_i}\leq 0}{\geq} \mathcal{L}(x, \lambda) \overset{(\delta)}{\geq} \mathcal{L}(\hat{x}, \lambda) \overset{(\gamma)}{=} f_0(\hat{x}).$$

Bewijs (ii): Zij $\bar{x}$ een Slater punt. Neem aan dat $\hat{x}$ aan de KKT voorwaarden voldoet voor λ en dat $\lambda_0 = 0$, dan

$$0 \overset{(\gamma)+\lambda_0=0}{=} \mathcal{L}(\hat{x}, \lambda) \overset{(\delta)}{\leq} \mathcal{L}(\bar{x}, \lambda) \overset{\lambda_0=0}{=} \sum_{i=1}^{m} \underbrace{\lambda_i}_{\geq 0} \underbrace{f_i(\bar{x})}_{\leq 0} \overset{\text{def. Slater punt}}{<} 0$$

waarbij we benutten dat uit $\lambda \neq 0_{m+1}^{\top}$ volgt dat en minstens één λ_i, $0 < i \leq m$, niet null is. Uit deze tegenspraak volgt $\lambda_0 \neq 0$. $\square$

9.2 Werken met het KKT theorema

Voorbeeld 4.2.5 uit B&V p.212:

$$f_0(x) = (x_1 - 2)^2 + (x_2 - 3)^2 \rightarrow \min$$
$$f_1(x) + x_1 + x_2 - 2 \leq 0, f_2(x) = x_1^2 - 4 \leq 0$$

Uitwerking:

$$\begin{aligned} f_0(x) &= (x_1 - 2)^2 + (x_2 - 3)^2 \rightarrow \min \\ f_1(x) &= x_1 + x_2 - 2 \leq 0 \\ f_2(x) &= x_1^2 - 4 \leq 0, \end{aligned}$$

$x \in \mathbb{R}^2$. Het probleem is convex (zo als in KKT aangegeven wij hebben dus geen Weierstrass nodig). Verder is $(0, 0)$ een Slaterpunt:

$$f_1(0, 0) = -2 < 0 \quad \text{en} \quad f_2(0, 0) = -4 \leq 0.$$

De Lagrangefunctie wordt dan

$$\begin{aligned} \mathcal{L}(x, \lambda) &= (x_1 - 2)^2 + (x_2 - 3)^2 + \lambda_1(x_1 + x_2 - 2) + \lambda_2(x_1^2 - 4) \\ \mathcal{L}_{x_1}(x, \lambda) &= 2(x_1 - 2) + \lambda_1 + 2\lambda_2 x_1 = 0 \\ \mathcal{L}_{x_2}(x, \lambda) &= 2(x_2 - 3) + \lambda_1 = 0 \end{aligned}$$

Mogelijke gevallen: ($f_1, f_2 < 0$ en $\lambda_1 = \lambda_2 = 0$), ($f_1 < 0, f_2 = 0$ en $\lambda_1 = 0, \lambda_2 \geq 0$), ($f_1 = 0, f_2 < 0$ en $\lambda_1 \geq 0, \lambda_2 = 0$) ($f_1, f_2 = 0$ en $\lambda_1, \lambda_2 \geq 0$).

Wij gaan verder met ($f_1 = 0, f_2 < 0$ en $\lambda_1 \geq 0, \lambda_2 = 0$).

$$\begin{aligned} 2(x_1 - 2) + \lambda_1 &= 0 \\ 2(x_2 - 3) + \lambda_1 &= 0 \\ x_1 + x_2 - 2 &= 0 \Leftrightarrow x_1 = 2 - x_2 \end{aligned}$$

Uit de eerste twee vergelijkingen volgt

$$2(x_1 - 2) = 2(x_2 - 3)$$

invullen van de deerde geeft

$$2(2 - x_2 - 2) = 2(x_2 - 3) \Leftrightarrow 6 = 4x_2 \Leftrightarrow x_2 = \frac{3}{2} \Rightarrow x_1 = \frac{1}{2}.$$

Het punt $(1/2, 3/2)$ is het uniek minimum de waarde van het minimum is $f_0(1/2, 3/2) = 4\frac{1}{2}$.

Probleem 4.3.1 uit B&V p.217. [**Optional** identiek met opgave 3] Voor $a \neq 0_n$ los het volgende probleem

$$ax = \sum_{i=1} x_i a_i \to \max,\ |x|^2 = \sum_{i=1}^{n} x_i^2 = 1$$

Uitwerking: Zij $a^\top \in \mathbb{R}^n$ met $a \neq 0_n^\top$.

$$ax = \sum_{i=1}^{n} a_i x_i \to \max$$

onder

$$|x|^2 = \sum_{i=1}^{n} x_i^2 \leq 1.$$

(Optimalisering op een sphere) Hiervan maken wij

$$f_0(x) = -ax = -\sum_{i=1}^{n} a_i x_i \to \min$$

onder

$$f_1(x) = |x|^2 - 1 = \sum_{i=1}^{n} x_i^2 - 1 \leq 0.$$

$$\mathcal{L}(x, \lambda) = \lambda_0(-ax) + \lambda_1(|x|^2 - 1)$$

Omdat $f_1(0_n) = -1 < 0$ is 0_n een Slaterpunt, dus $\lambda_0 = 1$. De KKT condities zijn dus

$$\begin{aligned}
\mathcal{L}_x(x, \lambda_1) &= -a + 2\lambda_1 x^\top \\
&= (-a_1 + 2\lambda_1 x_1, -a_2 + 2\lambda_1 x_2, \ldots, -a_n + 2\lambda_1 x_n) \\
&= 0_n^\top, \qquad \lambda_1 \geq 0.
\end{aligned}$$

Zij $f_1(x) < 0$, dan is $\lambda_1 = 0$ en

$$\mathcal{L}_x(x, \lambda_1) = 0_n^\top = -a$$

en dit kan allen als $a = 0_n^\top$ wat een tegenstelling met $a \neq 0_n^\top$ is.

Zij $f_1(x) = 0$, dan

$$\begin{aligned}
\mathcal{L}_x(x, \lambda_1) &= -a + 2\lambda_1 x^\top \\
&= (-a_1 + 2\lambda_1 x_1, -a_2 + 2\lambda_1 x_2, \ldots, -a_n + 2\lambda_1 x_n) \\
&= 0_n^\top, \qquad \lambda_1 \geq 0.
\end{aligned} \tag{6}$$

Dus,

$$\frac{a_i}{2\lambda_1} = x_i, \quad 1 \leq i \leq n$$

Omdat $f_1(x) = 1$ volgt hieruit

$$\sum_{i=1}^{n} \frac{a_i^2}{4\lambda_1^2} = 1 \Leftrightarrow \lambda_1 = \sqrt{\frac{\sum_{i=1}^n a_i^2}{4}} = \frac{|a|}{2}.$$

Naar invullen in (6) volgt dat $x^\top = \frac{a}{|a|}$ (= de vector a op lengte 1 gescaleerd). De waarde van het maximum(!) is

$$\frac{a^\top a}{|a|} = \frac{\sum_{i=1}^n a_i^2}{\sqrt{\sum_{i=1}^n a_i^2}} = \sqrt{\sum_{i=1}^{n} a_i^2} = |a|$$

in vector notatie

$$\frac{a^\top a}{|a|} = \frac{|a|^2}{|a|} = |a|.$$

9.3 Bewijs van het Kuhn-Tucker Theorema

Vector w is een tangente van $C \subset \mathbb{R}^n$ in $\overline{x}$, in notatie $w \in T_C(\overline{x})$, als er een reeks $w(k)$ bestaat met $w(k) \to w$ als $k \to \infty$ en een reeks $t(k) \in \mathbb{R}$ met $t(k) \searrow 0$ zo dat

$$\overline{x} + t(k)\, w(k) \in C.$$

Voorbeeld: Als $\overline{x}$ een inwendig punt van C is dan geldt $T_C(\overline{x}) = \mathbb{R}^n$, en als $\overline{x}$ een randpunt is en C convex, dan is $T_C(\overline{x})$ de verzameling van alle vectoren die vanuit $\overline{x}$ in richting C wijzen. (De hyperruitme op de "C-kant" van $\overline{x}$ gegeven door de tangentialhypervlakte in het punt $\overline{x}$.

Vector w is een normale van $C \subset \mathbb{R}^n$ in $\overline{x}$, in notatie $w \in N_C(\overline{x})$, als er een reeks $w(k)$ bestaat met $w(k) \to w$ als $k \to \infty$ en $x(k)$ met $x(k) \to \overline{x}$ zo dat

$$< w(k), x(k) - \overline{x} > \leq \ o(|x(k) - \overline{x}|)$$

De vector $v \in \mathbb{R}^n$ is een *regulaire normale* van C in $\overline{x}$ als

$$< v, x - \overline{x} > \leq \ o(|x - \overline{x}|) \quad \forall x \in C,$$

waarbij $< x, y >= x^\top y$.

Voorbeeld: Als $\overline{x}$ een inwendig punt van C is dan geldt $N_C(\overline{x}) = \{0_n\}$. Bewijs: uit

$$| < v, x - \overline{x} > | = |x - \overline{x}|\, |v| \ \cos(\eta)$$

volgt

$$\frac{| < v, x - \overline{x} > |}{|x - \overline{x}|} = |v| \ \cos(\eta)$$

en dus $< v, x - \overline{x} > \leq \ o(|x - \overline{x}|)$ voor ieder $x \in C$ als $|v| = 0$ oftewel $v = 0_n$. Als $\overline{x}$ een randpunt van C is en C convex, dan geldt $N_C(\overline{x}) = \{v\}$ waarbij v "loodrecht op C in $\overline{x}$ staat".

Zij $f : \mathbb{R}^n \to \mathbb{R}$ een differentieerbare functie op $C \subset \mathbb{R}^n$. Als $\overline{x}$ voor f in een lokaal minimum in C is, dan is $\nabla f(\overline{x})$ is een regulaire normale voor C in $\overline{x}$.

Supporting hyperplane theorem (Theorem 4.6 p.231 uit B&V) en Sectie 4.4.2 uit B&V.

Theorem 14 *(Supporting Hyperplane Theorem) Zij $C \subset \mathbb{R}^n$ een convexe verzameling en 0_n een randpunt van C. Dan bestaat er $\lambda \in (\mathbb{R}^n)^T$ zó dat $\lambda x \geq 0$ voor iedere $x \in \overline{C}$. In kort: $-\lambda \in N_{\overline{C}}(0_n)$.*

De hypervlakte λx ondersteunt ('supports') de verzameling C in 0_n: de verzameling C ligt op een kant van de hypervlakte. Bewijs van de stelling:

Zij int$(C) = \emptyset$ dan is C onderdeel van de hyperruimte $\mathcal{H} = \{x \in \mathbb{R}^n | \lambda x = 0\}$ voor een niet nul vector $\lambda \in (\mathbb{R}^n)^\top$[6].

Zij $c \in C$ en zij $\hat{x}$ een punt in $D = \text{cl}(\mathbb{R}^+ C)$[7] met minimale afstand tot het punt $-c$. Dan voldoet $\lambda = (\hat{x} + c)^\top$ aan de stelling. N.B.: Waarom bestaat een punt met minimale afstand? Voeg de nevenvoorwaarde $|x + c| \leq |c|$ aan C toe en maak de verzameling daardoor begrensd. Ook de nieuwe verzameling $\hat{C}$ is weer niet leeg (0_n is een element) en afgesloten. De stelling van Weierstrass levert dan dat er een punt met minimale afstand bestaat: $|x(-c)| \to$ min, $x \in \hat{C}$.

Bewijs van KKT: Neem aan dat $f_0(\hat{x}) = 0$. Zij

$$C = \{y \in \mathbb{R}^{m+1} \,|\, \exists x \in \mathbb{R}^n : y_i \geq f_i(x),\ 0 \leq i \leq m\}.$$

[6] λ is normale van $\mathcal{H}$, uit $\lambda^\top x > 0$ volgt $|\lambda||x| \cos(\eta) > 0$ waarbij η de hoek tussen λ en x is. Uit het feit dat $\cos(\eta) > 0$, volgt $-\pi/2 < \eta < \pi/2$.

[7] De verzameling $\mathbb{R}^+ C$ is gedefinieerd door $\mathbb{R}^+ C = \{yc \in \mathbb{R}^n : y \in \mathbb{R}^+, c \in C\}$.

Nota bene: omdat $f_0(\hat{x}) = 0$ en $f_i \leq 0$ $(1 \leq i \leq m)$ volgt dat $0_{m+1} \in C$.
De verzameling C is convex: zijn $x, y \in C$ dan bestaan x' en y' zó dat

$$x \geq (f_0(x'), \ldots, f_m(x')) \quad \text{en} \quad y \geq (f_0(y'), \ldots, f_m(y')).$$

Er geldt

$$\begin{aligned} \alpha x + (1-\alpha)y &\geq \alpha(f_0(x'), \ldots, f_m(x')) + (1-\alpha)(f_0(y'), \ldots, f_m(y')) \\ &= (\alpha f_0(x') + (1-\alpha)f_0(y'), \ldots, \alpha f_m(x') + (1-\alpha)f_m(y')) \end{aligned}$$

en omdat f_i, $0 \leq i \leq m$, convex zijn geeft dit

$$\geq (f_0(\alpha x' + (1-\alpha)y'), \ldots, f_m(\alpha x' + (1-\alpha)y')).$$

Het definitiegebied van f_i $(0 \leq i \leq m)$ is convex, dus als x' en y' in D liggen dan ook $w := \alpha x' + (1-\alpha)y'$ voor $\alpha \in [0,1]$. We hebben nu aangetoond dat $\alpha x + (1-\alpha)y \geq (f_0(w), \ldots, f_m(w))$. De verzameling C is dus convex.

Er geldt $0_{m+1} \notin intC$: Stel dat $0_{m+1} \in \text{int}C$. Dan bestaat $\varepsilon > 0$ zó dat $y = (-\varepsilon, 0, \ldots, 0)^\top \in C$. Uit de definitie van C volgt dan dat $-\varepsilon \geq f_0(y)$ en dit een contradictie is met de veronderstelling dat $f_0(\hat{x})$ het minimum van f_0 is.

Het Supporting-Hyperplane-Theorem levert nu dat er een $\lambda \in (\mathbb{R}^{m+1})^\top$ bestaat zó dat

$$\forall y \in C : \quad \lambda y \geq 0. \tag{7}$$

(β) Uit de definitie van C volgt dat 0_{m+1} met als gevolg dat $\mathbb{R}_+^{m+1} \subset C$. Dit leidt direct tot het bewijs van (β) [kies $(0, \ldots, 1_{\text{positie } j}, \ldots, 0) \in \mathbb{R}_+^{m+1}$, $1 \leq j \leq m$, en vul in (7) in; met dit argument volgt ook $\lambda_0 \geq 0$].

(γ) Stel dat $f_i(\hat{x}) < 0$ voor $i \geq 1$. Dan geldt

$$y = (0, \ldots, 0, f_i(\hat{x}), 0, \ldots, 0) \in C.$$

Uit (7) volgt dus dat $\lambda_i = 0$.

(α) Er geldt

$$\forall x \in \mathbb{R}^n : \quad (f_0(x), \ldots, f_m(x)) \in C$$

hieruit volgt met (7)

$$\mathcal{L}(x, \lambda) = \sum_{i=0}^{m} \lambda_i f_i(x) \geq 0.$$

Wij hebben verondersteld dat $f_0(\hat{x}) = 0$, uit (γ) volgt dan dat $\mathcal{L}(\hat{x}, \lambda) = 0$. Dus ligt bij $\hat{x}$ het globale minimum van $\mathcal{L}(\cdot, \lambda)$. Het theorema van Fermat geeft in dit geval $\mathcal{L}_x(\hat{x}) = 0$.

10 Practicum Week 5

Opdracht 1 (Exercise 4.6.7 uit B&V p.251)

A person is planning to divide her savings among three mutual funds, having expected returns of 10 %,10 % and 15 %. Her goal is a return of at least 12 %, while minimizing her risk. The risk function for an investment in this combination of funds is

$$200x_1^2 + 400x_2^2 + 100x_1x_2 + 899x_3^2 + 200x_2x_3,$$

where x_i is the proportion of her savings in fund i. Determine the proportions that should be invested in each fund. Would it help if she could go short, that is, if the x_i are allowed to be negative?

Opdracht 2 (Exercise 4.6.1 uit B&V p.250)

$$x_1^2 + x_2^2 + x_3^2 \to \text{extr},\ x_1 + x_2 + x_3 \leq 12,\ x_1, x_2, x_3 \geq 0$$

Opdracht 3 (Probleem 4.3.1 uit B&V p.217) Voor $a \neq 0_n$ los het volgende probleem

$$ax = \sum_{i=1} x_i a_i \to \max,\ |x|^2 = \sum_{i=1}^{n} x_i^2 = 1$$

11 De Lagrange Methode

11.1 Interpretatie van de λ-vector

Probleem $(P'_{3.1})$ Zij $f : D \to \mathbb{R}$, met $D \subset \mathbb{R}^n$:

$$f_0(x) \to \text{ extr}, \quad f_i(x) = b_i, \ 1 \le i \le m,$$

Stel dat de Lagrange methode kan worden toegepast op het bovenstaande probleem met $\hat{f}_i = f_i - b_i = 0$, $1 \le i \le m$. De oplossing y^* is een stationair punt van $\mathcal{L}(x, \lambda)$. De waarde functie $f_0(b) := f_0(y_1^*(b), \ldots, y_n^*(b))$ hangt van de waarde van de restrictie b af. Veronderstel dat $y^*(b)$ differentieerbaar is in b, dan is

$$\frac{\partial}{\partial b_j} f_0(b) = \lambda_j(b)\,, \quad 1 \le j \le m.$$

De Lagrangemultiplicator λ_j geeft dus aan hoe de waardefunctie in de optimale oplossing y^* van de constante b_j afhangt. Het getal $\lambda_j(b)$ word ook *schaduw prijs* of *marginale waarde* genoemd.

Het bovengenoemde resultaat is een voorbeeld van een *envelop theorema*. In het volgende wordt een volledig envelop theorema gegeven.

Theorem 15 *(Envelop theorema) De voorwaarden in Lagrange stelling zijn vervult. Voor $1 \le i \le m$, zij restrictie f_i afhankelijk van r_i en continu differentieerbaar in r_i. De Lagrangefunctie heeft een uniek stationair punt $x^* = x^*(r^*)$ voor $r = r^*$. Er geldt:*

$$\frac{\partial}{\partial r_i} f_0(r) = \left(\frac{\partial}{\partial r_i} \mathcal{L}(x, r) \right)_{x = x^*(r), r = r^*}, \quad 1 \le i \le m.$$

(Zonder bewijs.)

11.2 De algemene Stelling

Wij voeren op de productruimte $D \times \mathbb{R}^{m+1}$ de Lagrangefunctie $\mathcal{L}$ in, gedefinieerd door

$$\mathcal{L}(x, \lambda) = \sum_{i=0}^{m} \lambda_i f_i(x).$$

De rijvector λ wordt de Lagrange multiplicator genoemd. Wij veronderstellen dat de functies f_i, $0 \le i \le m$, differentieerbaar zijn in $\hat{x}$. Wij schrijven $\mathcal{L}_x(y, \lambda)$ voor de rijvector van partiële afgeleiden naar x in y:

$$\mathcal{L}_x(y, \lambda) = \left(\frac{\partial \mathcal{L}}{\partial x_1}(y, \lambda), \ldots, \frac{\partial \mathcal{L}}{\partial x_n}(y, \lambda) \right).$$

Probleem (P_5) Zij $f_0 : D \to \mathbb{R}$, met $D \subset \mathbb{R}^n$ een convexe verzameling:

$$\begin{array}{ll} & f_0(x) \to \text{ extr} \\ \text{onder} & f_i(x) \le 0, \ 1 \le i \le s, \\ & f_i(x) = 0, \ s+1 \le i \le m, \end{array}$$

noemen wij een n-dimensionaal optimaliseringsprobleem met nevenvoorwaarden.

Theorem 16 *Zij f_0 differentieerbaar in $y \in D$ en f_i, $1 \le i \le m$, continu differentieerbaar in y. Zij y een inwendig punt van D zó, dat $f(y)$ een lokaal extremum is van f op de verzameling*

$$T = \{x \in D : f_i(x) \le 0, 1 \le i \le s, f_i(x) = 0, s+1 \le i \le m\}.$$

Er geldt: er bestaat een Lagrange multiplicator $\lambda \neq 0_{m+1}$ zó, dat y een stationair punt is van $\mathcal{L}(\cdot, \lambda)$, oftewel

$$\mathcal{L}_x(y, \lambda) = 0_n^\top \Leftrightarrow \sum_{i=0}^{s} \lambda_i \frac{\partial f_i}{\partial x_j}(y) = 0\,, 1 \le j \le n,$$

en voor $1 \le i \le s$

$$\mathcal{L}_{\lambda_i}(y, \lambda) \begin{cases} = 0 & \text{voor} \lambda_i > 0 \\ \le 0 & \text{voor } \lambda_i = 0 \end{cases} \qquad \textit{(KKT)}$$

en voor $s+1 \le i \le m$

$$\mathcal{L}_{\lambda_i}(y, \lambda) = 0 \qquad (Lagrange).$$

N.B.: de vectoren $f_i'(y)$, $s+1 \le i \le m$, en voor $1 \le i \le s$, f_i' zo dat $\lambda_i > 0$ zijn lineair afhankelijk.

12 Iteratieve Methoden (voor Fermat/Lagrange)

Het doel van een optimaliseringsprobleem is het minimaliseren dan wel maximaliseren van een doelfunctie $f(x)$ door een vector van parameters x aan te passen. De verzameling $\mathbf{X} \subset \mathbb{R}^n$ bevat alle mogelijke waarden die de n-dimensionale vector x aan mag nemen, het toegelaten gebied. De doelfunctie heeft een één-dimensionale waarde, ofwel $f : \mathbb{R}^n \to \mathbb{R}^1$. Met deze notatie kunnen we het probleem kortweg schrijven als:

$$\min_{x \in \mathbf{X}} f(x). \tag{8}$$

Een minimalisatieprobleem kan eenvoudig omgezet worden naar een maximalisatieprobleem door het teken van de doelfunctie te veranderen. Een alternatieve manier om het probleem te beschrijven is het zoeken van de stationaire punten. Voorwaarde is dan dat f differentieerbaar is. Stationaire punten bevinden zich bij de nulpunten van de gradiënt $Df(x)$, de afgeleide naar x, zo dat

$$(Df(x))_i = \frac{\partial f(x)}{\partial x_i} = 0 \qquad 1 \le i \le n.$$

Een stationair punt is niet altijd het absolute, ofwel globale optimum van de doelfunctie. Er kan dan ook een lokaal optimum gevonden zijn. Als de functie echter convex is, dan zal er slechts een optimum zijn.

Als $f(x)$ en $Df(x)$ niet bekend zijn dan wordt de optimalisatie wordt doorgaans uitgevoerd met behulp van een algoritme dat de parameters stap voor stap aanpast vanaf een beginwaarde, naar een verbeterde waarde van de doelfunctie. De beginwaarde is zelden gegeven en wordt daarom vaak heuristisch gekozen. De gradiënt wordt gebruikt om de zoekrichting te bepalen en het probleem beter hanteerbaar te maken.

12.1 Newton-Raphston

Sectie 4.6 in R. Adams, p. 245.

12.2 Steepest Decent Algorithmen

Steepest Decent Algorithmen (SDA) is een klasse van optimalisatie algoritmes die door Robbins en Monro geïntroduceerd werd. In het SDA algoritme wordt de parameter iteratief bijgewerkt om een nulpunt van de gradiënt van $f(x)$ te vinden. Indien de doelfunctie convex is, wordt het globale optimum gevonden, anders een lokaal optimum. De schatting voor x^*, de lokatie van het optimum, bij de k-de iteratie zal genoteerd worden met x_k , $k = 0, 1, 2, \ldots$, zodat

$$\lim_{k\to\infty} x_k = x^*. \tag{9}$$

De algemene vorm van het SDA is als volgt:

$$x_{k+1} = x_k - a_k Df(x_k). \tag{10}$$

De rij $\{a_k\}$ wordt de *gain sequence* genoemd. Bij het kiezen van de niet-negatieve rij $\{a_k\}$ moet een rij gevonden worden die naar nul gaat naarmate de oplossing dichterbij komt, maar waarbij vroegtijdige (valse) convergentie wordt vermeden. Als a te klein is kruipt het algoritme naar het optimum. Als a echter te groot gekozen wordt kunnen er extreme schommelingen plaatsvinden, wat leidt tot een instabiel algoritme. Het algoritme convergeert indien de rij voldoet aan twee voorwaarden:

$$\sum_{k=0}^{\infty} a_k = \infty \qquad\qquad \sum_{k=0}^{\infty} a_k^2 < \infty. \tag{11}$$

Een geschaalde harmonische rij van de vorm $\{a/k+1\}$ met $a > 0$ is een veel gebruikt voorbeeld van een rij die aan deze voorwaarden voldoet.

Indien het toegelaten gebied niet de gehele $\mathbb{R}^n$ omvat, kan het gebeuren dat het algoritme een oplossing x^* vindt die buiten het toegelaten gebied $\mathbf{X}$ valt. Vaak vindt optimalisatie plaats met randvoorwaarden, waardoor θ bijvoorbeeld binnen een bepaald interval moet vallen. Om deze reden wordt de volgende variatie op het algoritme gebruikt:

$$x_{k+1} = \Pi_{\mathbf{X}}(x_k - a_k Df(x_k)). \tag{12}$$

Hierin is de functie $\Pi_{\mathbf{X}}$ een projectie van x_k terug in het toegelaten gebied $\mathbf{X}$ als het algoritme x_k hierbuiten plaatst.

12.3 Approximatie/Schatting van de Gradiënt

Er zijn vele modellen waarbij het niet mogelijk is de gradiënt van de doelfunctie met behulp van wiskundige analyse te vinden. Voor zulke modellen bestaan er methoden die beter geschikt zijn dan (12). De indirecte (of geschatte waarde) van de gradiënt is daarbij enkel gebaseerd op metingen van de doelfunctie. Het SDA algoritme in methoden met een gradiëntschatter heeft doorgaans de volgende twee eigenschappen:

1. Er wordt enkel een *benadering* geschat van de ware waarde van de gradiënt.
2. Er wordt enkel gekeken naar de output van het systeem dat bestudeerd wordt, zonder hierbij de onderliggende afhankelijkheid van de input te analyseren.

Op basis van het aantal iteraties convergeren algoritmen waarbij enkel doelfunctiewaarden gebruikt worden om de gradiënt te schatten vaak langzamer. Echter, omdat het lastig kan zijn om de exacte relatie tussen de parameters (input) en de doelfunctie (output) te vinden is een vergelijking in snelheid niet enkel te baseren op het aantal iteraties. Er kunnen grote besparingen in rekentijd zijn door de doelfunctiewaarde te berekenen in plaats van de gradiënt.

12.4 Zuivere schatters (Stochastische Variabelen)

Naast bovengenoemde deterministische versie van het SDA algoritme van Robbins en Monro bestaat er een stochastische variant. Het resultaat in (10) geldt ook als $Df(x)$ vervangen wordt door een zuivere gradiëntschatter $g(\theta)$ voor $Df(x)$. Dat wil zeggen dat $E[Df(x)] = g(x)$. In geval van een zuivere schatter leest het algoritme als volgt:

$$x_{k+1} = x_k - a_k g(x_k). \tag{13}$$

De convergentie in (9) is dan met kans 1. Het algorithme in (13) heet stochastische approximatie (SA).

12.5 Onzuivere schatters (Stochastische en Deterministische Variante)

Indien een schatter niet zuiver is, wordt deze onzuiver genoemd. In dat geval geldt dat

$$E[g(x)] = g(x) + \text{ fout }.$$

De klassieke aanpak voor de benadering van de gradiënt vector is via ’finite differences’ en volgt uit de definitie hiervan als een vector van p partiële afgeleiden. De definitie van een partiële afgeleide van de functie $f : \mathbb{R}^n \to \mathbb{R}$ van de variabelen $x_1, ..., x_n$ naar variabele x_i is als volgt:

$$\frac{\partial}{\partial x_i} f(x_1, \ldots, x_n) = \lim_{c \to 0} \frac{f(x_1, \ldots, x_{i-1}, x_i + c, x_{i+1}, \ldots, x_n) - f(x_1, \ldots, x_n)}{c}.$$

Er bestaan zowel één- als twee-zijdige finite difference gradiënt schatters. Bij de eerste soort wordt de gradiënt geschat aan de hand van metingen van de doelfunctie bij x en bij (x + een perturbatie). Bij tweezijdige schatters wordt de gradiënt geschat aan de hand van metingen van de doelfunctie bij ($x\pm$ een perturbatie). Bij n parameters zijn er respectievelijk $n+1$ en $2n$ metingen van de doelfunctie nodig per schatting. Ondanks de langere rekentijd die nodig is voor een tweezijdige schatter, wordt deze wel vaak gebruikt omdat de uitkomsten hierbij stabieler zijn. Hieronder worden verschillende SDA algoritmen die een onzuivere gradiëntschatter gebruiken besproken.

12.5.1 Kiefer-Wolfowitz Stochastische Aproximatie

Het prototype algoritme dat werkt met een onzuivere gradiëntschatter is het finite difference SDA (FDSDA) algoritme. Het heeft dezelfde vorm als het algoritme van Robbins en Monro (10). Het essentiële verschil tussen de algoritmen zit in de benadering van de gradiënt.

In het algoritme van Kiefer en Wolfowitz wordt gebruik gemaakt van een onzuivere tweezijdige schatter. Analoog aan de definitie van de partiële afgeleide wordt in het algoritme bij iedere iteratie één van de elementen in de vector met parameters aangepast, ofwel gepertubeerd, terwijl de andere waarden gelijk gehouden worden. Op dit punt wordt de waarde van de doelfunctie bepaald. Stel dat we kijken naar het i-de component van de vector met n parameters. Het i-de component van de gradiënt schatter wordt dan berekend door het verschil in functiewaarden te delen door het differentie interval, de perturbatie. In formulevorm ziet de gradiëntschatter voor het i-de component in het punt x er als volgt uit:

$$\frac{\partial}{\partial x_i} f(x_1, \ldots, x_p) \approx g_{k,i}(x_k) = \frac{\hat{f}(x_k + c_k e_i) - \hat{f}(x_k - c_k e_i)}{2c_k}. \tag{14}$$

In het stochastisch geval is $\hat{f}$ een zuivere schatter voor f en anders $\hat{f} = f$. De vector e_i is de i-de eenheidsvector, dus het i-de element heeft de waarde 1 en op de andere plaatsen staan nullen. Deze vector geeft de zoekrichting in de i-de richting. De vector c bevat de verschillen, waarbij c_k de perturbatie in de k-de stap van het algoritme is. De rij $\{c_k\}$ moet net als de rij $\{a_k\}$ uit (10) naar nul gaan op een gepast tempo. Door $\{c_k\}$ steeds kleiner te laten worden zal de fout die gemaakt wordt door $Df(x)$ te vervangen door bovengenoemde $g(x)$ naar 0 gaan als $k \to \infty$. Voor een partiële afgeleide wordt immers de limiet van $c \to 0$ genomen. Men zegt dan dat de gradiëntschatter asymptotisch in k zuiver is. Wanneer $\{a_k\}$ aan (11) voldoet, dan is $\{c_k\} = c/(k+1)^\gamma$ zó dat voldaan wordt aan de volgende voorwaarden:

1. $c_k \to 0 \quad \text{als} \quad k \to \infty$
2. $$\sum_{k=0}^{\infty} a_k c_k < \infty \qquad\qquad \sum_{k=0}^{\infty} a_k^2/c_k^2 < \infty \tag{15}$$

De asymptotisch optimale waarde voor γ is $1/6$.

12.5.2 Simultane Perturbaties

Een alternatieve onzuivere gradiëntschatter wordt gebruikt in de zogenaamde Simultane Perturbatie Stochastische Approximatie (SPSA) methode. Het grote voordeel van dit algoritme zit in de benodigde rekentijd. In tegenstelling tot het algoritme van Kiefer en Wolfowitz is deze voor dit algoritme niet afhankelijk van het aantal parameters n. De parameters worden hier namelijk niet stuk voor stuk geperturbeerd, maar gelijktijdig, ofwel simultaan. Aan de hand van tweezijdige random perturbaties uit de n-dimensionale random perturbatie vector Δ_k wordt de doelfunctie twee keer berekend. Elk component van de geschatte gradiënt-vector wordt bepaald door het verschil tussen deze twee metingen te delen door de perturbatie van de individuele componenten van de perturbatievector. De teller is hierbij voor elke parameter uit de vector gelijk. In formulevorm:

$$\frac{\partial}{\partial x_i} f(x_1, \ldots, x_p) \approx g_{k,i}(x_k) = \frac{\hat{f}(x_k + c_k \Delta_k) - \hat{f}(x_k - c_k \Delta_k)}{2 c_k \Delta_{k,i}}. \qquad (16)$$

Deze gradiëntschatter wordt vervolgens gebruikt in het SA algoritme(10).

De gain sequences hebben de volgende vorm:

$$a_k = a/(A + k + 1)^{\alpha} \qquad\qquad c_k = c/(k + 1)^{\gamma}$$

De rij $\{a_k\}$ uit het SDA algoritme (10) bepaalt wederom de aanpassing van de parametervector in de tegengestelde richting van de gradiënt en moet voldoen aan (11). Vaak gebruikt men hiervoor een aangepaste reeks, waarin hij een stabiliteitsconstante A toevoegt. Daarbij voegt hij een parameter α toe. De asymptotisch optimale waarde voor α is 1, maar kan in praktijk soms beter anders gekozen worden. De uit het FDSA algoritme bekende c bepaalt de grootte van de perturbatie voor elk component. In het geval van SPSA is dit geen vector, maar een diagonaalmatrix met de waarden uit de vector c op de diagonaal en voldoet aan (15). Hierdoor is ook SPSA asymptotisch in k zuiver. Voor convergentie moeten er behalve aan beide gain-sequences ook eisen gesteld worden aan de verdeling van de perturbatie vector Δ_k. De vector Δ_k bepaalt de zoekrichting in de k-de stap van het algoritme. De $\Delta_{k,i}$'s moeten onafhankelijk en symmetrisch verdeeld zijn rond 0 met eindigende inverse tweede momenten ($E[|\Delta_{k,i}|^{-1}]$) voor alle k en i. Een symmetrische Bernoulli verdeling waarbij met kans 0.5 de waarde 1 en met gelijke kans de waarde -1 wordt aangenomen voldoet aan deze eisen. De veelgebruikte uniforme- en normale verdeling voldoen niet. Deze verdelingen hebben namelijk geen eindige inverse momenten.

Het feit dat het aantal berekeningen van de doelfunctie bij SPSA onafhankelijk is van het aantal parameters, maakt deze methode zeer interessant voor problemen waarbij over een groot aantal parameters geoptimaliseerd moet worden. Dit voordeel geldt mits het aantal benodigde iteraties om tot een optimum te komen niet sterk toeneemt.

13 Iteratieve Methoden (voor Lagrange)

Probleem (P_6) Zij $f_0 : D \to \mathbb{R}$, met $D \subset \mathbb{R}^n$ een convexe verzameling:

$$\begin{array}{ll} & f_0(x) \to \ \max \\ \text{onder} & f_i(x) \leq 0,\ 1 \leq i \leq s, \\ & f_i(x) = 0,\ s+1 \leq i \leq m, \end{array}$$

noemen wij een n-dimensionaal optimaliseringsprobleem met nevenvoorwaarden.

Wij voeren op de productruimte $D \times \mathbb{R}^m$ de Lagrangefunctie $\mathcal{L}$ in, gedefinieerd door

$$\mathcal{L}(x, \lambda) = f_0(x) - \sum_{i=1}^{m} \lambda_i f_i(x).$$

Een punt (x^*, λ^*) heet zadelpunt las geldt

$$\mathcal{L}(x, \lambda^*) \leq \mathcal{L}(x^*, \lambda^*) \leq \mathcal{L}(x^*, \lambda)$$

voor ieder $(x, \lambda) \in D \times [0, \infty)^m$. Als (x^*, λ^*) een zadelpunt is dan is (x^*, λ^*) een oplossing voor (P_6).

Één oplossing voor (P_6) kan dus worden gevonden door simultaan de x-componenten te maximaliseren en in de λ-componenten te minimaliseren.

Zij (x^k, λ^k) gegeven als volgt:

$$x_i^{k+1} = x_i^k - \epsilon_k \left(\frac{\partial}{\partial x_i} f(x^k) + \sum_{i=1}^{m} \lambda_i^k \frac{\partial}{\partial x_i} f_i(x^k) \right)$$

en

$$\lambda_i^{k+1} = \lambda_i^k + \rho_k f_i(x^{k+1})$$

modulo projectie op $D \times [0, \infty)^m$, met

$$0 < \sum_k \epsilon_k = \infty, \sum_k (\epsilon_k)^2 < \infty$$

en

$$0 < \sum_k \rho_k = \infty, \sum_k (\rho_k)^2 < \infty.$$

Dan geldt

$$\lim_{k \to \infty} (x^k, \lambda^k) = (x^*, \lambda^*).$$

14 Practicum Week 6 (Tentamen, 2007)

Opgave 1 De functie f op $\mathbb{R}$ is gedefinieerd door

$$f(x, y) = y^2 + xy.$$

(a) Stel vast of de functie differentieerbaar is. Bepaal telkens de gradiënt van f in het punt $(2, 1)$.

(b) Bepaal de tangentenvlakte van f in het punt $(2, 1)$.

(c) Bepaal de stationaire punten van f. Bepaal vervolgens met behulp van figuur 1 de aard van de punten.

Opgave 2 De functie f op $\mathbb{R}^2$ is gedefinieerd door

$$f(x, y) = \ln(1 + x^2 + y^2).$$

(a) Bepaal met behulp van de vier stappen methode de extreme waarden van f en de (x, y)-waarden waarin zij worden aangenomen.

(b) Pas het SDA algorithme toe met gain sequence $a_k = a/k + 1$ voor $a = 2$ en begin waarde $z_0 = (x_0, y_0) = (1, 1)$. Bereken z_1, z_2.

Opgave 3 De functie f is gedefinieerd door

$$f(x, y) = e^x \ln(1 + y).$$

(a) Bepaal de gradiënt en de Hessiaan van f. Voor welke waarden van (x, y) is de functie differentieerbaar?

(b) De stelling van Taylor geeft aan dat

$$f(x, y) \approx f(0, 0) + Df(0, 0)(x, y)^\top + \frac{1}{2}(x, y)Hf(0, 0)(x, y)^\top.$$

Wat zijn de voorwaarden om de stelling van Taylor toe te kunnen passen? Bepaal de benadering voor de gegeven functie.

Opgave 4 De functie f op $\mathbb{R}^2$ is gedefinieerd door

$$f(x, y) = (x + y)^2 - 2xy.$$

Bepaal de minimale waarde van f

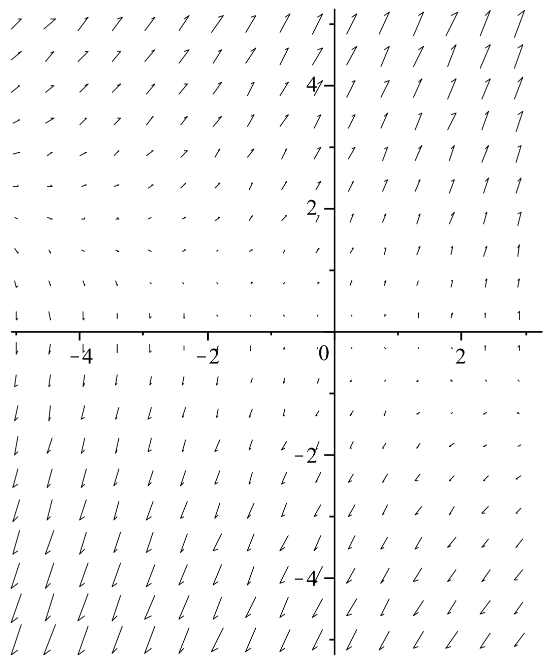

Figure 1: Gradiëntveld van $f(x,y) = y^2 + xy$

(a) onder de nevenvoorwaarde

$$x^2 + xy + y^2 - 3 = 0\,,$$

(b) onder de nevenvoorwaarde

$$x^2 + xy + y^2 - 3 \leq 0\,.$$

Opgave 5 De functie f op $\mathbb{R}^2$ is gedefinieerd door

$$f(x, y) = (1 + y)^3 x^2 + y^2.$$

Laat zien dat $(0, 0)$ het unieke stationair punt van f is en dat $(0, 0)$ een locaal minimum is. Heeft f een globaal minimum?